Diplomarbeit

3D–Objekterkennung: Repräsentation und Erkennungsverfahren hinsichtlich der Verwendung unterschiedlicher Farbmerkmale

CV–Bericht 55

Wolfram J. Schimke
Matrikel–Nr. 131 760

Berlin, 19. September 1996

Prof. Dr. rer. nat. Reinhard Klette
Fachgruppe Computer Vision
Institut für Technische Informatik
Fachbereich 13 (Informatik)
Technische Universität Berlin

Betreuer: Dr.–Ing. Gunter Bellaire

Kurzfassung

In dieser Arbeit wird das auf *Surface Normal Images* basierende Objekterkennungssystem PolyRec vorgestellt. Das System wurde zur Sortierung von gebrauchten Verpackungen aus Kunststoffen entwickelt.

Es wird eine 2.5D–Rekonstruktion des Objekts mittels *Photometric Stereo* berechnet. Die Objekterkennung basiert auf dem Vergleich der Aufdrucke und Texturen planarer Oberflächenstücke mit in der Datenbank gespeicherten Bildern. Es werden verschiedene Farbmerkmale, u.a. Farbhistogramme und ihre Invarianten sowie statistische Eigenschaften von Farbverteilungen, teilweise auch unter Einbeziehung von Information über die räumliche Verteilung der Farben, verwendet.

Ein Schwerpunkt der Arbeit liegt bei der Entwicklung, der Auswahl und der Untersuchung von Algorithmen zum Vergleich von Bildern, speziell auch unter dem Gesichtspunkt der Toleranz gegenüber Helligkeitsschwankungen und Änderungen der Beleuchtungsfarbe.

Ergebnisse ausführlicher Experimente mit synthetischen und realen Bildern werden gezeigt.

Es wird ein Verfahren zur Bildaufnahme der Oberflächentextur von 3D–Objekten basierend auf einem vollständigen 3D–CAD–Datensatz des bearbeiteten Objekts vorgestellt und getestet. Dieses Verfahren dient zur automatischen Erstellung der Bilder der Datenbank für das Objekterkennungssystem PolyRec.

Ich danke allen, die mich bei der Anfertigung dieser Arbeit unterstützt haben, mit der Bereitstellung von Quelltexten, Programmen oder Aufnahmen, der Erläuterung von Algorithmen und nicht zuletzt durch hilfreiche Kommentare zum vorliegenden Text.

Abstract

In this paper the object recognition system PolyRec based on surface normal images is introduced. The system is designed for the automatic sepration of plastic salvage.

A 2.5D reconstruction of the object's surface is computed, currently using photometric stereo. Imprints and texture of planar patches of the object's surface are matched against images stored in the database. Different color features, e.g. color histograms and their invariants and statistical properties of color distributions, partly together with spatial information, are used for the recognition.

Special emphasis is put on the development, the selection and the comparative study of image matching algorithms, especially with concern to tolerance for deviations in lightness and color of the illumination.

Results of experiments with synthetic and real images are presented.

A 3D texture scan method based on a complete 3D CAD data set of the examined object is proposed and tested. The method is used for the automated generation of the images of the database for the PolyRec object recognition system.

Inhaltsverzeichnis

Abbildungsverzeichnis

Tabellenverzeichnis

1 Einleitung

1.1 Objekterkennungssysteme

Die Erkennung dreidimensionaler Objekte ist ein Teilgebiet der Szenenanalyse (*image understanding*) und ein Hauptforschungsgebiet von Computer Vision. Unter Objekterkennung versteht man die Markierung von Merkmalen in einer Szene als Instanz eines Objekts (oder auch einer Klasse von Objekten), siehe auch [3]. Die Merkmale können z.B. geometrischer oder topologischer Art sein, wie Ecken und Kanten oder Formen von Flächen. Eine weitere wichtige Klasse von Merkmalen sind die Farb- und Intensitätsmerkmale, wie z.B. Aufdrucke auf der Objektoberfläche.
Bei einer modellbasierten Objekterkennung werden die aus der Szene extrahierten Merkmale einer digitalen Repräsentation des Objekts — dem Modell — zugeordnet. Dabei wird zwischen einer datengesteuerten (*data–driven*) und einer modellgesteuerten (*model–driven*) Strategie unterschieden. Bei der datengesteuerten Strategie wird von den Szenendaten ausgehend ein „passendes“ Objekt gesucht. Bei der modellgesteuerten Strategie hingegen werden die zu einem Modell eines Objekts gespeicherten Merkmale in der Szene gesucht. Diese Strategie eignet sich besonders zur Lagebestimmung von Objekten, wobei oftmals die eigentliche Objektidentifikation mit einer datengesteuerten Zuordnung erfolgt.

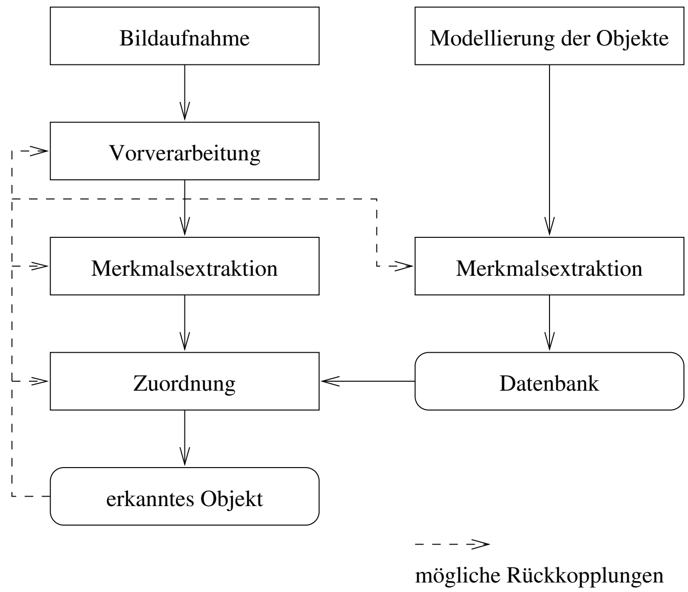

Abbildung 1: Verarbeitungspipeline und Komponenten eines Objekterkennungssystems: Der linke Teil wird für jedes zu erkennende Objekt durchlaufen, der rechte Teil stellt die Erzeugung der Datenbank dar.

Den allgemeinen Ablauf der Objekterkennung zeigt Abb. 1 auf der vorherigen Seite. Aufgrund der Struktur des Datenpfades in einem Objekterkennungssystem wird der Ablauf auch als Pipeline (Erkennungs- oder Verarbeitungspipeline) bezeichnet.
Ein typisches Merkmal einer Pipeline ist, daß ein Engpaß in einer Verarbeitungsstufe zu einem Engpaß im Gesamtsystem führt. Daher sind bei der Optimierung alle Komponenten des Systems zu beachten. Lediglich die Verfahren zur Bereitstellung der Datenbank sind in der Regel unkritisch bezüglich der Optimierung.
Rückkopplungen im Objekterkennungssystem dienen der Adaption an sich ändernde oder im voraus nicht bekannte Bedingungen. Bei der Zuordnung kann die Rückkopplung z.B. in einer Anpassung der Strategie der Suche im Merkmalsraum an die aktuelle bzw. tatsächliche Merkmalsverteilung bestehen.
Wichtige Unterscheidungsmerkmale von Objekterkennungssystemen sind auch ihre Einschränkungen bezüglich der Beleuchtung (natürliche Beleuchtung oder unter Laborbedingungen), der Komplexität der Szene und der Art der Objekte.

Repräsentationen für Modelle dreidimensionaler Objekte

Für die bei einer modellbasierten Objekterkennung erforderlichen Modelle der dreidimensionalen Objekte sind verschiedene digitale Repräsentationen möglich. Jede hat ihre speziellen Vorteile und somit auch bestimmte Anwendungsgebiete, für die sie geeignet ist.
3D–Modelle (z.B. *volume representation*, *surface boundary representation*, *generalized cylinders*) sind kompakt und lassen sich einfach verwalten. Sie enthalten jedoch keine projektive Information, die für den Vergleich mit den zweidimensionalen Kamerabildern erforderlich ist. Wird die projektive Information modelliert, so erhöht sich meist die zu verwaltende Datenmenge. Ein typisches Beispiel sind charakteristische Ansichten (*characteristic views*). Bei den Normallagen–Bildern (*surface normal images*) wird hingegen trotz der Modellierung projektiver Information eine recht kompakte Darstellung erzielt.
Eine vergleichende Darstellung verschiedener Repräsentationen dreidimensionaler Objekte und ihrer Eignung für Objekterkennungssysteme findet sich in [3].

1.2 Aufgabenstellung für das Objekterkennungssystem

Ziel war die Entwicklung eines Systems zur automatischen Erkennung von gebrauchten Kunststoffverpackungen (siehe [29]), speziell von solchen mit dem „Grünen Punkt“. Gemäß der Verpackungsverordnung müssen in Deutschland gebrauchte Verpackungen der Wiederverwertung zugeführt

werden. Dazu werden die gebrauchten Verpackungen vom Dualen System (DSD) bei den Haushalten abgeholt. Für diese Dienstleistung ist pro Verpackung eine dem Aufwand entsprechende Gebühr durch den Hersteller der Verpackung zu zahlen. Als Beleg ist der Aufdruck des „Grünen Punktes" auf der Verpackung anzusehen.
In den Haushalten findet nur eine grobe Sortierung der Verpackungen nach den Materialen (Papier und Pappe, Glas, Metall und Kunststoffe) statt. Die Kunststofffraktion muß aufgrund der Verschiedenartigkeit der Materialien noch weiter sortiert werden, um sie dem Recycling zuführen zu können. Eine Möglichkeit dazu ist, zunächst die Verpackung (das Objekt) zu erkennen. In einer Datenbank könnte dann die Stoffklasse ermittelt und somit die Verpackung den entsprechenden Prozessen zugeführt werden.
In einem Projekt [29] wurde ein Prototyp eines Objekterkennungssystems für diese Aufgabe entwickelt und untersucht. Das Objekterkennungssystem soll die Objekte anhand ihrer farbigen Aufdrucke identifizieren. Dazu wird eine 2.5D–Rekonstruktion des Objekts durchgeführt, um die Aufdrucke der einzelnen Flächen des Objekts getrennt den Bildern der Aufdrucke in der Datenbank zuzuordnen. Die Zuordnung erfolgt anhand von Farbmerkmalen der Bilder. Aus der Zuordnung auf der Ebene der Aufdrucke bzw. Bilder wird dann eine Zuordnung auf der Ebene der Objekte bestimmt.
Diese Arbeit dient der vertiefenden Untersuchung von Teilaspekten dieses Objekterkennungssystems. Es werden weitere Verfahren zur Zuordnung von Bildern rekonstruierter Flächen zu Bildern aus der Datenbank und die Einführung von Verfahren zum Ausgleich von Helligkeits- und Farbschwankungen der Beleuchtung untersucht. Weiterhin wird die weitgehend automatische Erstellung der Datenbank mit Hilfe eines 3D–Texture–Scanning–Verfahrens vorgestellt.

1.3 Ablauf der Objekterkennung im realisierten System

Ausgehend von der Aufnahme des zu erkennenden Objekts wird in diesem Abschnitt kurz der Ablauf des Objekterkennungsverfahrens beschrieben[1]. Eine Skizze des Ablaufs zeigt Abb. 2 auf der nächsten Seite.
Das hier vorgestellte Objekterkennungsverfahren basiert auf einer zumindest teilweisen 2.5D–Rekonstruktion des Objekts. Es wird ein Farbbild des Objekts mit der zugehörigen Linienzeichnung der rekonstruierten Flächen[2] und deren Orientierung benötigt.
In [29] haben wir für die 2.5D–Rekonstruktion ein PHOTOMETRIC–STEREO–Verfahren (*photometric stereo method*, PSM–Verfahren) eingesetzt. PSM–Verfahren haben den Vorteil, daß sie für jeden ausreichend hellen Objektpunkt die Orientierung berechnen können. Die Rekonstruktion der einzel-

[1] Eine genaue Beschreibung des Objekterkennungssystems enthält Abschnitt 6.

[2] Unter „Flächen" sind hier die einzelnen planaren Oberflächenstücke des Objekts zu verstehen.

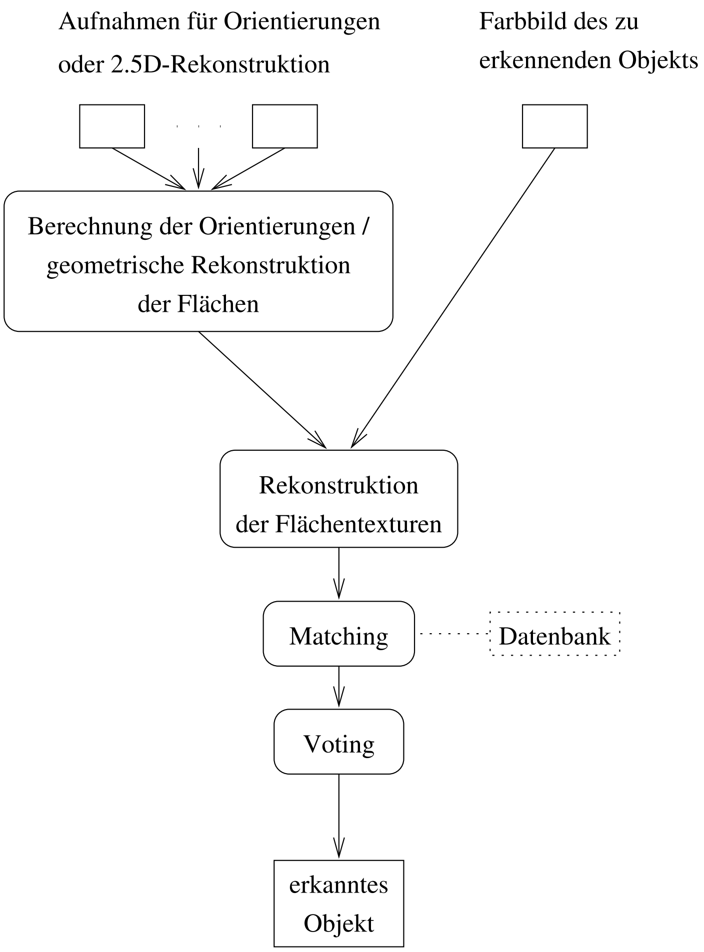

Abbildung 2: „Datenfluß“ bei der Erkennung von Objekten

nen Flächen (geometrische Rekonstruktion) erfolgte dann mit dem in [17] beschriebenen Verfahren von Arno Mitritz. Eine genaue Beschreibung ist in [29] zu finden.

Die für die Tests des vorgestellten Verfahrens erforderlichen Aufnahmen der zu erkennenden Objekte und die Ergebnisse der geometrischen Rekonstruktion sind bereits im Rahmen von [29] erstellt worden und wurden hier ebenfalls verwendet. Daten zur Aufnahmekonfiguration etc. sind dieser Referenz zu entnehmen. Alternative Verfahren zur Bereitstellung dieser Daten werden in Abschnitt 3.5 kurz dargestellt.

Die Ergebnisse der geometrischen Rekonstruktion der Flächen einschließlich der Bestimmung ihrer Orientierungen werden zusammen mit der Farbaufnahme derselben Ansicht des zu erkennenden Objekts zur Rekonstruktion der Aufdrucke (Texturen) der Flächen verwendet. Diese werden dann den Algorithmen zum Vergleich mit den in der Datenbank gespeicherten Bil-

dern der Objektflächen (Matching–Prozeß) zugeführt. Die einzelnen (in Abschnitt 5 vorgestellten) Algorithmen berechnen und verwenden verschiedene Farbmerkmale der Bilder zum Vergleich.
Die im Matching–Prozeß erhaltenen Daten werden anschließend ausgewertet, um das aufgenommene Objekt einem Objekt der Datenbank zuzuordnen (Voting). Es wird dabei eine Rangliste der Objekte der Datenbank erstellt.

1.4 Bezug zu anderen Arbeiten

Die im vorgestelleten Objekterkennungssystem verwendete Objektrepräsentation basiert auf den von J. Park, T. Chang und J. Choi in [30] vorgestellten *surface normal images*, die für diese Anwendung erweitert wurden. Die Untersuchungen stützen sich auf Arbeiten, die zuvor an der TU Berlin erfolgt sind [17, 29, 4].
Die Verfahren zum Vergleich von Bildern (rekonstruierte Aufdrucke mit Bildern der Datenbank) basieren zum Teil auf dem von M. Swain und D. Ballard in [46] dargestellten *color indexing*, das bereits in früheren Arbeiten an der TU Berlin untersucht wurde [36, 29]. Es wurden insbesondere auch von M. Stricker et al. diskutierte Erweiterungen [43] untersucht. Weitere untersuchte Verfahren sind die ebenfalls von M. Stricker et al. vorgeschlagene Verwendung statistischer Eigenschaften der Farbverteilung [43], die von G. Healey und D. Slater dargelegte Verwendung von Invarianten von Farbhistogrammen [18] und die Integration von Information über die räumliche Verteilung der Farben [42, 34]. Die Verfahren werden in dieser Arbeit in bezug auf die vorgestellte Anwendung im Objekterkennungssystem verglichen. Erweiterungen werden vorgestellt und diskutiert.
Der Vergleich der Aufdrucke mit den Bildern der Datenbank kann auch als eine Abfrage einer Bilddatenbank aufgefaßt werden. Das Ziel dabei ist, aus der Datenbank diejenigen Bilder zu selektieren, die hinreichend viele ähnliche Merkmale wie das Abfragebild haben (*content–based image retrieval*, *query by example*). Siehe dazu auch die Abhandlung im Ausblick, Abschnitt 8.5.

2 Allgemeine Grundlagen

Zunächst werden Grundlagen zu Farbe und der Farbbildverarbeitung behandelt. Danach folgt die Vorstellung von Normallagen–Bildern (*surface normal images*, SNI) und ihren Erweiterungen (ESNI). Darstellungen zu den in dieser Arbeit verwendeten Koordinatensystemen und weitere grundlegende mathematische Definitionen befinden sich im Anhang A.

2.1 Farbe, Farbenraum und Farbmodelle

2.1.1 Der Begriff „Farbe"

Farbe ist ein Merkmal von Körpern bzw. deren Oberflächen[3]. Es gibt zwei Herangehensweisen an den Begriff „Farbe". Er kann zum einen physikalisch–technisch über die physikalischen Eigenschaften des Körpers in bezug auf elektromagnetische Wellen (Emission, Absorption und Reflexion) in einem bestimmten Wellenlängenbereich definiert werden und zum anderen sinnesphysiologisch als Wahrnehmung durch das menschliche Auge[4]. Aber auch bei einer physikalisch–technischen Definition wird der betrachtete Wellenlängenbereich oft als der Bereich des „sichtbaren Lichts" bezeichnet, was wieder auf eine durch die menschliche Wahrnehmung beeinflußte Definition hinausläuft.
Gerade bei der Objekterkennung, deren Aufgabe nicht selten der Ersatz oder die Ergänzung menschlicher Tätigkeiten ist, ist die sinnesphysiologische Bedeutung der Farbe relevant. Die Farbgebung der Objekte ist z.B. bei Verpackungen stark am menschlichen Farbempfinden ausgerichtet. Nichtzuletzt ist auch die Verwendung trichromatischer Systeme[5] in der Technik eine direkte Folge des trichromatischen menschlichen Auges. Eine ausschließliche Definition über das menschliche Sehsystem und die menschliche Wahrnehmung ist jedoch aufgrund ihrer Subjektivität für ein Objekterkennungssystem ungeeignet.
Man unterscheidet in der Farbmetrik selbstleuchtende und nicht selbstleuchtende Körper in Abhängigkeit davon, ob der Körper Licht emittiert oder nicht. Bei den sogenannten Nichtselbstleuchtern ist eine Wahrnehmung der Farbe nur bei Beleuchtung möglich. Die Geometrie und die spektrale Energieverteilung der Beleuchtung beeinflussen die wahrgenommene (bzw. gemessene) Farbe. Von Interesse ist hier eine beleuchtungsunabhängige Bestimmung der Reflexionseigenschaften des Körpers (bzw. der Körperoberfläche).

[3] In der deutschen Sprache wird der Begriff „Farbe" auch als Bezeichnung für färbende Substanzen (Pigmente und Farbstoffe) verwendet. Diese Bedeutung ist jedoch nicht gemeint, wenn in dieser Arbeit von Farbe die Rede ist.

[4] In der Farbmetrik wird in der Regel nicht von Farbwahrnehmungen sondern von Farbvalenzen gesprochen.

[5] Systeme mit drei Klassen von Sensoren von unterschiedlicher spektraler Empfindlichkeit

Weitere Ausführungen zum Begriff „Farbe“ siehe [33] und zur menschlichen Farbwahrnehmung siehe [24, 19, 33].

2.1.2 Farbmischung, Farbkoordinaten und Farbenraum

Viele, aber nicht alle Farben lassen sich — bezüglich der menschlichen Wahrnehmung oder der üblichen Messung mit drei Sensorklassen — als Mischung von drei Grundfarben[6] darstellen. Darauf basiert z.B. die Farbwiedergabe im Fernsehen oder beim Drei- bzw. Vierfarbdruck. Allgemein gilt jedoch: Jede Farbe läßt sich als Linearkombination von drei Grundfarben darstellen (*1. Graßmann'sches Gesetz*, siehe [24, 13, 19, 33, 49, 50]), d.h. es können auch *negative* Anteile oder Anteile über 100% einer Grundfarbe vorkommen. Dies trifft z.B. auf sehr reine Farben, wie die von Laserlicht, zu. Die Koeffizenten der Linearkombination werden als Farbkoordinaten bezeichnet. Die Grundfarben sind die Basis (nicht unbedingt orthonormal) des zugehörigen Farbenraums.

Die spektrale Reflektanzkurve enthält mehr Information als die Farbkoordinaten. Es werden mehrere verschiedene spektrale Reflektanzkurven auf einen Punkt im Farbenraum abgebildet. Genau diese Informationsreduktion tritt auch bei einer Messung der Farbe mit der Kamera oder dem menschlichen Auge ein. Drei verschiedene Sensorklassen (bei der Kamera Rot, Grün und Blau) bestimmen die Koordinaten der Farbe des einfallenden Lichts bezüglich des durch die spektralen Empfindlichkeiten der Sensoren definierten Farbenraums.

Darauf aufbauend lassen sich zwei Definitionen für die Gleichheit von Farben angeben. Zum einen läßt sie sich als Gleichheit der Farben bezüglich einer Beobachtung (durch das menschliche Auge oder ein technisches Meßgerät, insbesondere eine Kamera) oder als Gleichheit der spektralen Energieverteilung definieren. Zwei bezüglich einer Beobachtung gleiche Farben mit unterschiedlicher spektraler Energieverteilung heißen metamere Farben.

Da die Farbenräume des Meßgeräts oder der Kamera und des menschlichen Auges sehr verschieden sein können, haben für das menschliche Auge ähnliche Farben nicht unbedingt ähnliche Farbkoordinaten im jeweils verwendeten technischen Koordinatensystem. Dies erschwert die Bestimmung der Ähnlichkeit von Farben mit technischen Mitteln.

2.1.3 Farbmodelle

RGB und rgb

Das RGB–Modell ist das Standardfarbmodell für die Bildaufnahme und die Bildwiedergabe. Daher wird es auch oft bei der Farbbildverarbeitung eingesetzt. Allerdings ist das RGB–Modell eher als ein Oberbegriff für alle Farb-

[6] auch Primärfarben genannt

modelle anzusehen, deren drei Grundfarben im roten, grünen bzw. blauen Spektralbereich liegen, denn je nach spektraler Empfindlichkeit der Sensoren bzw. spektraler Energieverteilung der Leuchtelemente (z.B. Leucht–Phosphore der Bildröhre) ergeben sich verschiedene Basen.
Im RGB–Modell läßt sich die Überlagerung zweier Farben als Addition ihrer Koordinaten (Vektoraddition) darstellen. Allerdings eignet sich der euklidische Abstand im RGB–Farbenraum nur bedingt zu einer dem menschlichen Empfinden angepaßten Definition der Ähnlichkeit von Farben.
Der Wertebereich für R, G und B ist in der Farbmetrik üblicherweise das abgeschlossene Intervall $[0, 1]$. In der Farbbildverarbeitung wird dagegen oft der Bereich der ganzen Zahlen von 0 bis 255 verwendet.

Das rgb–Modell ist eine auf dem RGB–Modell basierende intensitätsfreie Farbdarstellung.

$$\begin{aligned} r &= R/(R+G+B) \\ g &= G/(R+G+B) \\ b &= B/(R+G+B) \end{aligned} \tag{1}$$

Für $R+G+B=0$ (also $R=G=B=0$) gilt $r=g=b=0$.

Gegenfarbenmodell ($I_1 I_2 I_3$)

Eine Annäherung an die Informationsverarbeitung im menschlichen Sehsystem stellt das Gegenfarbenmodell dar. Dabei wird die Farbe in Form von drei Gegensätzen (Komplementärfarben) charakterisiert: schwarz vs. weiß, rötlich vs. grünlich und bläulich vs. gelblich. Gleichzeitig wird eine verringerte Korrelation zwischen den Kanälen erreicht.
I_1 ist der Rot–Grün–Kanal (auch RG), I_2 der Blau–Gelb–Kanal (auch BY) und I_3 der Schwarz–Weiß–Kanal (auch WK oder WB).

$$I_1 = R - G \tag{2}$$

$$I_2 = 2B - R - G \tag{3}$$

$$I_3 = \frac{R+G+B}{3} \tag{4}$$

Anschließend können die Werte der einzelnen Kanäle noch auf den beim RGB–Modell verwendeten Wertebereich transformiert[7] werden.

XYZ und CIELAB

Um eine von den spektralen Empfindlichkeiten der verwendeten Sensorklassen bzw. der spektralen Energieverteilung der Grundfarben unabhängige Farbdarstellung zu erhalten, wurde von der CIE[8] der XYZ–Farbenraum

[7] Verschiebung und Skalierung
[8] Commission Internationale de l'Eclairage

definiert. Er besitzt eine virtuelle Basis, d.h. seine Grundfarben sind physikalisch nicht realisierbar.
Der Hellbezugswert Y ist relativ zum Umgebungslicht bzw. zur Beleuchtung. Der Wert 100 bezeichnet diese Helligkeit. Bei unbunten Farben sind im XYZ–Farbenraum alle drei Koordinaten gleich. Auch im XYZ–Farbenraum wird die Farbmischung durch die Addition der Koordinatenvektoren dargestellt.
Der XYZ–Farbenraum ist mit dem RGB–Farbenraum durch eine (umkehrbare) Koordinatentransformation verknüpft. Die genaue Transformation hängt von der verwendeten Basis des RGB–Farbenraums ab.

Das CIELAB–Modell, auch als L*a*b*–Modell bezeichnet[9], wurde als „gleichabständiges“ oder (annähernd) uniformes Farbmodell entwickelt. Es behebt den Nachteil des XYZ–Farbenraums, daß der Abstand zweier Farborte im Farbenraum keinen Bezug zur visuellen Ähnlichkeit der Farben hat. Das CIELAB–Modell eignet sich daher sehr gut zur Definition von Farbabständen.
L* ist die Helligkeit, a* und b* bezeichnen die Grün–Rot– bzw. Blau–Gelb–Buntheit. Diese Gegenfarbendarstellung gleicht zwar prinzipiell dem oben dargestellten Gegenfarbenmodell, ist jedoch sehr eng an die Farbwahrnehmung des Menschen angelehnt.
Das CIELAB–Modell wird vorwiegend zur Beschreibung von Körperfarben (Nichtselbstleuchter) benutzt. Für die Beschreibung der Farbe von Selbstleuchtern wird meist das verwandte CIELUV–Modell eingesetzt.
Für die Umrechnung vom XYZ–Farbenraum in das CIELAB–Modell gelten folgende Gleichungen (nach [24]):

$$X* = \begin{cases} \sqrt[3]{\frac{X}{X_n}} & \text{für} \frac{X}{X_n} > 0.008856 \\ 7.787\frac{X}{X_n} + 0.138 & \text{für} \frac{X}{X_n} \leq 0.008856 \end{cases} \tag{5}$$

(entsprechend für $Y*$ und $Z*$)

$$L* = 116Y* - 16 \tag{6}$$

$$a* = 500(X* - Y*) \tag{7}$$

$$b* = 200(Y* - Z*) \tag{8}$$

(X_n, Y_n, Z_n) ist dabei der Unbuntpunkt, der einen vollkommen mattweißen Körper unter der beleuchtenden Lichtart repräsentiert.

HSI

Ein ebenfalls an das menschliche Farbempfinden angepaßtes Farbmodell ist das HSI–Modell. Die Buchstaben stehen für Buntton (hue), Sättigung (saturation) und Intensität (intensity). Es ist aber vom Ursprung her ein eher technisches Farbmodell.

[9] nicht zu verwechseln mit dem älteren Lab–Modell

Der Buntton B wird als Winkel zwischen 0° und 360° angegeben, wobei 0° Rot, 120° Grün und 240° Blau entsprechen. Die Sättigung S wird relativ zum intensitätsabhängigen, maximal möglichen Wert angegeben. In die Berechnung der Intensität I gehen bei der Umrechnung vom RGB–Modell R, G und B gleichgewichtet ein. Das HSI–Modell wurde auch in dieser Form hier verwendet. In der Literatur werden aber auch Transformationen für ungleich gewichtete Weißpunkte angegeben (siehe Hinweise in [24]).
Beispielhaft wird hier eine mögliche Umrechnung vom RGB– zum HSI–Modell[10] angegeben:

$$H = \begin{cases} \theta & \text{falls} B \leq G \\ 2\pi - \theta & \text{falls} B > G \end{cases} \tag{9}$$

$$\theta = \arccos \frac{((R-G)+(R-B))/2}{\sqrt{(R-G)^2+(R-B)\cdot(G-B)}} \tag{10}$$

$$S = 1 - \frac{\min(R,G,B)}{R+G+B} \tag{11}$$

$$I = \frac{R+G+B}{3} \tag{12}$$

Der Wertebereich für H ist 0° bis 360° und der für S und I das abgeschlossene Intervall $[0,1]$.
Geometrisch ist der HSI–Farbenraum ein Doppelkegel, auf dessen Umfang der Buntton abgetragen wird. Seine Achse ist die Unbuntachse, auf der die Intensität abgetragen wird. Die Sättigung ist die relative Entfernung des Farbortes von der Achse. Die Orte maximaler Sättigung liegen auf der Mantelfläche des Kegels.
Bei der Intensität 0 ist die Sättigung nicht definiert; bei undefinierter Sättigung und bei einer Sättigung von 0 ist der Buntton nicht definiert. Ist die Sättigung ungleich 0 jedoch sehr klein, so ist die Berechnung des Bunttons mit starken Fehlern behaftet.
Mit dem HSI–Modell verwandte Farbmodelle sind das HSV– und das HLS–Modell. Der Vorteil der letzten beiden Farbmodelle ist die einfache Umrechnung aus dem RGB–Modell. Sie kommt, im Gegensatz zum HSI–Modell, ohne trigonometrische Funktionen aus. Nachteil beider Farbmodelle ist, daß die Werte für V bzw. L nicht die Helligkeit der Farbe widerspiegeln, da sie auch vom Buntton abhängig sind.
Weitere Ausführungen zum HSI–Modell, seiner Ableitung und zu verwandten Farbmodellen finden sich in [24] und den dort genannten Quellen.

[10] Die Umkehrung wird in dieser Arbeit nicht benötigt.

2.2 Erweiterte Normallagen–Bilder (*Extended Surface Normal Images*, ESNI)

Normallagen–Bilder (*surface normal images*, SNI) sind eine fundamentale Grundlage des vorgestellten Objekterkennungsverfahrens. Sie werden hier kurz vorgestellt, und anschließend wird auf Erweiterungen zur Integration von Farbinformation eingegangen.
Ausführlichere Darstellungen zu SNI finden sich in [30] und [17], bzgl. der Erweiterungen siehe auch [4].

2.2.1 Normallagen–Bilder (*Surface Normal Images*)

J. Park, T. Chang und J. Choi diskutieren in [30] die Möglichkeiten, dreidimensionale Objekte in einem System geeignet zu repräsentieren. Bisherige Darstellungsvarianten haben einen hohen Verwaltungsaufwand oder erfordern viel Rechenleistung beim Matching. Bei Verwendung von 3D–Modellen sind für den Vergleich mit der 2D–Information des Kamerabildes jeweils perspektivische Repräsentationen zu berechnen. Charakteristische Ansichten (*characteristic views*) hingegen erfordern zwar keine Projektion beim Vergleich mit den Szenenmerkmalen, verursachen jedoch schon für etwas komplexere Objekte einen relativ hohen Aufwand, da viele Ansichten zu verwalten sind. Die Beschränkung auf charakteristische Ansichten führt auch zu dem Problem der Festlegung, welche Eigenschaften eine Ansicht als charakteristisch auszeichnen. Ein ausführlicher Vergleich verschiedener Repräsentationen findet sich in Kapitel 2 in [3].
Um diese Nachteile zu beseitigen, wird im oben genannten Artikel die Objektrepräsentation durch Normallagen–Bilder (*surface normal images*, SNIs) beschrieben. Dabei ist ein SNI die 2D–Ansicht (unter orthogonaler Parallelprojektion) eines 3D–Objekts, wobei eine Fläche, die Basisfläche, parallel zur Bildebene ausgerichtet und damit ohne perspektivische Verzerrung abgebildet wird. Diese spezielle, durch die Basisfläche definierte Ansicht wird als Normalansicht (*surface normal view*, SNV) bezeichnet. Die Oberflächennormale der Basisfläche ist parallel zur Blickrichtung. Jeder Fläche eines Objekts ist genau ein SNI zugeordnet, d.h. für jedes Objekt ist die Anzahl der SNIs gleich der Anzahl der Flächen.
Somit ist die Anzahl zu verwaltender Ansichten in der Regel deutlich geringer als bei der Verwendung charakteristischer Ansichten. Gleichzeitig werden nur genau definierte Ansichten, die Normalansichten (SNVs), miteinander verglichen. Nachteilig ist, daß, um ein Objekt in beliebiger Ansicht erkennen zu können, diese Ansicht erst in eine Normalansicht transformiert werden muß.
Die hierarchische Struktur, die durch die einzelnen Normalansichten eines Objekts und die einzelnen Flächen jeder Normalansicht definiert wird, heißt SNI–Struktur. Die folgende Grafik (Abb. 3) zeigt die SNI–Struktur eines

Objekts mit sieben Flächen:

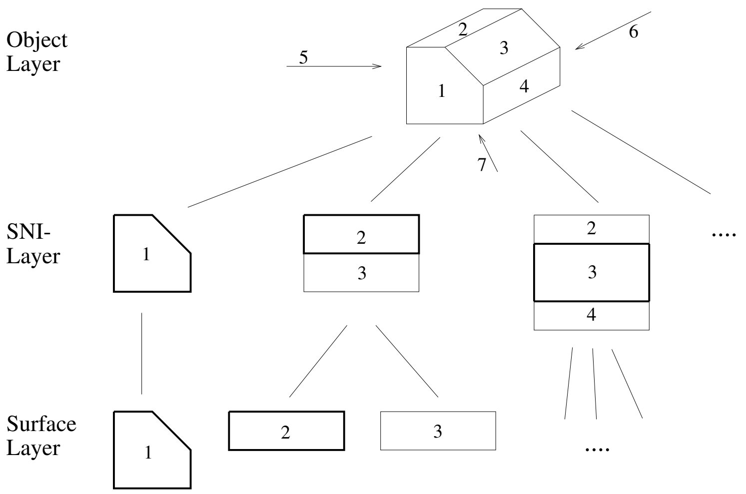

Abbildung 3: SNI–Struktur eines Objekts mit sieben Flächen (nach [17]). Dick umrandete Flächen sind die Basisfläche der jeweiligen Ansicht.

Mit dieser Struktur läßt sich eine leicht zu verwaltende Datenbank aufbauen, wobei in diese Baumstruktur weitere Daten integriert werden können (z.B. zu jeder Fläche deren Kantenanzahl, Kantenlängen, etc.), siehe auch Abschnitt 2.2.2.

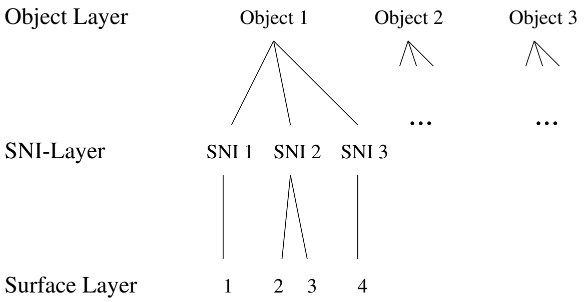

Abbildung 4: Baumstruktur der SNI–Daten (nach [17])

Um ein zu erkennendes Objekt mit der Datenbank zu vergleichen, werden die rekonstruierten 3D–Koordinaten[11] des zu erkennenden Objekts so gedreht, daß jeweils eine Fläche parallel zur Bildebene liegt. Solche Bilder werden *Rotated Input Images* (RIIs) genannt. Wenn sie — wie im vorgestellten Objekterkennungssystem — aus 2.5D–Daten erzeugt werden, fehlen die bei der gegebenen Ansicht von der Kamera abgewandten Flächen des Objekts,

[11]Für die Rotation reicht die Bestimmung der Orientierungen der rekonstruierten Flächen.

so daß die RIIs im allgemeinen nicht topologisch gleich zum entsprechenden SNI sind. Die RIIs enthalten dann unter Umständen weniger Flächen als die entsprechenden SNIs.
Aufgrund von Ungenauigkeiten bei der Bestimmung der Winkel zwischen den Flächen kann es auch vorkommen, daß Flächen, die im RII vorhanden sein müßten, nicht vorhanden sind oder umgekehrt. Dies ist in der Regel bei Flächen mit Winkeln in der Umgebung von 90° zur Basisfläche der Fall. Wird z.B. durch Fehler der Winkel mit 89° statt 90° bestimmt, so wird die Fläche im RII fälschlicherweise auftreten.

Abbildung 5: Erzeugung eines RII (nach [17])

Die für die Erzeugung der RIIs notwendige 2.5D– oder 3D–Rekonstruktion begrenzt die Anwendbarkeit des Verfahrens. Sie hat einen entscheidenden Anteil am Rechenzeitbedarf des Gesamtsystems und bedingt, daß die Objekterkennung i.a. nicht mit nur einem Kamerabild des zu erkennenden Objekts durchgeführt werden kann.

Gekrümmte Flächen

Gekrümmte Flächen werden durch planare Flächenstücke approximiert, d.h. Oberflächenpunkte, bei denen die Oberflächennormale in einem bestimmten Toleranzbereich liegt, werden zu einer planaren Fläche zusammengefaßt. Je kleiner die Toleranz gewählt wird, desto mehr planare Flächen erhält man und desto genauer wird die Krümmung modelliert aber umso komplexer gestaltet sich die Verwaltung.
Die folgende Abbildung zeigt die Einteilung der Oberfläche einer Halbkugel (im Schnitt):

Abbildung 6: Approximation gekrümmter Flächen (nach [17])

Wichtig bei dieser Repräsentation gekrümmter Flächen ist die Unterscheidung der Kanten in reale Kanten und Approximationskanten, da letztere keine Entsprechung auf dem realen Objekt haben. Bei der Verwendung gekrümmter Flächen bei der Objekterkennung ist zu beachten, daß im Gegensatz zu planaren Flächen, gekrümmte Flächen auch teilweise sichtbar sein können. Die SNI–Repräsentation ist daher für gekrümmte Flächen weniger geeignet.

2.2.2 Erweiterungen

G. Bellaire, K. Oppermann und W.J. Schimke haben eine Erweiterung der SNI–Struktur zur Repräsentation von Farbinformation entwickelt [29, 4, 3]. Dabei wird jedes SNI mit einem Farbbild und Texturinformation der Basisfläche verknüpft (erweiterte Normallagen–Bilder, *Extended* SNI, ESNI).
Falls nur das Farbbild und/oder die Texturinformation der Basisfläche (eventuell zusammen mit geometrischen und topologischen Merkmalen der Basisfläche, die ebenfalls unabhängig von den Nachbarflächen bestimmbar sind) für die Objekterkennung herangezogen wird, so entfällt gleichzeitig das Problem der im allgemeinen nicht vorhandenen topologischen Ähnlichkeit zwischen RII und SNI (aufgrund fehlender Flächen im RII). Die Basisfläche im RII ist der Basisfläche im entsprechenden SNI mathematisch ähnlich und unterscheidet sich von dieser nur durch Skalierung, Translation und Rotation um die optische Achse.
Die oben beschriebene Approximation einer gekrümmten Fläche durch planare Flächen führt zu einem Synchronisationsproblem: Wie kann sichergestellt werden, daß die Grenzen der planaren Flächenstücke im RII auf demselben Ort auf der Objektoberfläche liegen wie im SNI? Da die gekrümmte Fläche nicht unbedingt vollständig sichtbar ist[12], ist eine Orientierung an ihren Kanten nicht möglich. Es ist daher ein Initialpunkt zu definieren, an dem die Approximation durch planare Flächenstücke ausgerichtet wird.

Die für das vorgestellte Objekterkennungssystem erstellte Datenbank ist in Anlehnung an die erweiterte SNI–Struktur (ESNI–Struktur) aufgebaut, wobei jeweils nur die Information über die Basisfläche jeder Ansicht gespeichert wird. Siehe dazu auch die Beschreibung in Anhang C.

[12] Für die Mantelfläche eines Zylinders z.B. existiert keine Ansicht, in der sie vollständig sichtbar ist.

3 Automatisches Scannen der Oberflächentextur von 3D–Objekten

Im folgenden wird ein Verfahren vorgestellt, das zum automatischen Erfassen von Oberflächentexturen und Aufdrucken auf 3D–Objekten dient. Voraussetzung dafür ist die Bereitstellung eines CAD–Datensatzes des Objekts, in dem die Koordinaten der Eckpunkte, die Kanten und die Flächen des Objekts eingetragen sind[13]. Unter Verwendung dieser Daten kann dann das Objekt computergesteuert so gedreht werden (Abschnitt 3.1), daß mit einer Kamera von jeder Fläche ein Bild erstellt werden kann (Abschnitt 3.2). Anschließend ist dann in diesem Bild die gewünschte Fläche zu extrahieren (Abschnitt 3.3).
Das Verfahren eignet sich nur für annähernd polyedrische Objekte, d.h. Objekte mit ausschließlich planaren Oberflächenstücken oder Oberflächenstücken geringer Krümmung. Notwendige Erweiterungen und Anpassungen an Objekte mit stark gekrümmten Flächen werden diskutiert.

3.1 Rotation der Objekte

3.1.1 Aufbau der Pan–Tilt–Unit (PTU)

Die Pan–Tilt–Unit (PTU) dient zur computergesteuerten Rotation von Objekten, Kameras etc. Sie hat zwei von Schrittmotoren bewegte Drehachsen. Die Pan–Achse ist parallel zur Z–Achse. Die Tilt–Achse rotiert um die Pan–Achse und ist in der Nullstellung der Pan–Achse parallel zur X–Achse (siehe Abb. 7 auf der nächsten Seite). Sofern die Schwenkbereiche der beiden Achsen keinen Einschränkungen unterliegen, ist damit für ein mit der PTU verbundenes Objekt bzgl. Rotation jede Lage im Raum erzielbar[14].
Auf der PTU wurde eine mit schwarzem Papier belegte runde Sperrholzplatte befestigt, auf die die aufzunehmenden Objekte gestellt werden (siehe Abb. 8 auf Seite 17). Eine Fixierung der Objekte auf der Platte ist nur für größere Tilt–Winkel erforderlich. Die Beschleunigung der PTU beim Beginn und am Ende der Bewegung führt zu keiner Lageveränderung der Objekte bzgl. der PTU.

3.1.2 Aufnahme der Objektflächen

Die Abbildung 8 auf Seite 17 zeigt die Anordnung der Kamera und der PTU. Eingezeichnet sind auch das Objektkoordinatensystem (X,Y,Z) und

[13] Für gekrümmte Flächen ist eine Approximation durch planare Flächenstücke zu wählen, siehe auch Ende von Abschnitt 3.4.2 auf Seite 32 sowie Abschnitt 2.2.1 auf Seite 13.

[14] Der dritte Rotationsfreiheitsgrad entspricht einer Rotation um die Y–Achse der PTU, also der optischen Achse. Diese Rotation entspricht einer 2D–Rotation im Bild und braucht daher nicht weiter betrachtet zu werden.

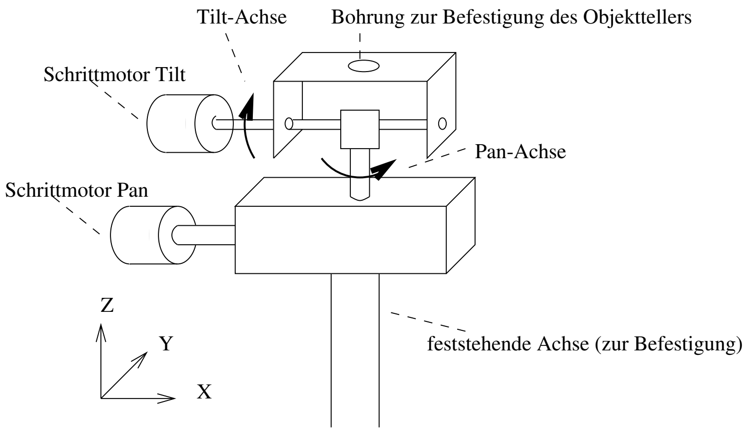

Abbildung 7: Schematischer Aufbau der PTU, in Nullstellung

das Bildkoordinatensystem (x,y). Zu beachten ist dabei, daß die optische Achse parallel zur Y–Achse ist.

Definition der Lage des Objekts auf der PTU

Um die Lage des Objekts zu beschreiben, genügt es, die Fläche, mit der das Objekt auf der Unterlage aufliegt (Grund- oder Auflagefläche), sowie eine Fläche, die parallel zur Bildebene ist, anzugeben. Für die zur Bildebene parallele Fläche muß jedoch noch angegeben werden, ob sie der Kamera zu- oder abgewandt ist.

Allerdings ist eine solche Lagebeschreibung eine recht große Einschränkung für die Objekte, denn die Normale der Grundfläche muß parallel zur Z–Achse sein und die Normale der anderen Fläche muß parallel zur optischen Achse sein, wodurch sie einen durch die Aufnahmegeometrie fest vorgegebenen Winkel einschließen müssen. Bei der verwendeten Aufnahmegeometrie (siehe Abb. 8 auf der nächsten Seite) beträgt der Winkel 90°.

Um nicht für einige Objekte, die keine 90°–Kanten aufweisen (wie z.B. die 1l–Eispackungen von Mövenpick und ähnliche Verpackungen), die Aufnahmegeometrie ändern zu müssen, ist es erforderlich, einen beliebigen Winkel zwischen der Grundfläche und einer zur Kamera gerichteten Fläche zuzulassen. Von den möglichen drei Basisrotationen der betrachteten Objektfläche zur Bildebene ist dabei nur eine Rotation um die X–Achse, im Prinzip ein Tilt–Offset, zu betrachten. Eine Rotation um die Z–Achse kann durch Drehung des Objekts auf der PTU ausgeglichen werden, und eine Rotation um die Y–Achse hat keinen Einfluß auf den Winkel zwischen optischer Achse und Normale der betrachteten Fläche.

Unter Voraussetzung der verwendeten Aufnahmegeometrie ist zur Bestimmung der Normalen (im Modellkoordinatensystem) einer zur Bildebene parallelen Ebene die Normale der betrachteten Fläche zunächst mit der durch

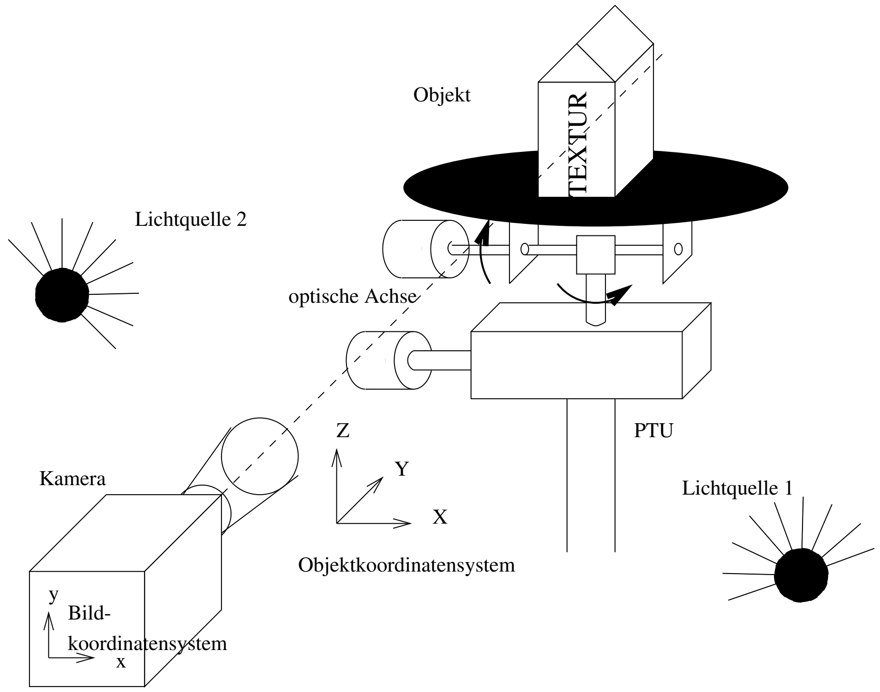

Abbildung 8: Aufnahmekonfiguration für das 3D–Texture–Scanning

die Abbildung der Normalen der Grundfläche auf die –Z–Achse[15] definierten Rotation zu transformieren. Durch Nullsetzen der Z–Komponente dieser transformierten Normalen und anschließendes Normieren erhält man dann die Normale einer zur Bildebene parallelen Fläche. Die Umrechnung in das Modellkoordinatensystem erfolgt mittels der Inversen der oben definierten Transformation. Zur Begründung: Nach der Anwendung der genannten Transformation liegen die Normalen aller möglicherweise zur Bildebene parallelen Flächen in der XY–Ebene, da sie bei der vorausgesetzten Aufnahmegeometrie senkrecht zur Grundfläche sind, also nach der Transformation senkrecht zur Z–Achse sind. Das Nullsetzen der Z–Komponente entspricht einer Projektion auf die XY–Ebene. Dieser Anteil des Normalenvektors bestimmt eindeutig die Rotation um die Z–Achse, die (zusammen mit der durch die Grundfläche definierten Transformation) die Objektkoordinatensysteme des Modells und der PTU in Übereinstimmung bringt. Die anderen beiden Rotationsfreiheitsgrade sind durch die Grundfläche festgelegt.
Eine weitere Überlegung betrifft die Grundfläche. Falls das Objekt in einer Lage aufgenommen werden soll, in der es einen stabilen Stand hat, aber nicht auf einer seiner Flächen aufliegt, so sollte die Angabe von drei nicht kollinearen Auflagepunkten zur Definition der nach „unten“ gerichteten Normalen im Objektkoordinatensystem des 3D–Modells genügen. Da auch die

[15] die der Z–Achse entgegengerichtete Achse

Normale einer Fläche aus drei ihrer Punkte (nicht kollinear) berechnet wird, ergeben sich damit aber keine Änderungen im Rechenablauf.
Es ist wichtig, daß eine genaue Ausrichtung des Objekts anhand der definierten Lage auf der PTU erfolgt, da hierbei entstehende Fehler sich auf alle Aufnahmen auswirken.

Berechnung der Koordinatentransformation

Es ist die Rotationsmatrix für die Abbildung des Objektkoordinatensystems des 3D–Modells des Objekts (CAD–Datensatz) auf das Objektkoordinatensystem der PTU zu bestimmen. Diese Rotation setzt sich aus zwei Teilrotationen zusammen; die erste dient zur Abbildung der Normalen der Grundfläche $\overrightarrow{n_G}$ auf den Vektor $-\overrightarrow{Z} = (0,0,-1)^{\mathrm{T}}$ und die zweite zur Abbildung des Normalenvektors in Richtung der Kamera $\overrightarrow{n_V}$ auf den Vektor $-\overrightarrow{Y} = (0,-1,0)^{\mathrm{T}}$. Bei der gegebenen Aufnahmegeometrie stehen die Vektoren $\overrightarrow{n_G}$ und $\overrightarrow{n_V}$ senkrecht aufeinander.
Die Achse der ersten Rotation ist das Kreuzprodukt der Vektoren $\overrightarrow{n_G}$ und $-\overrightarrow{Z}$, der Rotationswinkel ist der von diesen Vektoren eingeschlossene Winkel. Die Drehrichtung ist in der Orientierung der Achse enthalten, so daß nur um einen nicht–negativen Winkel gedreht wird.
Die so erhaltene Rotation wird auf den Vektor $\overrightarrow{n_V}$ angewandt, und man erhält den Vektor $\overrightarrow{n_{V\mathrm{r}}}$. Die zweite Rotation erfolgt um die durch das Kreuzprodukt der Vektoren $\overrightarrow{n_{V\mathrm{r}}}$ und $-\overrightarrow{Y}$ definierte Achse mit dem von diesen Vektoren eingeschlossenen Winkel. Da sowohl der Vektor $\overrightarrow{n_{V\mathrm{r}}}$ als auch der Vektor $-\overrightarrow{Y}$ in der XY–Ebene liegen[16], ist die Rotationsachse die Z– oder die –Z–Achse[17], so daß die Rotation auch um die Z–Achse mit dem orientierten Winkel zwischen den beiden Vektoren erfolgen kann.

Bestimmung von Pan und Tilt

Um eine beliebige Fläche (außer der Grundfläche) aufzunehmen, ist das Objekt so zu drehen, daß die aufzunehmende Fläche parallel zur Bildebene und der Kamera zugewandt ist. D.h. also, daß eine Drehung um die Pan– und Tilt–Achse der PTU zu finden ist, so daß der rotierte (Rotationen R_G und R_V) Normalenvektor $\overrightarrow{n_{\mathrm{R}}}$ der Fläche auf den Vektor $-\overrightarrow{Y}$ im Objektkoordinatensystem abgebildet wird. Eine solche Drehung existiert immer und ist eindeutig definiert.

[16] Es gilt $\overrightarrow{n_G} \perp \overrightarrow{n_V}$. Somit gilt nach der ersten Teilrotation R_G: $-\overrightarrow{Z} \perp \overrightarrow{n_{V\mathrm{r}}}$, d.h. der Vektor $\overrightarrow{n_{V\mathrm{r}}}$ liegt in der XY–Ebene. Der Vektor $-\overrightarrow{Y}$ liegt trivialerweise ebenfalls in der XY–Ebene.

[17] die der Z–Achse entgegengerichtete Achse

Seien

$$\overrightarrow{n_{\mathrm{R}}} = \begin{pmatrix} x \\ y \\ z \end{pmatrix} \text{ und} \tag{13}$$

$$\overrightarrow{t} = \begin{pmatrix} 0 \\ y \\ z \end{pmatrix}, \tag{14}$$

dann ist der Tilt der orientierte Winkel zwischen dem Vektor $\overrightarrow{t}$ und der −Y–Achse (Vektor $-\overrightarrow{Y}$). Die Rotationsachse ist dabei die X–Achse, da die Tilt–Rotation R_T in Nullstellung der Pan–Achse ausgeführt wird.
Sei $\overrightarrow{n_{\mathrm{RT}}}$ der Vektor $\overrightarrow{n_{\mathrm{R}}}$ nach Anwendung der Rotation R_T und

$$\overrightarrow{n_{\mathrm{RT}}} = \begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} \text{ und} \tag{15}$$

$$\overrightarrow{p} = \begin{pmatrix} x' \\ y' \\ 0 \end{pmatrix}, \tag{16}$$

dann ist der Pan–Winkel der orientierte Winkel zwischen dem Vektor $\overrightarrow{p}$ und der optischen Achse, also der −Y–Achse. Die zugehörige Rotation (Pan–Rotation) wird mit R_P bezeichnet.
Abschließend erfolgt eine Normalisierung der so erhaltenen Winkel: Ist der Tilt größer als 90° oder kleiner als −90° so werden 180° subtrahiert bzw. addiert und der Pan–Winkel entsprechend angepaßt (Addition von 180°). Danach muß der Pan–Winkel wieder in den Bereich von −180° bis +180° gebracht werden.

Generierung der synthetischen Linienzeichnung der Ansicht

Auf die Objektkoordinaten im 3D–Koordinatensystem des CAD–Datensatzes (pol–Datei) sind die Rotationen R_G, R_V, R_T und R_P anzuwenden. Anschließend erfolgt eine Projektion auf die XZ–Ebene des Objektkoordinatensystems der PTU, die parallel zur xy–Ebene des Bildkoordinatensystems ist.
Es wird Parallelprojektion angenommen. Dies kann hier, soweit das Objekt exakt ausgerichtet wurde, ohne Fehler getan werden, da sich die aufzunehmende Fläche in einer Ebene konstanter Tiefe befindet und für diesen Fall Parallel- und Zentralprojektion sich nur durch eine Skalierung der Bilder unterscheiden. Falls bei späteren Betrachtungen die sichtbaren angrenzenden Flächen der aufzunehmenden Fläche einbezogen werden, so ist der durch die Annahme von Parallelprojektion entstandene Fehler gering, wenn die Tiefenänderung klein gegenüber dem Abstand zur Kamera ist.

In die Linienzeichnung werden nur die sichtbaren Flächen der Ansicht übernommen. Der Normalenvektor einer sichtbaren Fläche bildet mit dem Normalenvektor der aufzunehmenden Fläche einen Winkel von weniger als 90°.

3.2 Bildaufnahme

Abb. 8 auf Seite 17 und Abb. 9 zeigen schematisch die Aufnahmegeometrie.

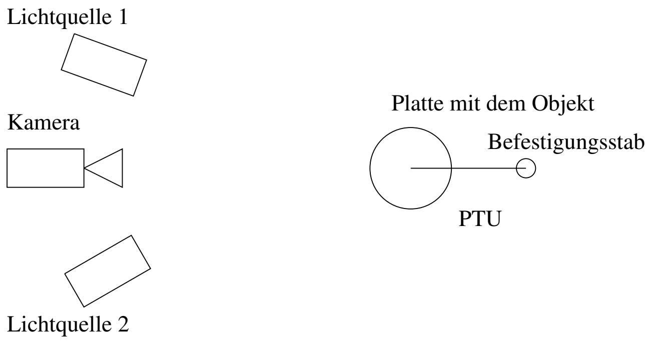

Abbildung 9: Verwendete Aufnahmegeometrie

Der Abstand der Objekte von der Kamera betrug ca. 135 cm. Die Kamera wurde 10 cm höher als die Platte der PTU angeordnet. Sie ist weder zur Horizontalen noch zur Vertikalen geneigt bzw. gedreht. Das Objekt wurde mit zwei Diaprojektoren mit Niedervolthalogenlampen beleuchtet. Die erste Lichtquelle war um 20° gedreht und hatte eine Neigung zur Horizontalen von 15° nach oben, die zweite war um 30° gedreht, mit einer Neigung von 10° nach unten.

Diese Beleuchtung erwies sich als hinreichend gut zur Vermeidung von Glanzlichtern[18]. Sie führt jedoch dazu, daß nur die zur Bildebene parallelen Flächen gleichmäßig und gut ausgeleuchtet sind. Flächen mit großer Neigung zur Bildebene erscheinen relativ dunkel. Dies ist bei der Segmentierung der Aufnahmen zu beachten.

Aus diesem Grund wurde jede Aufnahme mit einer Verstärkung von 9dB wiederholt, um eine gute Trennung von Hintergrund und Objekt zu ermöglichen. Die bei einer Verstärkung des Bildsignals eintretende Verschlechterung des Signal–Rausch–Verhältnisses ist für die Hintergrundsegmentierung unproblematisch. Alternativ könnte die zweite Aufnahme auch mit weiter geöffneter Blende erfolgen, allerdings besteht bei der verwendeten Kamera keine Möglichkeit zur automatischen Einstellung definierter Blendenwerte.

[18]Bei der Annahme von Lambert–Reflexion (perfekt diffus) ist die Beleuchtungsrichtung im Prinzip nahezu beliebig, da die einfallende Intensität auf zur Bildebene parallelen Flächen immer konstant ist und somit die gemessene Intensität nur von der lokalen Albedo abhängt. Leider haben die untersuchten Objekte jedoch einen gewissen Anteil an spiegelnder Reflexion.

Einer weiter geöffneten Blende standen auch die dann stärker in Erscheinung tretenden Reflexionen sehr heller Objekte auf dem schwarzen Papier des Untergrunds entgegen.
Wichtige zu betrachtende Punkte sind der Abstand des Objekts zur Kamera und die Schärfe des Bildes. Durch die Drehung der Objekte auf der PTU ändert sich bei einigen Objekten bei bestimmten Lagen der Abstand der aufzunehmenden Fläche zur Kamera. Ist der mittlere Abstand zur Kamera groß bzw. die Brennweite des Objektivs klein, ist der Einfluß dieser Tiefenänderung zwar klein, aber es lassen sich keine bildfüllenden Aufnahmen der Objekte machen, so daß die erzielbare Auflösung für die aufzunehmende Fläche gering ist. Daher ist ein relativ kleiner Abstand zur Kamera und eine große Brennweite zu wählen.
Steht keine Autofocus–Kamera zur Verfügung, so kann durch Einstellung einer möglichst kleinen Blendenöffnung eine große Tiefenschärfe erzielt werden, die kleine Schwankungen der Tiefe bei der Drehung des Objekts ausgleicht.

3.3 Segmentierung

Aufgabe der Segmentierung ist die Extraktion der Textur der Basisfläche der jeweiligen Ansicht. Dazu ist zunächst das Objekt vom Hintergrund zu separieren. Da die Aufnahmen unter Laborbedingungen erfolgen, wurde der Hintergrund komplett schwarz ausgekleidet, so daß ein rein farbbasiertes Verfahren angewandt werden kann. Anschließend ist aus der gesamten Ansicht des Objekts die Basisfläche zu extrahieren. Da es sich um eine genau definierte Ansicht handelt, kann dem Abbild des Objekts die Linienzeichnung der Ansicht überlagert werden (siehe Abbildungen 10 auf der nächsten Seite und 11 auf Seite 26).

Ablauf der Segmentierung in Stichworten:

```
schwarzen Hintergrund segmentieren (Region Growing)
Objektbereich (Segment, das das Objekt repräsentiert)
    finden und markieren
Löcher im Hintergrund schließen
Objektbereich vermessen
synthetische Linienzeichnung überlagern
Basisfläche extrahieren
```

Hintergrundsegmentierung

Wie oben dargestellt, kann der Hintergrund rein farbbasiert segmentiert werden. Dazu wird ein Region Growing verwendet, das den Hintergrund vom Ursprungspunkt aus beginnend markiert. Es wird also a priori angenommen, daß der Punkt $(0, 0)$ zum Hintergrund gehört. Man könnte statt

Abbildung 10: Aufnahme des Objekts `Reichelt_Schlagsahne_0.5l`

des Region Growings auch für jeden Pixel einzeln anhand seiner Farbe entscheiden, ob er zum Hintergrund gehört. Das führt jedoch dazu, daß auch alle schwarzen Bereiche der Objekttextur als Hintergrund klassifiziert werden. Bei Verwendung von Region Growing hingegen werden nur diejenigen schwarzen Gebiete der Textur fälschlicherweise als Hintergrund klassifiziert, die mit dem Hintergrund zusammenhängend sind. (Das Region Growing verwendet die für den Hintergrund übliche 4–Nachbarschaft.)
Ob beim Region Growing ein Pixel zum Hintergrund gehört, wird über die Intensitäten der drei Farbkanäle definiert. Liegen die Werte sowohl im Rot– als auch im Grün– und Blau–Kanal unter einem bestimmten Wert, so gilt der Punkt als schwarz und somit dem Hintergrund zugehörig. Es ist auch eine adaptive Berechnung des Schwellenwertes denkbar, da aufgrund der bereits erwähnten Reflexionen auf dem schwarzen Papier bei einigen sehr hellen Objekten auch größere Hintergrundintensitäten zu messen sind. Allerdings ist dies mit einem hohen Rechenaufwand verbunden, da nur ein iteratives Verfahren erfolgversprechend ist. Daher wurde für die Hintergrundsegmentierung dunkler Objekte eine um 9dB verstärkte Aufnahme verwendet. In der Praxis hat sich ein Schwellenwert von 30 bei den unverstärkten und von 80 bei den verstärkten Aufnahmen als günstig erwiesen, wobei bei der verwendeten Kamera und den verwendeten Einstellungen die Intensität in allen Kanälen den Wert 17 nicht unterschreitet. Bei einem etwas zu niedrigen Schwellenwert, so daß noch vereinzelte hellere Hintergrundpunkte nicht als solche klassifiziert werden, steigt der Rechenzeitbedarf des Verfahrens stark, da das verwendete Region Growing bei den dann auftretenden stark zerklüfteten Flächen sehr ineffizient wird.
Ein weiteres Kriterium, daß sich die Intensitäten der einzelnen Kanäle um nicht mehr als einen bestimmten Wert (z.B. 10) unterscheiden dürfen, hat

sich als unnötig erwiesen. Aufgrund des hohen zusätzlichen Rechenzeitaufwands kam dieses Kriterium nicht zum Einsatz. Es könnte zur Unterscheidung dunkler Farben von „echtem Schwarz“ dienen.
Da der Hintergrund nicht vollständig schwarz ist (z.B. Metallteile der PTU, deren Verkleidung recht aufwendig wäre), weist der per Region Growing erhaltene Hintergrund nicht nur an der Stelle des Objekts Löcher auf. Es ist nun zu entscheiden, welches der Löcher den Ort des Objekts repräsentiert. Dazu wird zuerst für den Bildmittelpunkt $(x/2, y/2)$, wobei x die Bildbreite und y die Bildhöhe ist, geprüft, ob er im schon klassifizierten Hintergrund liegt. Falls nein, wird angenommen, daß er zum Objekt gehört. Anderenfalls wird der Punkt $(x/2, 3y/4)$ untersucht und, falls sich auch dort Hintergrund befindet, der Schwerpunkt der noch nicht klassifizierten Punkte.
Man kann auch sofort mit dem Schwerpunkt als Kandidaten beginnen. Die obige Variante ist jedoch günstiger, da so oft der Aufwand für die Berechnung des Schwerpunktes entfällt. In der Regel sind die Objekte bildfüllend aufgenommen, so daß der Bildmittelpunkt auf das Objekt trifft. Nur bei liegenden Objekten ist das meist nicht der Fall, weshalb der zweite Kandidaten–Punkt geprüft wird. In der Praxis haben sich keine Fehler aufgrund der vorrangigen Überprüfung der beiden festen Punkte gezeigt, da oberhalb der Objekte keine hellen Stellen im Hintergrund auftraten.
Der so gefundene Objektbereich wird dann mittels Region Growing markiert. Alle bis dahin noch nicht klassifizierten Pixel gelten als Hintergrund. Dies sollten nur nicht–schwarze Teile des Hintergrunds sein.
Eine perfekte Verkleidung des Hintergrunds, so daß alle helleren Stellen zum Objekt gehören, ergäbe einen einfacheren Ablauf der Hintergrundsegmentierung. Aus Gründen der Robustheit, sollte jedoch trotzdem von mehreren Löchern im Hintergrund nach dem ersten Schritt ausgegangen werden.

Überlagerung von Bild und synthetischer Linienzeichnung

Die Extraktion der Basisfläche aus dem gesamten Bereich des Objekts erfolgt durch Überlagerung des Gebietes des Bildes, in dem das Objekt vermutet wird, mit der aus dem entsprechenden CAD–Datensatz gewonnenen synthetischen Linienzeichnung dieser Ansicht. Dazu ist eine Abbildung zu finden, mit der die Linienzeichnung paßgenau mit dem Bild des Objekts in Übereinstimmung gebracht wird. Die Koordinaten der Eckpunkte der Basisfläche im Bild sind dann die transformierten Koordinaten der Basisfläche in der synthetischen Linienzeichnung.
Die Bestimmung der Abbildung zur Überlagerung der Linienzeichnung erfolgt anhand eines bestimmten Punktpaares. Dies können entweder die linken und rechten oberen oder die linken und rechten unteren Extrempunkte des Objektbereichs bzw. der Linienzeichnung sein. Extrempunkt ist dabei derjenige Punkt mit dem geringsten euklidischen Abstand zur entsprechenden „Ecke“, wobei sich die Koordinaten der „Ecke“ aus minimal bzw.

maximal vorkommender x–Koordinate und minimal bzw. maximal vorkommender y–Koordinate ergeben.
Da sich jedoch aufgrund teilweise stark abgerundeter Ecken der Objekte die Extrempunkte des Objektbereichs meist nur unzureichend zum Überlagern eignen, wurde ein Verfahren entwickelt, das die gesamte untere bzw. obere Begrenzungstrecke des Objektbereichs zur Bestimmung des Punktpaares verwendet. An der Bestimmung der Extrempunkte in der Linienzeichnung wird dabei festgehalten, da sie fehlerfrei erfolgt.
Es werden alle Punkte am unteren bzw. oberen Rand des Objektbereichs betrachtet, deren y–Koordinate um maximal 10 kleiner[19] (bzw. größer) ist als die größere (kleinere) der y–Koordinaten der beiden Extrempunkte. Alternativ kann für Objekte, bei denen die y–Koordinaten der Extrempunkte sich stark von der minimal (bzw. maximal) vorkommenden y–Koordinate unterscheiden, die minimal (bzw. maximal) vorkommende y–Koordinate herangezogen werden. Mittels linearer Regression wird aus der so erhaltenen Punktmenge eine Geradengleichung bestimmt. Die x–Koordinaten der Extrempunkte werden dann in die Geradengleichung eingesetzt, und die so erhaltenen y–Koordinaten werden zusammen mit den x–Koordinaten der Extrempunkte zur Bestimmung der Abbildung verwendet. Bei ungeeigneter Lage der Extrempunkte kann auf ihre Verwendung ganz verzichtet werden. Stattdessen werden dann die minimalen und maximalen x–Koordinaten der oben beschriebenen Punktmenge verwendet. Dabei ist jedoch der obige Schwellenwert von 10 auf 3 zu senken.
Die Wahl, ob die Überlagerung anhand des oberen oder des unteren „Randes“ erfolgen soll, ist ein Parameter des Verfahrens. Je nach Objekt und dessen Lage können beide Punktpaare die genaueren Ergebnisse liefern.

Seien nun (x_1, y_1) und (x_2, y_2) die Koordinaten der beiden Punkte in der Linienzeichnung sowie (x'_1, y'_1) und (x'_2, y'_2) die Koordinaten der korrespondierenden Punkte im Bild. Dann ist die Abbildung der Punkte (x_1, y_1) auf (x'_1, y'_1) sowie (x_2, y_2) auf (x'_2, y'_2) zu bestimmen.
In homogenen Koordinaten ausgedrückt:

$$\begin{pmatrix} x_1 \\ y_1 \\ 1 \end{pmatrix} \rightarrow \begin{pmatrix} x'_1 \\ y'_1 \\ 1 \end{pmatrix} \qquad \text{und} \qquad \begin{pmatrix} x_2 \\ y_2 \\ 1 \end{pmatrix} \rightarrow \begin{pmatrix} x'_2 \\ y'_2 \\ 1 \end{pmatrix}$$

mit der Abbildungsmatrix

$$M = \begin{pmatrix} a & b & e \\ c & d & f \\ 0 & 0 & 1 \end{pmatrix}$$

[19]Der Ursprungspunkt des Koordinatensystems ist links oben, siehe Anhang A.1.

Da als Abbildungen nur Rotation, Translation und Skalierung mit Beibehaltung der Seitenlängenverhältnisse (d.h. Beibehaltung des *aspect ratio*) gestattet sind, folgt: $b = -c$ (keine Spiegelung oder Scherung) und $a = d$ (Erhaltung des *aspect ratio*).
Somit ergibt sich folgendes Gleichungssystem:

$$ax_1 + by_1 + e = x'_1 \qquad ax_2 + by_2 + e = x'_2 \tag{17}$$

$$-bx_1 + ay_1 + f = y'_1 \qquad -bx_2 + ay_2 + f = y'_2 \tag{18}$$

Durch Subtraktion der Gleichungen (17.1) und (17.2) sowie (18.1) und (18.2) erhält man

$$a(x_1 - x_2) + b(y_1 - y_2) = x'_1 - x'_2 \tag{19}$$

$$-b(x_1 - x_2) + a(y_1 - y_2) = y'_1 - y'_2 \tag{20}$$

Durch Umstellen der Gleichung (19) nach a

$$a = \frac{x'_1 - x'_2 - b(y_1 - y_2)}{x_1 - x_2} \tag{21}$$

und Einsetzen in Gleichung (20) sowie Umstellen nach b

$$-b(x_1 - x_2) + \frac{x'_1 - x'_2 - b(y_1 - y_2)}{x_1 - x_2}(y_1 - y_2) = y'_1 - y'_2$$

$$-b\frac{(x_1 - x_2)^2 + (y_1 - y_2)^2}{x_1 - x_2} = y'_1 - y'_2 - \frac{x'_1 - x'_2}{x_1 - x_2}(y_1 - y_2)$$

$$b = -\frac{(y'_1 - y'_2)(x_1 - x_2) - (x'_1 - x'_2)(y_1 - y_2)}{(x_1 - x_2)^2 + (y_1 - y_2)^2} \tag{22}$$

erhält man Gleichungen zur Bestimmung von a und b. Gleichungen für e und f ergeben sich durch Umstellen der Gleichungen (17.1) und (18.1)

$$: e = x'_1 - ax_1 - by_1 \tag{23}$$

$$f = y'_1 + bx_1 - ay_1 \tag{24}$$

Mit den Gleichungen (21), (22), (23) und (24) ist die Matrix M bestimmt.

Ablauf der Überlagerung in Stichpunkten:

```
Bestimmung der in Übereinstimmung zu bringenden Punktpaare
Berechnung von b, a, e und f
Aufstellen der Transformationsmatrix
für jeden Punkt der Basisfläche in der Linienzeichnung
  Transformation der Koordinaten der Linienzeichnung in
      Koordinaten des Bildes
  Eintragen in den Clippath
```

Am Ende der Überlagerung mit der Linienzeichnung steht das Begrenzungspolygon der Basisfläche fest, und die Segmentierung ist damit abgeschlossen.

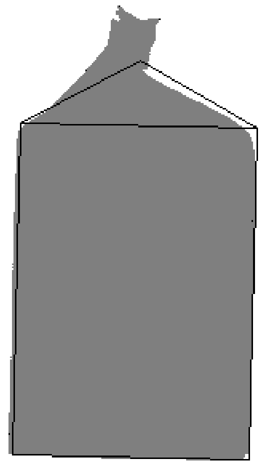

Abbildung 11: Überlagerung der synthetischen Linienzeichnung für die Aufnahme aus Abbildung 10

Zur Überprüfung des Ergebnisses kann noch untersucht werden, inwieweit sich als Hintergrund erkannte Bildteile innerhalb des Begrenzungspolygons befinden. Gegebenenfalls könnte die Abbildungsmatrix zur Überlagerung der Linienzeichnung korrigiert und so iterativ die optimale Segmentierung erreicht werden.
Zur Verringerung des Speicherplatzbedarfs wird abschließend das Bild der Basisfläche auf das kleinste das Begrenzungspolygon (Clippath) vollständig enthaltende Rechteck, das parallel zum Bildrand ist, begrenzt.

3.4 Ergebnisse

Das Verfahren wurde an 27 Objekten getestet (siehe Tabelle 1). Die Objekte wurden manuell vermessen und mit Hilfe eines CAD–Programms (Polygraph) modelliert. Die Objekte lassen sich in die Klassen quaderförmig (Tetrapaks), Giebelpackung, dreiseitiges gerades Prisma und approximiert quaderförmig einteilen.
Die Objekte `Odol` und `Pril` sind zwar Flaschen von komplexer Gestalt (u.a. nicht rotationssymmetrisch), besitzen aber nur auf zwei relativ großen, gegenüberliegenden und wenig gekrümmten Flächen Aufdrucke. Daher wurden beide als Quader modelliert, deren Dimensionen in etwa der jeweiligen Größe der Flächen mit dem Aufdruck entsprechen.

Objekt	Klasse
`Uelzena_Vollmilch_0.5l`	Giebelpackung
`Reichelt_Schlagsahne_0.5l`	Giebelpackung
`Reichelt_FettarmeMilch_2l`	Giebelpackung
`Brandenburg_Vollmilch_2l`	Giebelpackung
`Reichelt_Vollmilch_1l`	Giebelpackung
`Reichelt_FettarmeMilch_1l`	Giebelpackung
`Mueritz_Buttermilch_1l`	Giebelpackung
`Emzett_FettarmeMilch_1l`	Giebelpackung
`Erlenhof_Vollmilch_1l`	Giebelpackung
`Milsani_Vollmilch_1l`	Tetrapak
`Penny_Vollmilch_1l`	Tetrapak
`Penny_FettarmeMilch_1l`	Tetrapak
`X_TrinkJoghurt_Pfirsich_1l`	Tetrapak
`X_TrinkJoghurt_Erdbeer_1l`	Tetrapak
`Pfanner_EisTee_Ananas_1l`	Tetrapak
`Ranjo_Nektar_Mix_1l`	Tetrapak
`Kirberg_Saft_TraubenRot_1l`	Tetrapak
`Valensina_Saft_Orange_0.75l`	Tetrapak
`Valensina_Saft_Multi_0.75l`	Tetrapak
`Valensina_Saft_Orange_0.2l`	Tetrapak
`Milsani_Kondensmilch_0.312l`	Tetrapak
`Selmi_Kondensmilch_0.312l`	Tetrapak
`Ravensberger_Buttermilch_1l`	Tetrapak
`Toblerone_400g`	dreiseitiges Prisma
`Milram_Quark_Pfeffer_200g`	quaderförmig
`Pril_Supra_0.5l`	Flasche, approx. quaderförmig
`Odol_75ml`	Flasche, approx. quaderförmig

Tabelle 1: Liste der verwendeten Objekte zur Erstellung der Datenbank

3.4.1 Ergebnisse der einzelnen Verarbeitungsstufen

Aufnahme der Bilder

Ein limitierender Faktor bei der Aufnahme der Texturen ist der eingeschränkte Bewegungsbereich der PTU. Für den Pan sind Winkel im Bereich von $-139.8°$ bis $+139.8°$ (relativ zur Nullstellung) und für den Tilt Winkel im Bereich von $-47.3°$ bis $+31.5°$ möglich. Das führt dazu, daß die der Grundfläche oder der zur Kamera gewandten Fläche gegenüberliegenden Flächen nicht aufgenommen werden können. Daher mußten nahezu alle Objekte in mindestens zwei Lagen auf die PTU gestellt werden, um ihre Textur vollständig (mit Ausnahme der zumeist einheitlich weißen Bodenfläche) zu erfassen.

Die Einschränkungen der Bewegungsbereiche lassen sich zwar abschalten, jedoch wird dann keinerlei Überprüfung der Bewegungen vorgenommen und es kann zu Beschädigungen der PTU kommen. Für den Tilt ergibt sich aufgrund mechanischer Beschränkungen auch kein deutlich größerer Bereich. Lediglich für den Pan wären dann alle Positionen erzielbar. Aufgrund von Sicherheitsbedenken bzw. aufgrund des geringen Tilt–Bereichs wurde daher mit mehreren Lagen der Objekte auf der PTU gearbeitet.
Die Giebelverpackungen wurden je zweimal aufgenommen. Bei der ersten Lage standen die Verpackungen auf der Bodenfläche mit dem EAN–Code von der Kamera aus gesehen rechts. Bei der zweiten Lage wurden die Verpackungen auf die Fläche mit dem EAN–Code mit dem Giebel zur Kamera gerichtet gelegt. Quaderförmige Objekte (Tetrapaks u.a.) wurden zwei- oder dreimal aufgenommen, je nachdem, ob sich die Aufdrucke auf Front- und Rückseite unterschieden. Bei der ersten Lage wurden die Objekte auf die Frontfläche geleget, mit der Deckfläche zur Kamera; bei der zweiten und dritten Lage standen die Objekte auf der Bodenfläche, wobei die Front- bzw. die Rückseite zur Kamera ausgerichtet wurden.
Die dreieckigen Seitenflächen des Objekts `Toblerone` wurden beim liegenden Objekt aufgenommen. Die rechteckigen Flächen konnten nur am stehenden Objekt aufgenommen werden, da beim liegenden Objekt eine Tilt–Neigung von $\pm 60.0°$[20] erforderlich wäre. Für die Objekte `Odol` und `Pril`, bei denen nur zwei gegenüberliegende Flächen aufzunehmen waren, ist nur eine Lage erforderlich. Beide Objekte sind approximativ als Quader modelliert. Die Objekte sind stehend mit einer der beiden zwischen den aufzunehmenden Flächen befindlichen Flächen zur Kamera auszurichten.

Die Brennweite des Kameraobjektivs wurde der Größe des aufzunehmenden Objekts angepaßt, so daß bildfüllende Aufnahmen entstanden. Die benutzten Brennweiten lagen im Bereich von 25 bis 38 mm. Die Blende wurde so eingestellt, daß weiße Objektflächen auf den maximalen Wert des linearen Bereichs der CCD–Sensoren der Kamera abgebildet wurden. Bei der verwendeten Kamera (Sony DXC–730P) beträgt dieser Wert 190 (Grauwertebereich 0 bis 255), siehe auch Abschnitt 4.3.1 in [24].
Aufgrund der Verwendung gebrauchter Objekte für die Aufnahmen, ergaben sich kleinere, unvermeidbare Glanzlichter. Sie entstehen an verformten (ausgebeulten bzw. eingedrückten) Oberflächen. Da die meisten Objekte in mehreren Lagen aufgenommen wurden, standen bei einer Reihe von Flächen[21] mehrere Aufnahmen zur Verfügung. Davon wurde diejenige ohne bzw. mit geringstem Anteil an Glanzlichtern ausgewählt. Diese Auswahl kann automatisch erfolgen, da Glanzlichter im Bild sehr dicht am Weißwert des Bildaufnahmesystems liegen. Die Glanzlichter übersteuern die CCD–

[20] falls die dreieckigen Flächen senkrecht zur Bildebene sind, ansonsten $\pm 90.0°$

[21] Flächen, die aus mehreren Lagen mit dem Bewegungsbereich der PTU erreichbar waren

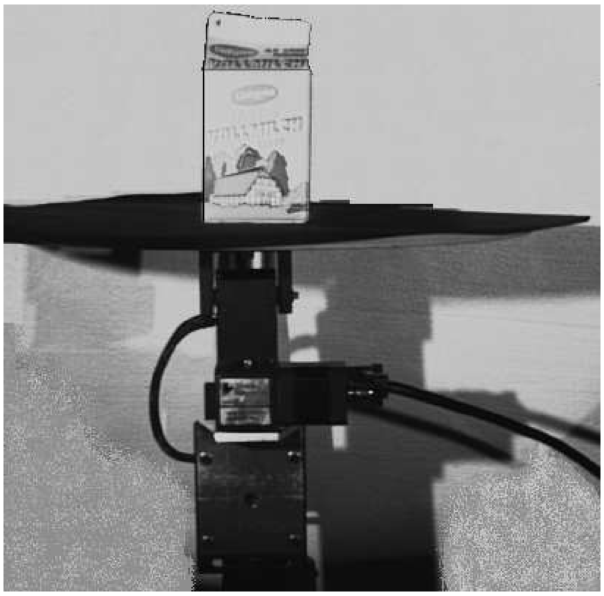

Abbildung 12: PTU mit aufzunehmendem Objekt

Sensoren der Kamera weit über den linearen Bereich, in dem, wie oben beschrieben, alle Werte nicht glanzlichtbehafteter Stellen liegen. Um das Glanzlicht–Problem ganz zu umgehen, sollten die Aufnahmen mit neuen, nicht deformierten Objekten erfolgen, die jedoch für die erstellten Aufnahmen nicht zur Verfügung standen.

Segmentierung

Bei der Hintergrundsegmentierung war der Schwellenwert einzustellen. Da es keine optimale Einstellung gibt, die sowohl hellere Bereiche des Hintergrunds als auch dunklere Bereiche der Objekte sicher erkennt, wurde der Schwellenwert für Objekte mit mittlerer bis hoher Helligkeit angepaßt. Für dunkle Objekte wurde die Hintergrundsegmentierung anhand einer verstärkten Aufnahme vorgenommen[22]. Für die Hintergrundsegmentierung konnten so, trotz des einfachen Verfahrens, sehr gute Ergebnisse erzielt werden. Dies ist natürlich auch den Aufnahmebedingungen (weitgehend schwarz ausgekleideter Hintergrund) zuzuschreiben. Sie stellen jedoch keine Einschränkung für die potentiellen Anwendungen des 3D–Texture–Scanning dar.
Der wichtigste Schritt der Segmentierung ist die Überlagerung mit der synthetischen Linienzeichnung. In den ersten Experimenten wurde die Abbildung zur Überlagerung nur anhand der unteren Extrempunkte definiert. Wie sich jedoch gezeigt hat, waren die Ergebnisse vielfach ungenau, da die

[22]Die verstärkte Aufnahme dient ausschließlich der Hintergrundsegmentierung.

Extrempunkte nicht immer mit den entsprechenden Punkten der Linienzeichnung übereinstimmen[23]. Daher wurde das Verfahren unter Verwendung der linearen Regression eingeführt, was zu einer deutlichen Verbesserung der Ergebnisse geführt hat.
Bei den Aufnahmen liegender Objekte, bei denen die am weitesten links und rechts liegenden sichtbaren Flächen eine stärkere Neigung zur Bildebene aufweisen[24], traten jedoch noch immer größere Fehler auf. Sie sind auf die Annahme von Parallelprojektion bei der Erzeugung der synthetischen Linienzeichnung zurückzuführen, siehe Abb. 13. Bei diesen Objekten wird jedoch bei der gewählten Aufnahmegeometrie die Oberkante auf die Bildmittellinie abgebildet. Hier tritt dann lediglich eine perspektivische Längenverkürzung auf, die relativ gering ist und somit auch nur zu relativ geringen Fehlern bei der Überlagerung führt.

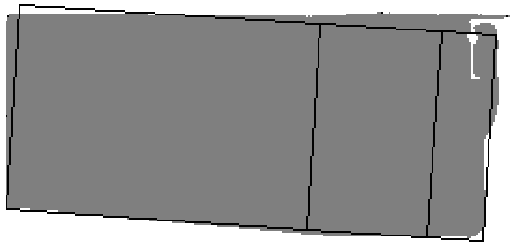

Abbildung 13: Fehler durch Annahme von Parallelprojektion

Bei einigen Aufnahmen wurde auf eine Verwendung der Extrempunkte ganz verzichtet, da sie mit den entsprechenden Punkten der synthetischen Linienzeichnung überhaupt nicht übereinstimmten, siehe Abb. 14 auf der nächsten Seite. Dies ist jedoch auf die hervorstehenden dreieckigen Flächen links und rechts der Basisfläche zurückzuführen, die im Modell nicht enthalten sind. Würde man die vorstehenden Dreiecke wieder zurückbiegen, wie es bei der ungeöffneten Verpackung der Fall ist, so könnte das ursprüngliche Verfahren uneingeschränkt verwendet werden.

Die im allgemeinen erzielte Genauigkeit des Verfahrens ist gut. Die Abweichungen liegen bei den meisten Aufnahmen unter fünf Pixeln. Lediglich für einige „Problemfälle“ sind größere Abweichungen zu verzeichnen. Die erzielbare Genauigkeit ist direkt von der Genauigkeit der CAD–Datensätze abhängig. Fehler bei der Angabe der Flächengrößen im CAD–Datensatz führen zu einer fehlerfhaften synthetischen Linienzeichnung, so daß die Basisfläche nicht korrekt extrahiert werden kann.
Bei vier Objekten entstanden durch Meßfehler bei der manuellen Erstellung der 3D–Modelle Ungenauigkeiten bei der Segmentierung. Ein Teil der Bilder

[23]Dieses Problem tritt vor allem bei stark abgerundeten Ecken auf.
[24]z.B. die Aufnahmen der oberen Flächen der Giebelpackungen bei der zweiten Lage

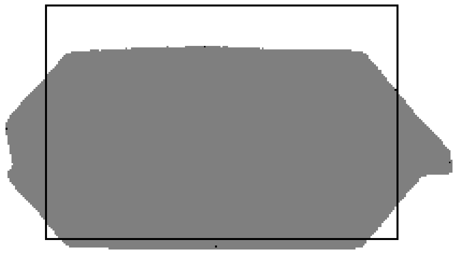

Abbildung 14: Fehler bei der Überlagerung unter Verwendung der Extrempunkte (Deckfläche eines geöffneten Tetrapaks

dieser Objekte enthält einen signifikanten Anteil an Hintergrund (bis zu 9%). Er tritt als schwarzer Streifen an einem der Bildränder auf.
Die Genauigkeit läßt sich eventuell durch die Zuordnung von Linien statt Punkten bei der Bestimmung der Abbildung zur Überlagerung der Linienzeichnung verbessern. Dafür ist auch eine Ausgleichsrechnung vorzusehen, da die Abbildung dann überbestimmt ist.

3.4.2 Zusammenfassende Beurteilung

Insgesamt ergibt sich für die untersuchte Klasse der annähernd polyedrischen Objekte ein praktikables Verfahren zur Erfassung der Oberflächentexturen. Manuelle Arbeiten fallen nur bei der Erstellung des 3D–Modells und der Positionierung der Objekte auf der PTU an. Für den ersten Punkt ist eine Automatisierung möglich (siehe Abschnitt 3.5). Prinzipiell könnte auch die Positionierung der Objekte auf der PTU durch Roboter erfolgen, so daß sich ein vollautomatisches System ergäbe.
Bezüglich der Überlagerung der synthetischen Linienzeichnung sind noch weitere Untersuchungen zur Verbesserung der Robustheit vorzunehmen. Insbesondere ist eine automatische Auswahl des besten Ergebnisses aus den Ergebnissen verschiedener Verfahren denkbar.

Konkavitäten

Bei Objekten mit Konkavitäten ergibt sich das Problem der Verdeckung sowie Interreflexionen[25] und Schatten. Schatten lassen sich durch eine entsprechende Beleuchtung (gleichmäßig aus allen Richtungen) vermeiden. Interreflexionen dagegen sind unvermeidbar. Allerdings ist ihr Anteil am gesamten reflektierten Licht eines jeden Objektpunktes sehr gering. Daher liegt die durch Interreflexionen verursachte Helligkeits- und Farbveränderung in der Regel im Toleranzbereich der nachfolgenden Verarbeitungsprozesse. Weiter-

[25] Intraobjekt–Interreflexionen

hin sind die Intraobjekt–Interreflexionen unabhängig vom Betrachterstandpunkt, so daß zwei zu vergleichende Aufnahmen des Objekts in gleicher Weise verfälscht werden.
Das Problem der Verdeckung ist etwas schwieriger. Im Extremfall kann für innenliegende Oberflächenteile auch keine Ansicht existieren, in der sie sichtbar sind. Tritt die Verdeckung der Basisfläche bei der Normalansicht auf, so sind die verdeckten Teile aus einer anderen Ansicht zu rekonstruieren. Dafür ist das Verfahren entsprechend zu erweitern.
Bei der gegebenen Anwendung stellt sich auch die Frage, ob die Konkavitäten verwertbare Aufdrucke besitzen. Das ist in der Regel nicht der Fall, da Konkavitäten auch vom Kunden schlechter wahrgenommen werden und somit als Informationsträger ungeeignet sind. Die untersuchten Giebelmilchtüten haben zwei Konkavitäten an den Giebelseiten. Sie besitzen an diesen Stellen nur unspezifische[26] (und daher nicht verwertbare) oder gar keine Aufdrucke. Daher sind die Konkavitäten der Einfachheit halber auch nicht im 3D–Modell erfaßt. Dort ist stattdessen eine der konvexen Hülle entsprechende dreieckige Fläche modelliert, die dann als nicht relevant markiert ist.

Gekrümmte Flächen

Für Objekte mit gekrümmten Flächen, die nicht durch eine einzige planare Fläche angenähert werden können oder sollen, ist das hier vorgestellte Verfahren in dieser Form nicht geeignet. Eine gekrümmte Fläche müßte durch relativ kleine planare Stücke approximiert werden, für die dann jeweils eine Aufnahme zu erstellen wäre[27]. Anschließend sind die einzelnen Teile der Textur der gekrümmten Fläche zusammenzusetzen. Dazu ist bei Flächen, die sich nicht auf eine Ebene abwickeln lassen, auch eine geeignete umkehrbare Abbildung auf eine ebene Fläche zu finden. Alternativ könnte man die Textur auch als nicht–planares Bild modellieren.

3.5 Automatische Gewinnung der Geometriedaten

Der Automatisierungsgrad des Verfahrens läßt sich durch eine automatische Gewinnung der Geometriedaten (CAD–Datensätze) der Objekte noch weiter erhöhen. Es ist zu erwarten, daß die Genauigkeit der Daten größer ist als die der manuell ermittelten. Andererseits ist dafür Sorge zu tragen, daß kleine Details der Objektgestalt nicht erfaßt werden, um die Komplexität des Modells nicht unnötig zu erhöhen. So sind z.B. aus technischen Gründen nahezu alle Kanten der Verpackungsobjekte abgerundet. Es ist jedoch nicht sinnvoll, diese Rundungen zu modellieren.

[26] drucktechnisch bedingte Aufdrucke
[27] Die planaren Stücke müssen so klein sein, daß durch die Krümmung auf dem entsprechenden Stück des Objekts noch keine größeren Verzerrungen auftreten.

Das für die 2.5D–Rekonstruktion der zu erkennenden Objekte eingesetzte PHOTOMETRIC–STEREO–Verfahren ist prinzipiell für diese Aufgabe geeignet. Es sind jedoch mehrere Ansichten[28] zu rekonstruieren und zu einem vollständigen 3D–Modell des Objekts zusammenzufügen. Weiterhin sind auch die erzielbaren Genauigkeiten für eine exakt parallele Ausrichtung der Objektflächen nicht hoch genug. Die dann auftretende perspektivische Verzerrung der Datenbankbilder würde jedoch von vielen Matching–Verfahren toleriert.
Weitere Möglichkeiten wären Strukturierte Beleuchtung (ebenfalls 2.5D–Rekonstruktion) und das von Marian Sänger entwickelte Laser–Rekonstruktionsverfahren [37] (3D–Rekonstruktion).

[28] Mit Hilfe der PTU können vollautomatisch Aufnahmen verschiedener Ansichten erstellt werden.

4 Beleuchtungsunbhängigkeit

In der Literatur wird Farbkonstanz (*color constancy*) vielfach im Zusammenhang mit dem menschlichen visuellen System untersucht, und darauf aufbauend werden Modelle entwickelt. Dabei liegt in der Regel kein genaues Wissen über die Beleuchtung vor, auch kann nicht immer von einer bekannten Gestalt der Objekte ausgegangen werden oder es sind Verdeckungen zu berücksichtigen.
Für das vorgestellte Objekterkennungssystem können sich die Verfahren im wesentlichen auf den zweidimensionalen Fall beschränken, da die Algorithmen für die Farbkonstanz erst auf die separierten und rekonstruierten Flächen angewandt werden. Die Neigung der einzelnen planaren Flächenstücke zur Bildebene bedeutet dabei nur eine Änderung der Lichtquellenrichtung relativ zur Normalen der Fläche. Das einfallende Licht ist an jedem Ort auf einem gegebenen planaren Flächenstück unter der Annahme einer globalen Beleuchtung konstant. Werden jedoch stärker gekrümmte Flächen durch planare Flächen approximiert, so ist die krümmungsbedingte Änderung der Schattierung zu berücksichtigen.
Idealerweise wird von einer absolut gleichmäßigen, ungerichteten Beleuchtung ausgegangen, die sich in der Praxis jedoch nicht realisieren läßt. In der Literatur wird oftmals auch angenommen, daß die betrachtete Fläche zeitlich oder örtlich getrennt von zwei verschiedenen Lichtquellen beleuchtet wird. Das trifft für das vorgestellte Objekterkennungssystem jedoch nicht zu.
Da für die Aufnahmen mit dem PHOTOMETRIC–STEREO–Verfahren bereits von Lambert–Reflexion ausgegangen werden mußte, könnte man diese Annahme auch im Zusammenhang mit der Erzielung von Farbkonstanz machen. Um jedoch in der Wahl des 2.5D–Rekonstruktionsverfahrens offen zu bleiben, wird auf diese Einschränkung verzichtet.

Für die Anwendungen des vorgestellten Objekterkennungssytems sind hauptsächlich Helligkeitsschwankungen zu berücksichtigen. Die Beleuchtung der Szene erfolgt immer mit (nahezu) weißem Licht. Unterschiede in der spektralen Verteilung der Intensität der Beleuchtung werden schon sehr gut durch den Weißabgleich der Kamera ausgeglichen. Abweichungen sind jedoch im Laufe des Betriebs durch die Erwärmung der CCD–Sensoren der Kamera (soweit nicht temperaturstabilisiert) und Schwankungen der Farbtemperatur der Lichtquellen möglich. Bei den verwendeten Diaprojektoren mit Niedervolthalogenlampen können insbesondere Schwankungen der Netzspannung zur Änderung der Farbtemperatur des erzeugten Lichts führen.
Der Einfluß von Fremdlicht kann sowohl im Labor als auch unter den in einer späteren Anwendung zu erwartenden Bedingungen ausgeschlossen werden.

4.1 Ausgleich von Helligkeitsschwankungen

Die Schwankungen in der gemessenen Helligkeit resultieren zum einen aus unterschiedlich heller Beleuchtung (Helligkeit und Anordnung der Lichtquellen) und zum anderen aus unterschiedlichen Einstellungen des bildgebenden Systems. Beide — Beleuchtung und bildgebendes System — können für zwei zu vergleichende Aufnahmen nicht immer in denselben Zustand gebracht werden. Die Anforderungen an beide sind für die Aufnahmen der Bilder für die Datenbank und die Aufnahmen für die Rekonstruktion auch verschieden, so daß von abweichenden Einstellungen auszugehen ist.

Eine eventuell vorhandene Blendenautomatik der Kamera ist für den Ausgleich der Helligkeitsschwankungen ungeeignet, da die Kamera dabei auch die Helligkeit des Hintergrunds bewertet[29] und alle sichtbaren Flächen des Objekts gemeinsam berücksichtigt. Man stelle sich nur den Extremfall eines Quaders vor mit Flächen, die entweder einheitlich weiß oder schwarz sind. Bei der Aufnahme der Bilder für die Datenbank würden die schwarzen Flächen als etwas dunkleres Grau und die weißen Flächen als etwas helleres Grau aufgenommen. Liegt der Quader als zu erkennendes Objekt vor der Kamera und seien zwei schwarze Flächen und eine weiße Fläche sichtbar, so erscheinen die schwarzen Flächen trotz unveränderter Beleuchtung dunkler als bei der Aufnahme für die Datenbank und die weiße Fläche erscheint entsprechend heller.

Solange die Intensität nicht so gering wird, daß die Bestimmung von Sättigung und Buntton sehr ungenau wird, und die Intensität auch nicht so hoch wird, daß ein Teil der drei Kamerasensoren den linearen Arbeitsbereich verlassen[30], könnte man in einem ersten Ansatz das Problem lösen, indem man beim Matching nur den Buntton und die Sättigung berücksichtigt. Gerade bei Verpackungen werden aber zu einem großen Anteil stark gesättigte Farben eingesetzt, so daß effektiv nur der Buntton zur Unterscheidung dienen kann. Dies führt jedoch zu schlechteren Ergebnissen (siehe [29]), da es für die Unterscheidung der Objekte sehr wohl von Bedeutung sein kann, ob es sich um eine helle oder dunkle Variante derselben Farbe handelt. Weiterhin ist die Genauigkeit der Bestimmung des Bunttons bei geringer Sättigung oder geringer Intensität sowie bei hoher Intensität bei gleichzeitig geringer Sättigung (stark verweißlichte Farben) sehr gering. Eine mögliche Lösung ist die in Abschnitt 5.2.6 beschriebene dynamische Quantisierung des Bunttons. Daher ist es notwendig, die gemessenen Intensitätswerte so zu korrigieren, daß eine von der Beleuchtung und den Kameraeinstellungen unabhängige Beschreibung der Objekteigenschaften vorliegt.

[29] Der Hintergrund ist zwar für alle Objekte gleich, hat aber je nach Größe und Gestalt des Objekts einen wechselnden Anteil am Bild.

[30] Um die Bildverstärker nicht zu übersteuern, wird ab einer bestimmten Intensität das Ausgangssignal der CCD–Sensoren abgeschwächt, siehe [24].

Eine Korrektur der Intensität läßt sich natürlich am einfachsten in einem Farbmodell beschreiben, das wie z.B. HSI die Intensität in einem unabhängigen Kanal kodiert. Das Gegenfarbenmodell besitzt zwar den Schwarz–Weiß–Kanal, die Intensität ist jedoch in allen Kanälen enthalten:

$$\begin{aligned} I_1 &= R - G \\ I_2 &= 2B - R - G \\ I_3 &= \frac{R + G + B}{3} \end{aligned}$$

Mit einer Intensitätsänderung $(R', G', B') = \lambda(R, G, B)$ ergibt sich:

$$\begin{aligned} I_1' &= \lambda R - \lambda G &&= \lambda I_1 \\ I_2' &= 2\lambda B - \lambda R - \lambda G &&= \lambda I_2 \\ I_3' &= \frac{\lambda R + \lambda G + \lambda B}{3} &&= \lambda I_3 \end{aligned}$$

D.h. das Gegenfarbenmodell verhält sich bezüglich der Intensität wie das RGB–Modell.
Im CIELAB–Modell ist die Intensität, wie beim HSI–Modell, in einem unabhängigen Kanal kodiert. Allerdings ist der Zusammenhang zwischen der Helligkeit L* und der Intensität nicht linear. Die Größe L* entspricht einer visuell gleichmäßigen Stufung.
Ein Ansatz zur Erzielung von Beleuchtungsunabhängigkeit bezüglich der Helligkeit basiert darauf, daß sich die gemessene Intensität in zwei Aufnahmen derselben Szene mit verschieden heller Beleuchtung oder verschiedener Blendenöffnung nur durch einen konstanten Faktor unterscheidet.
Es wird die mittlere Intensität (arithmetisches Mittel) des Bildes bestimmt. Dann wird der Quotient aus der gemessenen mittleren Intensität und einer festen Intensität (vorzugsweise in der Mitte des Wertebereichs der Intensitäten) berechnet. Anschließend wird die Intensität in jedem Bildpunkt durch Multiplikation mit diesem Quotienten normiert.
Beim RGB–Modell ist die Intensität gleich der Summe der drei Farbkanäle. Gegebenenfalls kann auch eine gewichtete Summe (wie bei der Berechnung der Intensität in einigen am menschlichen visuellen System orientierten Farbmodellen) verwendet werden. Es werden alle Kanäle mit dem oben beschriebenen Quotienten normiert. Beim HSI– und beim CIELAB–Modell ist die Intensität in einem getrennten Kanal kodiert, der unabhängig von den anderen beiden Kanälen entsprechend normiert wird. Dabei ist beim CIELAB–Modell noch zu beachten, daß der $L*$–Kanal nicht linear mit der Intensität zusammenhängt, so daß die Normierung nicht direkt erfolgen kann. Der Schwarz–Weiß–Kanal beim Gegenfarbenmodell ist, wie bereits oben dargestellt, nicht der alleinige Träger von Intensitätsinformation, so daß (wie beim RGB–Modell) alle Kanäle normiert werden müssen.

4.2 Farbkonstanz

Einfache Methoden zur Erzielung von Farbkonstanz bedienen sich einer in jeder Szene befindlichen Referenztafel mit bekanntem spektralen Reflexionsfaktor oder bestimmen die Farbe der Beleuchtung aus Glanzlichtern, da diese nur wenig vom spektralen Reflexionsfaktor des Objekts beeinflußt werden. Dies war im Rahmen dieser Arbeit nicht möglich, da bereits vorhandene Aufnahmen verwendet werden sollten und die Erstellung einer Referenztafel mit den vorhandenen Meßgeräten nicht möglich war. Die Aufnahmen wurden auch (aufgrund der Anwendung) so erstellt, daß möglichst keine Glanzlichter vorkamen.
Bei der Anwendung allgemeiner Verfahren sind keine allzu guten Ergebnisse zu erwarten, da keine genauen Kenntnisse über spektrale Reflektion der Objekte und die spektrale Energieverteilung der Lichtquellen vorausgesetzt werden kann.

Wie bereits oben diskutiert, treten bei den gegebenen Anwendungen des vorgestellten Objekterkennungssystems nur geringe Farbabweichungen auf. Die Anforderung, Objekte unter stark in der Farbe variierender Beleuchtung wiederzuerkennen, wie sie bei den Betrachtungen zur Farbkonstanz in der Literatur üblich ist, ist hier nicht gegeben. Die Abweichungen vom Weiß sind so gering, daß ein regelmäßiger Weißabgleich bereits zufriedenstellende Ergebnisse liefert.
Auch in einer anderen Hinsicht ist für die gegebenen Anwendungen nur eine abgeschwächte Form der Farbkonstanz nötig: Es ist keine Rekonstruktion des tatsächlichen spektralen Reflexionsfaktors der Oberfläche erforderlich, sondern lediglich eine beleuchtungsinvariante Beschreibung zu finden. Diese muß sich jedoch aus Aufnahmen mit nur einer, unbekannten Beleuchtung berechnen lassen. Dazu sind beleuchtungsinvariante Merkmale der spektralen Reflektanz der Oberflächen der Objekte zu finden. Ein Verfahren dieser Art beschreibt [18]. Dieses Verfahren zeigt auch eine weitere Idee: die Integration der Farbkonstanz in das Matching–Verfahren.
In dem vorgestellten Objekterkennungssystem wird Farbkonstanz ausschließlich als integraler Bestandteil von Matching–Verfahren betrachtet. Ein separater Verarbeitungsschritt zur Erzielung von Farbkonstanz verursacht einen zu großen Rechenzeitaufwand.

5 Matchingverfahren

In diesem Abschnitt werden die Verfahren zur Berechnung und zum Vergleich der Farbmerkmale der Bilder vorgestellt. Der erste Unterabschnitt widmet sich den Verfahren, die bereits in der Literatur beschrieben wurden. Im zweiten Unterabschnitt werden dann Erweiterungen und Weiterentwicklungen sowie die Kombination verschiedener Verfahren vorgestellt.

5.1 Verwendete Verfahren aus der Literatur

5.1.1 Farbindizierung (*Color Indexing*)

M. Swain und D. Ballard haben ein Verfahren entwickelt und in [46] beschrieben, daß Objekte ausschließlich anhand ihrer Farbe erkennt. Der Begriff *color indexing* bezeichnete ursprünglich nur das Verfahren von M. Swain und D. Ballard. Inzwischen wird *color indexing* jedoch oft auch als Oberbegriff für alle Verfahren verwendet, die ausschließlich Farbmerkmale verwenden.
Beim Verfahren von M. Swain und D. Ballard wird für zwei zu vergleichende Bilder (bzw. Bildausschnitte) jeweils ein Farbhistogramm berechnet. Die Ähnlichkeit der beiden Histogramme wird dann mit Hilfe der Histogrammschnitt–Methode (*histogram intersection*) bestimmt.
Als Farbmodell schlagen M. Swain und D. Ballard das Gegenfarbenmodell vor, wobei sie eine Auflösung von 16 Stufen für Rot–Grün (RG) und Blau–Gelb (BY) verwenden sowie von 8 Stufen bei Weiß–Schwarz (WB). Die verringerte Farbauflösung dient vor allem dazu, das Verfahren toleranter gegenüber kleineren Farbschwankungen zu gestalten und die zu verarbeitende Datenmenge zu reduzieren.

Histogrammschnitt–Methode

Bei der Histogrammschnitt–Methode wird die Ähnlichkeit zweier Histogramme bestimmt. Das Ergebnis ist eine reelle Zahl aus dem abgeschlossenen Intervall $[0, 1]$, wobei 1 vollständige Übereinstimmung bedeutet.
Es wird für jedes Bin das Minimum beider Histogramme bestimmt und über alle Bins summiert. Dieser Wert ist durch die Summe über alle Bins eines Histogramms zu teilen. Dies ist jedoch nur möglich, falls beide Histogramme über die gleiche Anzahl von Pixeln gebildet wurden.
In [29] schlagen K. Oppermann und W.J. Schimke vor, relative Histogramme zu benutzen. Das hat den Vorteil, daß beim Histogrammschnitt keine Normierung erforderlich ist. Das Ähnlichkeitsmaß ergibt sich direkt aus der Summe der Minima. Der Rechenzeitbedarf steigt durch die Verwendung von Gleitkommaarithmetik auf heutigen Prozessoren nicht merklich an.

Sei $H(c)$ das Farbhistogramm eines Bildes B, c ein Farbvektor (c_1, c_2, c_3) (bei einem dreidimensionalen Farbenraum), N die Anzahl der Pixel von B,

p ein Pixel und $C(p)$ sein Vektor im Farbenraum, dann gilt:

$$H(c) = \frac{1}{N} |\{p : p \in B, C(p) = c\}|, \tag{25}$$

wobei $|\{p : p \in B, C(p) = c\}|$ die Anzahl der Pixel mit der Farbe c ist.
Der Schnitt zweier Histogramme H_1 und H_2 (*histogram intersection*) ist dann definiert als

$$d_{\mathrm{HI}} = \sum_c \min(H_1(c), H_2(c)). \tag{26}$$

Der Histogrammschnitt ist äquivalent zur L_1–Metrik (City–Block–Metrik, Summe der absoluten Differenzen).

5.1.2 Metriken und Summenhistogramme

Das Hauptproblem bei der Farbindizierung nach Swain/Ballard ist die Einstellung der Quantisierung der Farben. Eine zu grobe Quantisierung läßt kaum noch Unterscheidungen verschiedener Bilder zu, eine zu feine Quantisierung ist nicht robust genug. Das Verfahren berücksichtigt nicht den Abstand der Bins. D.h., fällt aufgrund der Quantisierung eine nur leicht vom Wert in der Datenbank abweichende Farbe in ein benachbartes Bin, so ist das Ergebnis des Histogrammschnitts nicht anders als wenn sich die Farbe stark von der in der Datenbank unterscheiden würde.
Daher schlagen M. Stricker und M. Orengo in [43] die Verwendung von Summenhistogrammen vor. Bei der oben dargestellten leichten Farbabweichung ergibt sich nur eine leichte Veränderung des Histogramms.
Das Summenhistogramm $SH(c)$ berechnet sich aus dem Histogramm $H(c)$ mit

$$SH(c) = \sum_{c' \leq c} H(c'), \tag{27}$$

wobei $(c_1, c_2, c_3) = c \leq c' = (c'_1, c'_2, c'_3)$ genau dann gilt, wenn $c_1 \leq c'_1 \wedge c_2 \leq c'_2 \wedge c_3 \leq c'_3$ erfüllt ist.
Zum Vergleich der Histogramme[31] verwenden M. Stricker und M. Orengo die L_1–, die L_2– und die L_∞–Metrik:

$$d_{L_1}(H_1, H_2) = \sum_c |H_1(c) - H_2(c)| \tag{28}$$

$$d_{L_2}(H_1, H_2) = \sqrt{\sum_c (H_1(c) - H_2(c))^2} \tag{29}$$

$$d_{L_\infty}(H_1, H_2) = \max_c |H_1(c) - H_2(c)| \tag{30}$$

[31] sowohl für Histogramme als auch für Summenhistogramme

5.1.3 Statistische Eigenschaften der Farbverteilung

Auch das in Abschnitt 5.1.2 vorgestellte Verfahren besitzt Quantisierungsparameter, deren optimale Werte schwer zu finden sind. M. Stricker und M. Orengo stellen daher in [43] noch ein weiteres Verfahren vor: die Verwendung von statistischen Eigenschaften der Farbverteilung.
Eine Wahrscheinlichkeitsverteilung kann eindeutig durch ihre Momente beschrieben werden. Faßt man nun die Farbverteilung als Wahrscheinlichkeitsverteilung auf, so hat man mit den Momenten eine alternative Beschreibungsmöglichkeit zum Histogramm. Zur Vereinfachung werden die Momente nicht über der dreidimensionalen Farbverteilung selbst, sondern getrennt für die einzelnen Farbkanäle berechnet.
M. Stricker und M. Orengo verwenden die ersten drei Momente (Mittelwert[32], Varianz und Kovarianz[33]) der Verteilungen eines jeden Farbkanals, da sie bereits genügend Information enthalten, um Farbverteilungen zu charakterisieren.

Sei $C_i(p)$ der Wert im i-ten Farbkanal ($1 \leq i \leq r$, $r =$ Anzahl der Farbkanäle) des Pixels p und N die Anzahl der Pixel im Bild B, dann berechnen sich der Mittelwert E_i, die Standardabweichung σ_i (Quadratwurzel der Varianz) und die Kubikwurzel der Kovarianz s_i folgendermaßen:

$$E_i = \frac{1}{N} \sum_{p \in B} C_i(p) \tag{31}$$

$$\sigma_i = \sqrt{\frac{1}{N} \sum_{p \in B} (C_i(p) - E_i)^2} \tag{32}$$

$$s_i = \sqrt[3]{\frac{1}{N} \sum_{p \in B} (C_i(p) - E_i)^3} \tag{33}$$

Die Ähnlichkeitsfunktion ist dann die gewichtete Summe der absoluten Differenzen der Momente.

Seien M_1 und M_2 die Verteilungen zweier Bilder mit den Momenten $E_{i,1}$, $\sigma_{i,1}$ und $s_{i,1}$ bzw. $E_{i,2}$, $\sigma_{i,2}$ und $s_{i,2}$, dann ist die Ähnlichkeitsfunktion definiert durch

$$d_{\text{mom}}(M_1, M_2) = \sum_{1 \leq i \leq r} w_{i,1}|E_{i,1} - E_{i,2}| + w_{i,2}|\sigma_{i,1} - \sigma_{i,2}| + w_{i,3}|s_{i,1} - s_{i,2}| \tag{34}$$

wobei w_{ij} die Gewichte sind.
Da nur eine endliche Menge von Momenten berücksichtigt wird, ist d_{mom} keine Metrik für die Verteilungen, weil zwei verschiedene Verteilungen den Abstand 0 haben können.

[32] Anfangsmoment erster Ordnung

[33] Zentralmomente zweiter und dritter Ordnung

5.1.4 Invarianten von Farbhistogrammen

G. Healey und D. Slater beschreiben in [18] ein Verfahren zum beleuchtungsunabhängigen Vergleich von Farbverteilungen. Bei einem trichromatischen System[34] wird angenommen, daß sich der spektrale Reflexionsfaktor der Oberfläche des Objekts in einem dreidimensionalen linearen Modell beschreiben läßt. Das Verfahren basiert auf einem von G. Taubin und D. Cooper in [48] vorgestellten Verfahren zur Berechnung algebraischer Invarianten (*moment invariants*), das G. Healey und D. Slater auf dreidimensionale Farbverteilungen anwenden.
Im folgenden wird die Herleitung aus [18] kurz wiedergegeben.

Modellierung von Farbverteilungen

Betrachtet man ein mehrkanaliges Bildaufnahmesystem mit n Kanälen, so werden für jedes Pixel i n Meßwerte bestimmt:

$$\rho_k^i = \int\limits_{\lambda} l(\lambda) s^i(\lambda) f_k(\lambda) d\lambda \qquad (1 \leq k \leq n), \tag{35}$$

wobei λ die Wellenlänge bezeichnet, $l(\lambda)$ die spektrale Energieverteilung der Beleuchtung, $s^i(\lambda)$ der spektrale Reflexionsfaktor im Urbildpunkt des Pixels i und $f_k(\lambda)$ die spektrale Empfindlichkeit der k–ten Sensorklasse. Für ein trichromatisches System (wie z.B. eine RGB–Kamera) ist $n = 3$.
Der spektrale Reflexionsfaktor im Urbildpunkt des Pixels i wird durch

$$s^i(\lambda) = \sum_{1 \leq j \leq n} \sigma_j^i S_j(\lambda) \tag{36}$$

angenähert, wobei $S_j(\lambda)$ eine feste Menge von n Basisfunktionen ist.
Sei $\overrightarrow{\rho^i}$ der Vektor der Sensormeßwerte $(\rho_1^i, \rho_2^i, \ldots, \rho_n^i)^{\mathrm{T}}$ und $\overrightarrow{\sigma^i}$ der Vektor der Koeffizienten des spektralen Reflexionsfaktors $(\sigma_1^i, \sigma_2^i, \ldots, \sigma_n^i)^{\mathrm{T}}$, dann gilt

$$\overrightarrow{\rho^i} = A\overrightarrow{\sigma^i}, \tag{37}$$

wobei A eine $n \times n$–Matrix der Form

$$A_{kj} = \int\limits_{\lambda} l(\lambda) S_j(\lambda) f_k(\lambda) d\lambda \tag{38}$$

ist. Es ist zu beachten, daß die Matrix A zwar von der Beleuchtung nicht jedoch vom Ort abhängt[35].

[34] Ein solches wurde sowohl von G. Healey und D. Slater als auch für die Experimente zu dieser Arbeit verwendet.
[35] Die Beleuchtung wird als ortsinvariant (global) angenommen.

Für zwei Bilder derselben Szene, die unter verschiedener Beleuchtung ($l(\lambda)$ bzw. $\tilde{l}(\lambda)$) aufgenommen wurden, gilt:

$$\overrightarrow{\rho^{i}} = A\overrightarrow{\sigma^{i}} \qquad \text{und} \qquad \overrightarrow{\tilde{\rho}^{i}} = \tilde{A}\overrightarrow{\sigma^{i}}. \tag{39}$$

Für jedes Paar von Beleuchtungen, bei denen sowohl A als auch $\tilde{A}$ nicht singulär werden, gilt somit:

$$\overrightarrow{\tilde{\rho}^{i}} = M\overrightarrow{\rho^{i}}, \tag{40}$$

mit $M = \tilde{A}A^{-1}$.
Seien H bzw. $\tilde{H}$ die zugehörigen n–dimensionalen Histogramme, dann gilt nach Gleichung (40):

$$\tilde{H}(M\overrightarrow{\rho}) = H(\overrightarrow{\rho}). \tag{41}$$

D.h., eine Änderung der Beleuchtung korrespondiert mit einer affinen Koordinatentransformation der Histogramme.

Momente von Farbverteilungen

Ein Zentralmoment eines Farbhistogramms $H(\overrightarrow{\rho})$ sei definiert durch

$$A^{\overrightarrow{\alpha}} = \frac{1}{|H(\overrightarrow{\rho})|}\int(\overrightarrow{\rho} - \overrightarrow{\overline{H}})^{\overrightarrow{\alpha}} H(\overrightarrow{\rho})d\overrightarrow{\rho}, \tag{42}$$

wobei $\overrightarrow{\overline{H}} = [\frac{1}{|H(\overrightarrow{\rho})|}]\int \overrightarrow{\rho} H(\overrightarrow{\rho})d\overrightarrow{\rho}$ das arithmetische Mittel von $H(\overrightarrow{\rho})$ ist, mit $|H(\overrightarrow{\rho})| = \int H(\overrightarrow{\rho})d\overrightarrow{\rho}$. Die Notation $\overrightarrow{\rho}^{\overrightarrow{\alpha}}$ bezeichnet dabei $\prod_{1\leq i\leq n}\rho_i^{\alpha_i}$ mit dem Vektor $\overrightarrow{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_n)$. Dieser Vektor aus nicht negativen ganzen Zahlen wird Multiindex genannt. Sein Betrag sei definiert durch $|\overrightarrow{\alpha}| = \sum_{1\leq i\leq n}\alpha_i$.
Sei die Fakultät von $\overrightarrow{\alpha}$ definiert durch $\overrightarrow{\alpha}! = \prod_{1\leq i\leq n}\alpha_i!$ und sei eine lexikographische Ordnung über Multiindizes folgendermaßen definiert: Für zwei Multiindizes $\overrightarrow{\alpha}$ und $\overrightarrow{\beta}$ gilt $\overrightarrow{\alpha} > \overrightarrow{\beta}$, wenn $\alpha_k < \beta_k$ für das erste k mit $\alpha_k \neq \beta_k$ gilt.
$\overrightarrow{X_{[d]}(\overrightarrow{\rho})}$ sei definiert als der Vektor der Menge $\{\overrightarrow{\rho}^{\overrightarrow{\alpha}}/\sqrt{\overrightarrow{\alpha}!} : |\overrightarrow{\alpha}| = d\}$ in lexikographischer Ordnung, z.B. (für $n = 3$):

$$\overrightarrow{X_{[2]}(\overrightarrow{\rho})} = \left(\frac{\rho_1^2}{\sqrt{2}}, \rho_1\rho_2, \rho_1\rho_3, \frac{\rho_2^2}{\sqrt{2}}, \rho_2\rho_3, \frac{\rho_3^2}{\sqrt{2}}\right)^{\mathrm{T}}. \tag{43}$$

Sei $X_{[k,l]}(\overrightarrow{\rho})$ die Matrix $\overrightarrow{X_{[k]}(\overrightarrow{\rho})}\,\overrightarrow{X_{[l]}(\overrightarrow{\rho})}^{\mathrm{T}}$, dann lassen sich Vektoren und

Matrizen von Zentralmomenten wie folgt definieren[36]:

$$\overrightarrow{A_{[d]}} = \frac{1}{|H(\overrightarrow{\rho})|} \int \overrightarrow{X_{[d]}(\overrightarrow{\rho} - \overrightarrow{\overrightarrow{H}})} H(\overrightarrow{\rho}) d\rho \qquad \text{und} \tag{44}$$

$$A_{[k,l]} = \frac{1}{|H(\overrightarrow{\rho})|} \int X_{[k,l]}(\overrightarrow{\rho} - \overrightarrow{\overrightarrow{H}}) H(\overrightarrow{\rho}) d\rho. \tag{45}$$

Invarianten

Ziel ist es, Invarianten zu finden, die unabhängig von der durch die Matrix M aus Gleichung (41) beschriebenen Beleuchtungsänderung sind. Falls die Matrix M orthogonal ist (entsprechend einer Rotation der Histogramme im Farbenraum), so sind ihre Eigenwerte Invarianten und unabhängig von M. Zur Bestimmung von Invarianten einer allgemeinen affinen Transformation ist diese in eine Koordinatentransformation und eine orthogonale Transformation zu zerlegen. Die Koordinatentransformation wird dabei durch die Matrix L beschrieben, die die Lösung der Gleichung

$$LA_{[1,1]}L^{\mathrm{T}} = I \tag{46}$$

ist[37]. L läßt sich aus der Cholesky–Zerlegung[38] von $A_{[1,1]}$ mit anschließender Invertierung bestimmen. Als Invarianten dienen dann die Eigenwerte der Matrix $A_{[d,d]}$ für das Histogramm $H(L\overrightarrow{\rho})$.
In der Implementierung wurden, wie von G. Healey und D. Slater vorgeschlagen, die Eigenwerte von $A_{[2,2]}$ verwendet. Bei einem trichromatischen System ergeben sich sechs nicht–negative reelle Zahlen.

5.1.5 Integration räumlicher Information

Das Fehlen von Information über die räumliche Verteilung der Farben im Bild ist ein Hauptgrund für Ungenauigkeiten in den Ergebnissen der bisher vorgestellten Verfahren.
M. Stricker und A. Dimai beschreiben in [42] ein Verfahren, das die in Abschnitt 5.1.3 beschriebenen Momente für fünf unscharf begrenzte Bildregionen (*fuzzy regions*) berechnet. Die erste Region ist eine zentrierte Ellipse, deren Halbachsen parallel zu den Koordinatenachsen sind. Die anderen vier Regionen entstehen, indem das Komplement der ersten Region entlang der horizontalen und der vertikalen Bildmittellinie geteilt wird. Die Übergangsbereiche der Regionen sind cosinusförmig abfallend, so daß leichte Verschiebungen von Bildteilen im Übergangsbereich der Regionen nicht zu starken Änderungen der berechneten Merkmale führen.

---

[36] $X_{[k,l]}(\overrightarrow{\rho})$ ist eine Matrix als Funktion von $\overrightarrow{\rho}$. $X_{[k,l]}(\overrightarrow{\rho} - \overrightarrow{\overrightarrow{H}})$ bedeutet daher, daß $\overrightarrow{\rho} - \overrightarrow{\overrightarrow{H}}$ anstelle von $\overrightarrow{\rho}$ als Funktionsargument verwendet wird. Diese Subtraktion von $\overrightarrow{\overrightarrow{H}}$ ist die „Zentrierung", d.h., der Erwartungswert von $\overrightarrow{\rho} - \overrightarrow{\overrightarrow{H}}$ ist 0.

[37] I ist die Einheitsmatrix.

[38] Die Matrizen sind symmetrisch und positiv definit.

Die Parameter der Regionen sind die Länge der Halbachsen (relativ zur Breite bzw. Höhe des Bildes) und die Breite des Übergangsbereichs (in Pixeln). Abbildung 15 zeigt ein Beispiel, bei dem der Übergangsbereich zur Veranschaulichung besonders breit gewählt wurde.

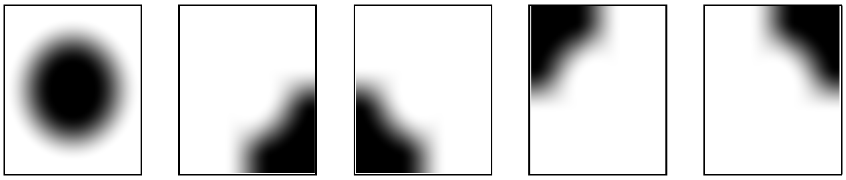

Abbildung 15: *Fuzzy Regions* nach M. Stricker und A. Dimai, Bildgröße 200×250 Pixel, Halbachsen jeweils 0.3, Übergangsbereich 70 Pixel. Die Regionenzugehörigkeit ist, soweit drucktechnisch möglich, proportional zur Schwärzung.

Bei der Berechnung der zu vergleichenden Merkmale, in diesem Fall arithmetisches Mittel, Varianz und Kovarianz, ist die Regionenzugehörigkeit der Pixel zu beachten, und die Merkmale sind für jede Region getrennt zu berechnen.

Sei R_l eine Region, dargestellt als Funktion von der Menge der Pixel in das Intervall $[0,1]$, wobei $R(p) = 1$ bedeutet, daß p zur Region R_i gehört, $R(p) = 0$, daß p nicht zur Region R_i gehört, und $0 < R(p) < 1$, daß p im Übergangsbereich liegt, dann berechnet sich z.B. der Mittelwert in Anlehnung an Gleichung (31) wie folgt:

$$E_{l,i} = \frac{1}{N \cdot |R_l|} \sum_{p \in B} R_l(p) C_i(p) \tag{47}$$

mit

$$|R_l| = \sum_p R_l(p). \tag{48}$$

Durch den Vergleich der vier Randregionen mit zyklischer Vertauschung wird eine eingeschränkte Rotationsinvarianz (für Rotationswinkel, die Vielfache von 90° sind) erreicht. Für die Verwendung im hier vorgestellten Objekterkennungssystem ist diese eingeschränkte Rotationsinvarianz nicht ausreichend, da beliebige Rotationen möglich sind (siehe Abschnitt 6.1)[39].

5.1.6 Tupel–Histogramme

Ein weiterer Ansatz zur Integration von Information über die räumliche Verteilung der Farben wurde von R. Rickman und J. Stonham entwickelt

[39] Das Verfahren wurde für die Suche in Bilddatenbanken entwickelt, wo genau diese eingeschränkte Rotationsinvarianz benötigt wird, siehe Abschnitt 8.5.

[34]. Dabei werden nicht Histogramme von einzelnen Bildregionen bestimmt, sondern ein Histogramm der räumlichen Beziehungen von Farben.

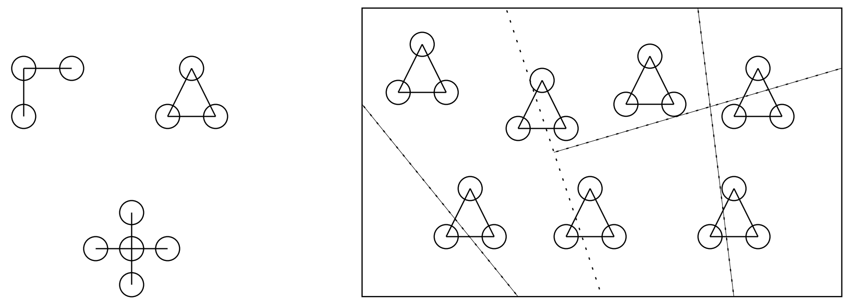

Abbildung 16: Tupelgeometrie: links: zwei mögliche Anordnungen für 3–Tupel und eine Anordnung für ein 5–Tupel, rechts: Beispiele für „Proben“ mit einem der 3–Tupel (Farbkanten sind durch gepunktete Linien angedeutet)

Es werden k „Proben“ aus einem Bild gezogen, wobei k von der Anzahl der Pixel im Bild abhängig sein sollte, z.B. $k = 25\%$ der Pixel. Die Orte der „Proben“ werden zufällig gewählt. Somit ist zu erwarten, daß die Ergebnisse auch von der Güte des Zufallszahlengenerators abhängen. Bei jeder „Probe“ werden n Pixel in einer festen geometrischen Anordnung betrachtet (siehe Abb. 16 links). Aus den Farbwerten aller n Pixel wird dann der Index im Histogramm berechnet: Seien zur Vereinfachung die Farbwerte als natürliche Zahlen aus dem abgeschlossenen Intervall $[0, G]$ angenommen, dann ist der Index im Histogramm eine n–stellige Zahl zur Basis G, wobei die i–te „Ziffer“ sich aus dem Farbwert des i–ten Pixels ergibt.
R. Rickman und J. Stonham benutzen nur den Buntton für die Berechnung des Histogramms. Da die alleinige Verwendung des Bunttons aber, wie in Abschnitt 4.1 dargestellt, zu größeren Ungenauigkeiten führt[40], wird in der Implementierung eine grobe Quantisierung des Farbenraums unter Verwendung aller Farbkanäle benutzt. Eine grobe Quantisierung ist erforderlich, weil die Histogrammgröße (Anzahl der Bins) bei gleicher Quantisierung deutlich größer ist als bei der Farbindizierung nach Swain/Ballard, siehe Abschnitt 5.3.1.
R. Rickman und J. Stonham verwenden in ihrer Arbeit 3–Tupel. Für 2–Tupel ergibt sich eine Nachbarschaft. Bei Verwendung von n–Tupeln (mit $n > 3$) entstehen sehr schnell Probleme mit der Histogrammgröße, da für eine handhabbare Histogrammgröße eine extrem grobe Quantisierung erforderlich wird.
In [34] werden nicht die Histogramme selbst gespeichert, sondern nur die Indizes der Bins, deren Wert einen bestimmten Schwellenwert überschreitet.

[40] R. Rickman und J. Stonham selbst betrachten die alleinige Verwendung des Bunttons nur als vorläufig.

Der Wert für die Übereinstimmung zweier Histogramme berechnet sich dann aus der Anzahl gleicher gespeicherter Indizes.
Um eine größere Robustheit gegen Rauschen zu erzielen, schlagen R. Rickman und J. Stonham eine Medianfilterung des Bildes vor der Berechnung des Histogramms vor.

5.2 Weiterentwicklungen und Kombinationen existierender Verfahren

5.2.1 Farbmodelle und Quantisierung

Die Autoren der in Abschnitt 5.1 dargestellten Verfahren haben die Experimente zumeist nur in einem Farbmodell durchgeführt. Es ist jedoch festzustellen, daß die Ergebnisse oft entscheidend vom verwendeten Farbmodell abhängen. Das liegt inbesondere an der unterschiedlich starken Korrelation der einzelnen Farbkanäle und den verschiedenartigen Möglichkeiten der Gewichtung[41].
Ein weiterer Einflußfaktor ist die Feinheit der Quantisierung bei Merkmalen, bei denen eine Berechnung mit der ursprünglichen Quantisierung zu speicherplatzaufwendig ist. Das betrifft alle Verfahren, die Histogramme benutzen. Auch bei Farbmodellen mit anscheinend gleichwertigen Kanälen, wie z.B. RGB oder der Rot–Grün– und der Blau–Gelb–Kanal im Gegenfarbenmodell, ist eine unterschiedliche Quantisierung unter Umständen von Vorteil, da die Information zur Unterscheidung der Bilder nicht gleichmäßig über den gesamten Farbenraum verteilt sein muß. Bei Landschaftsaufnahmen ist z.B. eine stärkere Relevanz des gelben und grünen Spektralbereichs zu erwarten.

5.2.2 Kachelung des Bildes

Ein einfacheres Verfahren zur Integration räumlicher Information ist die gleichmäßige Einteilung des Bildes in rechteckige Regionen (gleichmäßige Kachelung, *tessalation*). Diese Regionen sind einfacher zu berechnen als die in Abschnitt 5.1.5 dargestellten. Sie haben jedoch den Nachteil, daß bei einer Verschiebung von Bildteilen im Bereich der Regionengrenzen sich die berechneten Merkmale von mindestens zwei Regionen deutlich ändern.
Bei der Anwendung des vorgestellten Objekterkennungssystems können solche lokalen Verschiebungen nur bei lokalen Deformationen der zu erkennenden Objekte auftreten. D.h., scharf begrenzte Regionen sollten erwartungsgemäß in diesem Fall (und nur in diesem Fall) etwas schlechtere Ergebnisse liefern. Dagegen steht jedoch eine erhebliche Einsparung an Rechenzeit bei der Bestimmung der Regionenzugehörigkeit eines Punktes. Bei geeigneter Wahl der Einteilung und geeigneter Bildgröße[42] kann die Region sogar ausschließlich mit Bitoperationen aus den Koordinaten bestimmt werden.

[41] Im RGB–Modell läßt sich z.B. die Intensität nicht gewichten
[42] jeweils Zweierpotenzen

5.2.3 Farbindizierung mit räumlicher Information

Die Farbindizierung mittels Histogrammen ist bzgl. der Rechenzeit den anderen vorgestellten Verfahren überlegen. Die Berechnung der ersten drei Momente ist wesentlich rechenintensiver, siehe Abschnitt 5.3.2. Im Gegensatz dazu steht aber die geringe Unterscheidungsfähigkeit der Farbindizierung: Der bei der Berechnung des Histogramms über das gesamte Bild eintretende Verlust an Information über die räumliche Verteilung der Farben ist zu groß. Es liegt daher nahe, die Unterscheidungsfähigkeit der Farbindizierung durch die Integration räumlicher Information zu verbessern.
Eine einfache und effiziente Methode ist die oben beschriebene Kachelung. Es wird aber auch die Einteilung in die in Abschnitt 5.1.5 beschriebenen unscharfen Regionen untersucht, da zu erwarten ist, daß die Einteilung in eine zentrale Region und Randregionen toleranter gegenüber Ungenauigkeiten in der Rekonstruktion der Flächengrenzen ist.

5.2.4 Erweiterung der Tupel–Histogramme

Neben der Einbeziehung von Sättigung und Intensität bietet sich auch die Untersuchung anderer Farbmodelle an. Ein Problem hierbei ist jedoch wieder die mit feinerer Quantisierung exponentiell zunehmende Größe des Histogramms.
Statt der Schwellenwert–Methode von R. Rickman und J. Stonham wird in den Tests die Histogrammschnitt–Methode verwendet. Der Rechenaufwand erhöht sich dabei kaum. Dafür ist jedoch keine Einstellung eines Schwellenwertes nötig. Weiterhin ist eine höhere Genauigkeit zu erwarten, da beim Vergleich der einzelnen Bins ihr tatsächlicher Wert verglichen wird und nicht nur, ob beide über einem Schwellenwert liegen. Nachteilig ist allerdings, daß deutlich mehr Daten zu speichern sind.

5.2.5 Vergleich von Invarianten von Farbhistogrammen

Der von G. Healey und D. Slater vorgeschlagene Vergleich der Eigenwerte mit Hilfe des euklidischen Abstands bevorzugt große Eigenwerte: Die Eigenwerte haben typischerweise ein Verhältnis von 1:15 zwischen dem größten und kleinsten Eigenwert. Eine kleine prozentuale Abweichung im größten Eigenwert beeinflußt den euklidischen Abstand deutlich stärker als eine prozentual gleichgroße Abweichung im kleinsten Eigenwert. Daher wird eine Ähnlichkeitsfunktion definiert, die die Verhältnisse der Eigenwerte verwendet.
Für je zwei zu vergleichende Eigenwerte wird das Verhältnis des kleineren zum größeren der beiden Eigenwerte berechnet. Die Verhältnisse werden aufaddiert und durch die Anzahl der Eigenwerte dividiert.

Sei n die Anzahl der Eigenwerte, $v_{j,i}$ der i-te Eigenwert zum j-ten Bild ($j \in \{1,2\}$) und v_j der Vektor der sortierten Eigenwerte $v_{j,i}$.

$$d(v_1, v_2) = \frac{1}{n} \sum_{i=1}^{n} R(v_{1,i}, v_{2,i}) \tag{49}$$

mit

$$R(v_{1,i}, v_{2,i}) = \begin{cases} v_{1,i}/v_{2,i} & \text{für } v_{1,i} < v_{2,i} \\ v_{2,i}/v_{1,i} & \text{sonst} \end{cases}$$

5.2.6 Dynamische Buntton–Quantisierung

Das Hauptproblem bei der Verwendung des HSI–Farbmodells ist der Fehler bei der Berechnung des Bunttons bei wenig gesättigten Farben. Es liegt daher nahe, den Buntton dynamisch in Abhängigkeit von der Sättigung zu quantisieren.
Zwei mögliche Verfahren zur dynamischen Buntton–Quantisierung wurden realisiert: Beim ersten Verfahren wird die Anzahl der für den Bunttonkanal zur Verfügung stehenden Bits mit der Sättigung skaliert.

Sei S die Sättigung, $S \in [0,1]$, dann ergibt sich die Anzahl der Bits für den Bunttonkanal zu:

$$\mathrm{Bits_H} = \mathrm{ROUND}(S(8-n)+n) \qquad (n \in \{0,\ldots,8\}) \tag{50}$$

Es stehen maximal acht Bit zur Verfügung. n ist ein Parameter des Verfahrens. Der Wert 8 für n ergibt keine Quantisierung, der Wert 0 führt bei hinreichend kleiner Sättigung ($S < 1/16$) zum völligen Wegfall des Bunttons (0 Bit).
Beim zweiten Verfahren entstehen mehr Stufen der Feinheit der Quantisierung.

$$H = \begin{cases} \lfloor H \cdot S \rfloor / S & \text{für } S \neq 0 \\ 0 & \text{für } S = 0 \end{cases} \tag{51}$$

($\lfloor x \rfloor$ bezeichnet die größte ganze Zahl g mit $g \leq x$.)

5.3 Vergleichende theoretische Betrachtungen zum Aufwand der Verfahren hinsichtlich Zeit- und Speicherplatzbedarf

5.3.1 Speicherplatz

Die Speicherung von Histogrammen unter Verwendung der vollen Farbauflösung (in der Regel 256 Graustufen in jedem der drei Farbkanäle, also ca. 16,7 Mio. verschiedene Farben) erfordert zuviel Speicherplatz. Eine gröbere Quantisierung des Farbenraums führt zu einer drastischen Verringerung des Speicherplatzbedarfs der Histogramme. Mit ihr einher gehen aber auch Ungenauigkeiten, falls — wie beim Verfahren von G. Healey und D. Slater (Abschnitt 5.1.4) — Merkmale aus den Histogrammen berechnet werden. Eine Reduktion des Speicherplatzbedarfs der Histogramme ist auch ohne eine gröbere Quantisierung des Farbenraums erzielbar, da die Histogramme nur sehr dünn besetzt sind. Die Anzahl der Bins ungleich null ist immer kleiner oder gleich der Anzahl der Pixel im Bild, also meist unter 300 000. Durch entsprechende Speicherstrukturen (z.B. Hash) kann eine kompakte Speicherung mit effizientem Zugriff erzielt werden[43].
Falls Histogramme direkt verglichen werden, kann bei Verwendung einer Schwellenwertmethode der Speicherplatz weiter verringert werden, da Bins mit einem Wert unterhalb des Schwellenwertes nicht berücksichtig werden und somit auch nicht gespeichert werden müssen. R. Rickman und J. Stonham gehen bei ihrem Verfahren (Abschnitt 5.1.6) noch einen Schritt weiter und speichern nur die Indizes der Bins mit Werten oberhalb des Schwellenwertes. So eine starke Reduktion ist bei diesem Verfahren auch notwendig, da die Anzahl der Bins gegenüber einem einfachen Histogramm mit der Anzahl der Punkte im Tupel potenziert wird.
Einen sehr geringen Speicherplatzbedarf bei gleichzeitig einfachen Datenstrukturen und somit effizientem Zugriff haben das Verfahren von M. Stricker und A. Dimai (Abschnitt 5.1.3) und, wenn man nur die letztendlich berechneten Merkmale betrachtet, auch das Verfahren von G. Healey und D. Slater (Abschnitt 5.1.4). Bei beiden werden nur neun bzw. sechs reelle Zahlen gespeichert. Beim Verfahren von M. Stricker und A. Dimai sind dabei weder Quantisierungen noch Zwischenwerte mit größerem Speicherplatzbedarf erforderlich.
Bei der Integration räumlicher Information vervielfacht sich der Speicherplatzbedarf aller Verfahren linear mit der Anzahl der Regionen.

5.3.2 Rechenzeit

Den geringsten Rechenzeitbedarf haben die einfachen histogrammbasierten Verfahren. Zur Bestimmmung des dem Farbwert des Pixels entsprechenden Histogramm–Bins sind (bei geeigneter Quantisierung des Farbenraums)

[43] Z.B. haben G. Healey und D. Slater bei der Implementierung ihres Verfahrens die Histogramme mit Hilfe eines Hashs in voller Auflösung verarbeitet.

lediglich Bitmanipulationen erforderlich, und der Vergleich der Histogramme umfaßt relativ wenige Rechenoperationen. Die einzelnen Metriken und der Histogrammschnitt unterscheiden sich nur wenig im Aufwand. Lediglich die Berechnung des Summenhistogramms verursacht einen merklichen zusätzlichen Aufwand.
Die Berechnung der Momente ist deutlich aufwendiger als die Berechnung der Histogramme, da hier für jeden Bildpunkt mehrere Additionen und Multiplikationen erforderlich sind.
Die Berechnung von Invarianten nach G. Healey und D. Slater ist das vom Rechenzeitbedarf gesehen aufwendigste der hier vorgestellten Verfahren ohne Berücksichtigung räumlicher Information. Es wird zweimal ein Histogramm berechnet aus dem jeweils Matrizen berechnet werden. Dabei wird eine Matrizenaddition pro Histogramm–Bin ausgeführt. Die anschließenden Matrixoperationen (Invertierung, Cholesky–Zerlegung) sind aufgrund der geringen Größe der Matrizen (3×3 bzw. 6×6) nicht sehr aufwendig. Lediglich die Bestimmung der Eigenwerte verursacht noch einen größeren Aufwand. Da es sich jedoch um reellwertige symmetrische positiv definite Matrizen handelt, ist ein optimiertes Verfahren anwendbar.
Bei der Berücksichtigung räumlicher Information erhöht sich der Aufwand entsprechend der Anzahl der Regionen. Für jede Region ist der Aufwand so hoch wie für ein Bild mit der gleichen Anzahl von Pixeln. Hinzu kommt noch der Aufwand für die Bestimmung der Regionenzugehörigkeit des jeweiligen Pixels. Bei einer gleichmäßigen Kachelung reicht dafür eine Division pro Koordinate[44]. Für die Bestimmung der unscharfen Region nach [42] ist ein deutlich höherer Aufwand erforderlich. Allerdings werden die Regionenzugehörigkeiten abgespeichert und brauchen für mehrere gleichgroße Bilder nur einmal berechnet werden.
Der Aufwand zur Berechnung der Tupel–Histogramme ist etwas kleiner als das n–fache des Aufwandes der Berechnung eines einfachen Histogramms mit derselben Anzahl von „Proben“, wobei n die Tupel–Größe ist. Allerdings werden für die Tupel–Histogramme deutlich weniger „Proben“ gezogen als Pixel im Bild sind, so daß der Rechenzeitaufwand insgesamt nur etwas höher oder auch geringer ausfällt, je nach Anzahl der Punkte pro Tupel.

[44]bei Zweierpotenzen genügen sogar Bitmanipulationen

6 Das Objekterkennungssystem PolyRec

Nachdem der Ablauf der Objekterkennung und damit auch die Struktur des realisierten Objekterkennungssytems in Abschnitt 1.3 bereits kurz dargestellt worden ist, sollen in diesem Abschnitt das Objekterkennungssytem und seine Komponenten ausführlich vorgestellt werden.
PolyRec ist ein modellbasiertes Objekterkennungssystem mit einer datengesteuerten Zuordnungsstrategie. Es gliedert sich in die Komponenten

- 2.5D–Rekonstruktion (geometrische Rekonstruktion)
- Rekonstruktion der Aufdrucke (Texturen)
- Vergleich der rekonstruierten Aufdrucke mit den Bildern der Datenbank (Matching)
- Identifizierung des bearbeiteten Objekts anhand der Ergebnisse des Matchings (Voting)

Das Objekterkennungssytem enthält keine Rückkopplungen, siehe auch Abschnitt 8.3 im Ausblick.
Die geometrische Rekonstruktion ist bereits in [29] dargestellt und wird hier nicht weiter betrachtet. Sie bestimmt die Koordinaten der Eckpunkte der einzelnen sichtbaren Flächen des Objekts im projektiven Abbild sowie die Orientierungen bzw. Normalen der Flächen. Mit diesen Daten kann dann die Rekonstruktion der Aufdrucke erfolgen. Die rekonstruierten Aufdrucke der einzelnen Flächen werden mit den Bildern der Datenbank verglichen. Anschließend wird aus den Ergebnissen des Vergleichs eine Rangliste der erkannten Objekte bestimmt.
Für das Objekterkennungssytem gelten folgende Einschränkungen: Es darf sich nur ein Objekt in der Szene befinden. Das Objekt muß annähernd polyedrisch sein[45]. Die Beleuchtung muß möglichst gleichmäßig und ungerichtet sein. Für die 2.5D–Rekonstruktion mittels des verwendeten Photometric–Stereo–Verfahrens sind zusätzlich noch drei gerichtete Lichtquellen erforderlich und Lambert–Reflexion für das Objekt anzunehmen.

6.1 Rekonstruktion der Aufdrucke

Die Rekonstruktion der Aufdrucke der einzelnen Flächen des Objekts ist die in Abschnitt 2.2.1 dargestellte Erzeugung der RIIs, wobei jedoch, wie in Abschnitt 2.2.2 dargelegt, nur die jeweilige Basifläche betrachtet wird.
Zur Erzeugung eines RII ist eine Rotation im 3D–Raum zu finden, so daß der Vektor $\lambda(p, q, -1)^{\mathrm{T}}$ (mit $\lambda = 1/\sqrt{p^2 + q^2 + 1}$) (Normale der Basifläche im Kamerakoordinatensystem) auf den Vektor $(0, 0, -1)^{\mathrm{T}}$ abgebildet wird.

[45] Erweiterungen für runde Objekte wurden bereits diskutiert.

Eine solche Drehung existiert immer und ist eindeutig definiert. Diese 3D–Rotation entspricht einer affinen Transformation des zweidimensionalen Abbilds der Basisfläche.
Durch diese affine Transformation (im folgenden auch Entzerrung genannt) wird die Lage und Größe der Basisfläche im RII nicht eindeutig bestimmt. Die Basisfläche im RII ist (bis auf Fehler bei der 2.5D–Rekonstruktion) mathematisch ähnlich mit dem Abbild der Basisfläche in der Datenbank und unterscheidet sich von diesem nur durch Skalierung, Translation und Rotation (2D).
Je nach Art der zu berechnenden Bildmerkmale, ist eventuell eine paßgenaue Überlagerung der rekonstruierten Fläche mit der Fläche der Datenbank erforderlich. Dazu dient das von K. Oppermann und W.J. Schimke in [29] bereits beschriebene Verfahren, bei dem die längste Kante der Basisfläche im RII parallel zur längsten Kante der Datenbankfläche ausgerichtet wird. Damit reduziert sich das Matching auf genau die Anzahl von Fällen, wie es gleichlange längste Kanten gibt. Bei einer rechteckigen Fläche sind es zwei Fälle, wie die Abb. 17 zeigt. Tatsächlich kann sich die Anzahl der Fälle erhöhen, da aus Gründen der Robustheit des Verfahrens alle diejenigen Kanten als längste Kanten aufgefaßt werden, die annähernd so lang sind wie die absolut längste Kante. Dadurch werden leichte Fehler bei der geometrischen Rekonstruktion ausgeglichen.

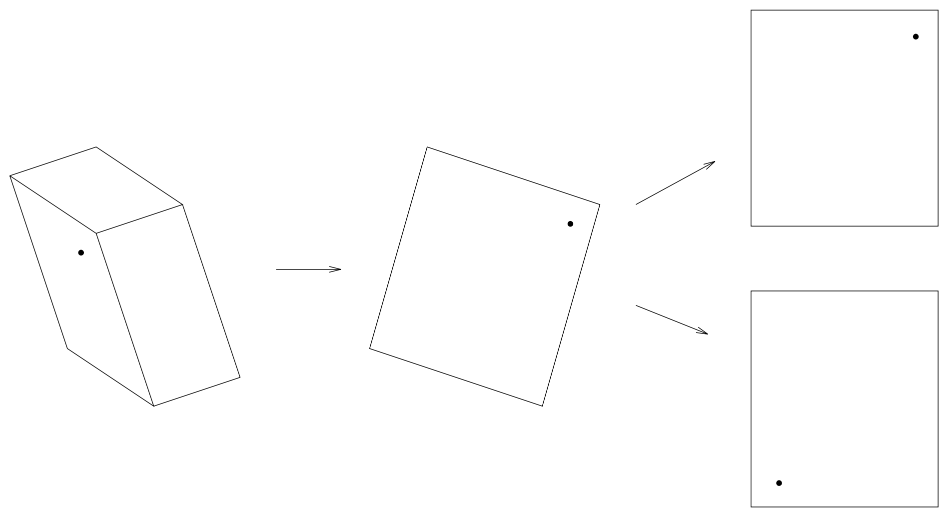

Abbildung 17: Entzerrung und Ausrichtung der Flächen

Eine Alternative zu diesem Verfahren ist die Ausrichtung entlang der Hauptachse der Fläche. Dabei wird der Schwerpunkt der Fläche in den Bildmittelpunkt verschoben und die Hauptachse parallel zur y–Achse ausgerichtet. Es sind dann nur noch zwei Fälle zu betrachten, die sich durch eine Drehung um 180° um den Bildmittelpunkt unterscheiden. Da der Rechenaufwand zur

Bestimmmung der Hauptachse erheblich höher ist als die Bestimmung der längsten Kante, wurde dieses Verfahren nicht realisiert[46].
Bei beiden Verfahren muß noch eine abschließende Skalierung von RII und SNI auf eine gemeinsame Größe erfolgen. Sinnvollerweise wird das RII auf die Größe der Datenbankfläche (SNI) skaliert. Wird bei der Berechnung der Bildmerkmale räumliche Information nur relativ zur Bildgröße berücksichtigt, so kann die Skalierung entfallen. Die beschriebenen Matching–Verfahren wurden so realisiert.
Bei allen Verfahren zur Ausrichtung wird vorausgesetzt, daß die SNIs, also die Bilder der Datenbank, bereits ausgerichtete Flächen enthalten.

Verzicht auf den Ausgleich der perspektivischen Verzerrung

Für Matching–Verfahren, die räumliche Information benutzen, ist ein Verzicht auf die Entzerrung der Flächen nicht möglich. Wird hingegen nur die Verteilung der Farben im Bild zum Matching herangezogen (Histogramme, Momente), so können die Merkmale näherungsweise auch von der mit perspektivischen Verzerrungen behafteten aufgenommenen Fläche berechnet werden. Auf diese Weise kann der Rechenaufwand für die affinen Transformationen eingespart werden. Allerdings sind auch größere Fehler zu erwarten, da die verzerrte Fläche aufgrund von Digitalisierungseffekten weniger Information enthält als die Datenbankfläche.

Alternative: Affine Transformation der Bilder der Datenbank

Die mit perspektivischer Verzerrung aufgenommenen Flächen enthalten (aufgrund der geringeren Pixelanzahl) weniger (Farb-)Information als die Flächen der Datenbank. Statt die aufgenommenen Flächen zu entzerren und die fehlenden Pixel zu interpolieren, können auch die Flächen aus der Datenbank jeweils entsprechend der Lage der aufgenommenen Fläche verzerrt werden. Der Informationsgehalt entspricht jetzt genau dem der verzerrten Aufnahme. Insofern ist ein besseres Ergebnis zu erwarten. Allerdings steigt auch der Rechenaufwand erheblich, denn statt einer geringen Anzahl von rekonstruierten Flächen muß nun die relativ große Anzahl von Flächen aus der Datenbank einer affinen Transformation unterzogen werden. Für diese Flächen müssen dann jeweils die zu vergleichenden Merkmale berechnet werden. Eine Vorab–Berechnung der Merkmale für die Bilder der Datenbank wäre nicht mehr möglich, so daß dieses Verfahren nur bei Verwendung von einfach und schnell zu berechnenden Merkmalen in Betracht zu ziehen ist.
Das Problem der Ausrichtung stellt sich hier ebenso. Eine Ausrichtung an der längsten Kante ist genauso vorstellbar. In Anbetracht der Abweichungen bei der Berechnung der Orientierungen der Flächen bei der geometrischen

[46] Das Verfahren unter Verwendung der längsten Kante erzeugt nur selten mehr als zwei Fälle, so daß kein höherer Aufwand für Transformationen entsteht.

Rekonstruktion ist eventuell die oben dargestellte Ausrichtung anhand von Schwerpunkt und Hauptachse günstiger.

6.2 Vergleich der rekonstruierten Aufdrucke mit den Bildern der Datenbank

Beim Matching ist für jede Paarung einer rekonstruierten Fläche mit einer Fläche aus der Datenbank ein Wert für die Übereinstimmung zu berechnen. Dazu werden die Verfahren aus Abschnitt 5 verwendet. Für die beiden Bilder werden die jeweiligen Merkmale berechnet und anschließend verglichen.

6.2.1 Skalierung der Werte der Ähnlichkeitsfunktionen

Um die verschiedenen Verfahren leichter vergleichen zu können, werden die Vergleichswerte auf das Intervall $[0,1]$ umgerechnet, wobei 0 keine Übereinstimmung und 1 die maximale Übereinstimmung bedeutet. Dabei ist zu beachten, daß die meisten Matchingverfahren den Wert für die maximale Übereinstimmung auch für nicht vollkommen inhaltsgleiche Bilder liefern. Die Skalierung der Werte im Intervall $[0,1]$ (im folgenden Match–Werte) ist vom jeweiligen Verfahren abhängig, so daß die Werte unterschiedlicher Verfahren nicht direkt miteinander verglichen werden können.
Der Histogrammschnitt liefert bereits Werte im Intervall $[0,1]$. Verfahren, die Abstände zwischen Merkmalen berechnen, liefern Werte im Bereich $[0,\infty)$, wobei 0 die maximale Übereinstimmung repräsentiert. Die Umrechnung auf das Intervall $[0,1]$ erfolgt mit folgender Formel

$$v = \frac{1}{\sqrt[n]{d+1}} \tag{52}$$

wobei d der berechnete Abstand, v der Matchwert und n ein Parameter ist. Geeignete Werte für den Parameter n sind 2 und 3.

6.2.2 Vorauswahl der zu vergleichenden Flächen

Da ein Vergleich aller rekonstruierten Flächen mit allen Flächen der Objekte der Datenbank zu zeitaufwendig ist, muß eine Vorauswahl geeigneter rekonstruierter Flächen getroffen werden und für jede rekonstruierte Fläche die Menge der Kandidaten aus der Menge der Flächen der Datenbank bestimmt werden (siehe auch [29]).
Kleine rekonstruierte Flächen tragen wenig zum Erkennungsprozeß bei und können ihn sogar negativ beeinflussen, da sie wenig Information enthalten und ihre geometrische Rekonstruktion besonders fehleranfällig ist. Daher werden rekonstruierte Flächen, die kleiner als ein bestimmter wählbarer Bruchteil der größten rekonstruierten Fläche sind, von vornherein vom Matching ausgeschlossen.

Als weiteres Kriterium wurde das sogenannte Edge–Area–Ratio, das Verhältnis zwischen längster Kante und Wurzel des Flächeninhalts der Fläche, gewählt. Es ist ein Maß für die „Länglichkeit“ der Fläche. Weichen die Edge–Area–Ratios einer rekonstruierten Fläche und einer Fläche der Datenbank um mehr als einen bestimmten Faktor voneinander ab, so wird diese Paarung von Flächen nicht weiter betrachtet, da sie offensichtlich nicht übereinstimmen können.
Eine Vorauswahl anhand des durchschnittlichen oder des häufigsten Bunttons, wie K. Oppermann und W.J. Schimke sie in [29] beschrieben haben, wurde nicht mehr benutzt, da die Berechnung des Bunttons (sofern nicht das HSI–Farbmodell benutzt wird) relativ aufwendig ist und die erzielte Reduktion zu vergleichender Flächen bei der getesteten Datenbank zu gering war. Bei Tests mit großen Datenbanken sollte dieses Vorauswahl–Kriterium jedoch wieder in Betracht gezogen werden.

6.3 Objektidentifizierung

Die beim Matching erhaltenen Werte beschreiben die Übereinstimmung auf der Ebene der Flächen. Beim Voting wird daraus die Übereinstimmung auf der Ebene der Objekte berechnet. Als Ergebnis erhält man eine Rangliste der Objekte, geordnet nach abfallender Übereinstimmung mit dem zu erkennenden Objekt.
Für das Voting werden zunächst die vom Matching erhaltenen Werte mit einem von der beteiligten rekonstruierten Fläche abhängigen Gewicht multipliziert[47]. Das Gewicht setzt sich aus einem unabhängigen Anteil und aus von der Größe und von der Orientierung der rekonstruierten Fläche abhängigen Anteilen zusammen. Die Matching–Werte werden umso schwächer gewichtet, je kleiner die beteiligte rekonstruierte Fläche ist bzw. je stärker die perspektivische Verzerrung des Aufdrucks bei der Aufnahme war, da bei kleinen Flächen und bei starker perspektivischer Verzerrung größere Fehler zu erwarten sind. Ein Schwellenwert legt die relative Größe einer rekonstruierten Fläche fest, bei deren Unterschreitung das Gewicht für diese Fläche auf null gesetzt und diese rekonstruierte Fläche somit ausgeschlossen wird. Aus Effizienzgründen geschieht dies bereits vor dem Matching. Mit einer kontinuierlichen Berücksichtigung der Größe der rekonstruierten Fläche bei der Gewichtung wird ein ähnlicher Effekt wie beim Ausschluß kleiner Flächen erzielt, jedoch kann auch der dann vorhandene geringe Einfluß kleiner Flächen noch stören. Außerdem müßten mehr Bilder verglichen werden.
In der vorliegenden Implementierung werden zu Testzwecken fünf verschiedene Votings parallel verwendet. In der späteren Anwendung sollte man sich auf ein oder zwei Votings beschränken. Bei der Verwendung von mehr

[47] Details zur Gewichtung siehe [29]

als einem Voting wäre die Behandlung von Widersprüchen in der Plazierung der Objekte zu definieren.
Die einzelnen Voting–Verfahren unterscheiden sich nur in der Art der Zusammenfassung der gewichteten Match–Werte. Die verschiedenen Votings wurden bereits in [29] vorgestellt. Sie werden hier nur der Vollständigkeit halber nochmals dargestellt.
Die Votings 0 und 1 werden als „Maximum Votings“ bezeichnet. Dabei wird pro rekonstruierter Fläche das Bild aus der Datenbank mit der besten Übereinstimmung, also dem höchsten Match–Wert bestimmt.
Beim Voting 0 wird der gewichtete Match–Wert zu den Werten aller zur Datenbank–Fläche gehörenden Objekte[48] addiert und ein Zähler inkrementiert. Am Ende des Votings werden die Werte für jedes Objekt durch den Wert des Zählers geteilt.
Beim Voting 1 wird für jedes zur Datenbank–Fläche gehörende Objekt das Maximum aus dem bisherigen Wert und dem gewichteten Match–Wert eingetragen. Somit erhält das bzw. die Objekt(e), zu dem bzw. denen das Datenbank–Bild mit dem besten Match–Wert paßt, den höchsten Wert.
Die Votings 2, 3 und 4 sind sogenannte „Selection Votings“. Für jedes Objekt werden dabei die n besten Paarungen von rekonstruierter und Datenbank–Fläche bestimmt, wobei n das Minimum aus der Anzahl der in der Datenbank erfaßten Flächen des Objekts und der Anzahl der rekonstruierten Flächen ist. Beim Voting 2 werden die gewichteten Match–Werte aufaddiert, beim Voting 3 werden die Kehrwerte der „globalen Platzierungen“ und beim Voting 4 die Kehrwerte der „Platzierungen“ für diese rekonstruierte Fläche aufaddiert. Anschließend werden die Werte durch n geteilt. Als „Platzierung“ wird die Anzahl größerer Match–Werte +1 definiert. Bei der „Platzierung“ für eine rekonstruierte Fläche werden dabei nur die Match–Werte der betreffenden rekonstruierten Fläche betrachtet.
Bei den Tests zu [29] haben sich Voting 2 und Voting 3 als sehr geeignet erwiesen.

[48] Normalerweise ist eine Datenbank–Fläche nur einem Objekt zugeordnet. Um aber gleichartige Flächen verschiedener Objekte in der Datenbank adäquat zu repräsentieren, wird dann nur eine Aufnahme erstellt, die mehreren Objekten zugeordnet wird.

7 Ergebnisse des Gesamtsystems

7.1 Allgemeine Umgebung zur Beurteilung von Algorithmen zum Vergleich von Bildmerkmalen

Zum Test der verschiedenen Verfahren zur Bestimmung und zum Vergleich der Bildmerkmale wurde eine spezielle Umgebung entwickelt. Diese Umgebung wurde zur schnellen Vorselektion guter Parameter für die einzelnen Verfahren entworfen. Sie eignet sich auch zur allgemeinen Beurteilung von Verfahren, die Bilder anhand von (Farb-)Merkmalen vergleichen, insbesondere zum Vergleich von Algorithmen zur bildbasierten Abfrage von Bilddatenbanken.
Es wird eine Menge von Bildern eingelesen, mit einem wählbaren Algorithmus zur Farb- bzw. Helligkeitskonstanz behandelt und in das vorgegebene Farbmodell gewandelt. Anschließend werden die Bilder mittels des oder der anzugebenden Verfahren verglichen, wobei auch die Parameter des oder der Verfahren wählbar sind.

Auswahl der Parameter für die Tests

Für die Auswahl von Parametern für den anschließenden Test des Objekterkennungssystems wurden 24 repräsentative Bilder aus der Datenbank ausgewählt und jeweils vier abgewandelte Bilder berechnet. Die erste Abwandlung bestand in der Überlagerung mit normalverteiltem Rauschen (getrennt für jeden Farbkanal, Streuung 15 Grauwerte), die zweite Abwandlung ist eine Rotation um 5°, die dritte eine Farbverfälschung (Verstärkung des Blaukanals um 1.1) und die vierte eine Intensitätserhöhung (um den Faktor 1.3).
Für so erhaltenen 120 teilweise synthetischen Testbilder wurden dann Tests mit allen Verfahren mit jeweils einer Reihe verschiedener Parameter durchgeführt. Dabei wurde beurteilt, wie gut die Zuordnung der Originalbilder zu den entsprechenden abgewandelten Bildern und die Zuordnung der abgewandelten Bilder desselben Originalbildes untereinander ist. Weiterhin wurde die Schärfe der Abgrenzung von den anderen Bildern beurteilt. Als Maß diente das Minimum, das Maximum, das arithmetische Mittel und die Standardabweichung der Ähnlichkeitswerte für zusammengehörige Bilder und für nicht zusammengehörige Bilder.
Die Parameter wurden für jede Variante eines jeden Verfahrens getrennt bewertet. Die einzelnen Parametersätze wurden nach der Differenz der Mittelwerte der Ähnlichkeitswerte für zusammengehörige Bilder und für nicht zusammengehörige Bilder sortiert. Anschließend wurden manuell unter den Parametersätzen, auch unter Berücksichtigung der Standardabweichungen, aussichtsreiche Kandidaten für die Tests des Objekterkennungssystems ausgewählt. Für eine vollautomatische Auswahl geeigneter Parameter besteht noch Forschungsbedarf.

7.2 Vergleich der vorgestellten Verfahren zur Beleuchtungsunabhängigkeit und zum Vergleich von Bildern und Bildmerkmalen

Die Experimente wurden mit dem in Abschnitt 6 vorgestellten Objekterkennungssystem PolyRec durchgeführt. Die Datenbank enthielt 21 Objekte mit insgesamt 114 Bildern. Es wurden 18 Datensätze von 2.5D–Rekonstruktionen von 14 verschiedenen Objekten verwendet. In zwei Fällen unterschieden sich zwei Objekte im wesentlichen nur in ihrer Größe (`Valensina_Saft_Orange_0.75l` und `Valensina_Saft_Orange_0.2l` sowie `Reichelt_Vollmilch_1l` und `Reichelt_Vollmilch_2l`); die Seitenlängenverhältnisse und Aufdrucke weichen jeweils nur gering voneinander ab. Da die absolute Größe im Objekterkennungssystem nicht berücksichtigt wird, ist eine Unterscheidung der beiden Objektpaare nicht zu erwarten. Dies ist für die Anwendung bei der Sortierung von Verpackungen auch nicht erforderlich, da in beiden Fällen die Objekte aus demselben Material bestehen.

Allgemeine Parameter

Für die in Abschnitt 6.2.2 beschriebene Vorauswahl wurden folgende Parameter verwendet: Rekonstruierte Flächen, deren Größe weniger als die Hälfte der Größe der größten rekonstruierten Fläche beträgt, wurden nicht berücksichtigt (`SEGMENT_FRACTION_LIMIT` 2.0). Für das Edge–Area–Ratio (`EA_RATIO_LIMIT`) wurde der Wert 2.5 verwendet, d.h. Paarungen einer rekonstruierten Fläche mit einer Fläche der Datenbank werden nicht betrachtet, wenn sich die Seitenlängenverhältnisse um einen Faktor größer als 2.5 unterscheiden.
Die in Abschnitt 6.3 genannten Gewichte wurden wie folgt gewählt: Der konstante Anteil des Gewichts (`WEIGHT_CONST`) betrug 1.0, die Gewichtung der Flächengröße (`WEIGHT_AREA`) 0.5. Die perspektivische Verzerrung (`WEIGHT_PQ`) wurde nicht mit einbezogen (Gewicht 0.0). Die Multiplikatoren für die Gewichtung der Flächengröße und der perspektivischen Verzerrung (`WEIGHT_MULTIPLIER_AREA` und `WEIGHT_MULTIPLIER_PQ`) wurden nicht benutzt (Wert 1.0). Es wurde das Voting–Verfahren 2 (siehe Abschnitt 6.3) verwendet. Verfahren 3 erwies sich als weniger geeignet.
Bei der in Abschnitt 5.2.6 beschriebenen dynamischen Buntton–Quantisierung wurde die zweite Variante aufgrund ihrer feineren Stufung der Quantisierung verwendet.

Darstellung der Ergebnisse

In den ersten Spalten jeder Ergebnistabelle werden die jeweiligen Parameter angegeben. Dazu gehören das verwendete Farbmodell (FM) und das Verfahren zur Helligkeitskonstanz (HKV). Die Abkürzungen für die Farbmodelle sind: RGB, III (Gegenfarbenmodell), HSI und LAB (CIELAB).

Für die Verfahren zur Helligkeitskonstanz werden folgende Abkürzungen verwendet: N kein Verfahren, S Verwendung des rgb–Modells (siehe Abschnitt 2.1.3) und I Verwendung des Verfahrens mit Normierung der Intensität (siehe Abschnitt 4.1). Unter „Quantisierung“ wird die Quantisierung des Farbenraums in Bit angegeben. Ist nur eine Zahl genannt, so bezieht sie sich auf alle Kanäle, ansonsten beziehen sich die drei Zahlen auf die einzelnen Kanäle in der Reihenfolge wie sie im Namen des Farbmodells vorkommen.
Für jeden Parametersatz eines Verfahrens wird der durchschnittliche Rang $\overline{R}$ der Objekte, der Anteil der Objekte mit Rang von eins bis drei T_3 (in Prozent) sowie der minimale $R_{\min}$ und maximale Rang $R_{\max}$ angegeben. Der Rang ist dabei die Position des Testobjekts in der nach absteigender Übereinstimmung sortierten Liste der Objekte, die PolyRec als Ergebnis liefert.

Problemfälle

Die Objekte `Reichelt_FettarmeMilch_1l` und `Reichelt_Fettarme-Milch_2l` wurden besonders häufig schlecht erkannt (Rang ≥ 10), wobei das Objekt `Reichelt_FettarmeMilch_1l` nahezu doppelt so oft betroffen war. Diese beiden Objekte sind auch in nahezu allen Fällen die Ursache für den hohen Wert bei $R_{\max}$. In den meisten übrigen Fällen liegt die Ursache beim Objekt `Odol_75ml`. Die Objekte `Odol_75ml`, `Penny_FettarmeMilch_1l` und `Reichelt_Schlagsahne_0.5l` erhielten zwar wesentlich seltener einen Rang größer oder gleich zehn, wurden aber dennoch unterdurchschnittlich erkannt. Die Ursachen für die schlechte Erkennung dieser Objekte sind hauptsächlich in Fehlern der geometrsichen Rekonstruktion zu suchen, die die Begrenzungen der Flächen nicht korrekt extrahiert. Ein weiteres Problem ist die geringe Genauigkeit der ebenfalls bei der geometrsichen Rekonstruktion ermittelten Orientierungen der Flächen. Sie führt zu Fehlern bei der Entzerrung der Aufdrucke.

7.2.1 Farbhistogramme mit Histogrammschnitt und Metriken

Farbindizierung mittels Farbhistogrammen und Histogrammschnitt wurde an der TU Berlin bereits in einem Projekt untersucht [36]. Sie war auch bereits im Vorläufer des realisierten Objekterkennungssystems implementiert [29].
Im Rahmen dieser Arbeit wurden weitere Farbmodelle und andere Auflösungen untersucht. Es wurde auch eine deutlich andere Auswahl von Objekten verwendet.
Die Tabellen 2, 3, 4 und 5 zeigen die Ergebnisse unter Verwendung von Farbhistogrammen mit Histogrammschnitt in allen verwendeten Farbmodellen und Verfahren zur Helligkeitskonstanz (HKV) und mit verschiedenen Quantisierungen. Die besten Ergebnisse werden im RGB–Modell mit dem

FM	HKV	Quantisierung	$\overline{R}$	T_3	$R_{\min}$	$R_{\max}$
RGB	N	2-2-2	3.4	77.8%	1	19
RGB	N	2-3-2	3.9	72.2%	1	20
RGB	N	2-3-3	3.7	83.3%	1	21
RGB	N	3-2-3	4.4	66.7%	1	21
RGB	N	3-3-2	4.4	72.2%	1	20
RGB	N	3-3-3	4.2	77.8%	1	21
RGB	N	3-4-3	3.9	72.2%	1	21
RGB	N	3-4-4	4.1	72.2%	1	18
RGB	N	4-3-4	4.1	72.2%	1	21
RGB	N	4-4-3	4.1	72.2%	1	20
RGB	N	4-4-4	3.8	72.2%	1	19
RGB	S	2-2-2	1.8	88.9%	1	6
RGB	S	2-3-2	1.7	94.4%	1	6
RGB	S	2-3-3	1.7	94.4%	1	6
RGB	S	3-2-3	1.5	94.4%	1	4
RGB	S	3-3-2	1.5	94.4%	1	5
RGB	S	3-3-3	1.4	94.4%	1	4
RGB	S	3-4-3	1.4	100.0%	1	3
RGB	S	3-4-4	2.4	77.8%	1	6
RGB	S	4-3-4	2.2	77.8%	1	7
RGB	S	4-4-3	1.4	100.0%	1	3
RGB	S	4-4-4	2.2	77.8%	1	6
RGB	I	2-2-2	4.4	66.7%	1	20
RGB	I	2-3-2	4.5	66.7%	1	19
RGB	I	2-3-3	4.9	66.7%	1	20
RGB	I	3-2-3	4.5	72.2%	1	20
RGB	I	3-3-2	4.2	72.2%	1	21
RGB	I	3-3-3	4.4	72.2%	1	20
RGB	I	3-4-3	4.9	61.1%	1	20
RGB	I	3-4-4	5.1	61.1%	1	20
RGB	I	4-3-4	4.7	61.1%	1	20
RGB	I	4-4-3	4.6	61.1%	1	20
RGB	I	4-4-4	4.7	61.1%	1	20

Tabelle 2: Ergebnisse unter Verwendung von Farbhistogrammen mit Histogrammschnitt, RGB–Modell. Zur Erläuterung der Darstellung der Ergebnisse siehe Abschnitt 7.2.

FM	HKV	Quantisierung	$\overline{R}$	T_3	$R_{\min}$	$R_{\max}$
III	N	3-3-2	5.1	50.0%	1	20
III	N	3-3-3	4.9	61.1%	1	20
III	N	4-4-3	3.8	72.2%	1	20
III	N	4-4-4	3.9	72.2%	1	20
III	S	3-3-2	7.2	44.4%	1	16
III	S	3-3-3	7.2	44.4%	1	16
III	S	4-4-3	3.7	72.2%	1	12
III	S	4-4-4	3.7	72.2%	1	12
III	I	3-3-2	4.6	66.7%	1	20
III	I	3-3-3	4.4	55.6%	1	20
III	I	4-4-3	3.2	72.2%	1	18
III	I	4-4-4	4.0	66.7%	1	19

Tabelle 3: Ergebnisse unter Verwendung von Farbhistogrammen mit Histogrammschnitt, Gegenfarbenmodell

FM	HKV	Quantisierung	$\overline{R}$	T_3	$R_{\min}$	$R_{\max}$
HSI	N	3-2-2	4.7	61.1%	1	21
HSI	N	3-3-2	5.3	66.7%	1	20
HSI	N	3-3-3	5.6	61.1%	1	20
HSI	N	4-3-2	5.3	66.7%	1	20
HSI	N	4-3-3	5.6	61.1%	1	20
HSI	N	4-4-3	5.1	66.7%	1	20
HSI	N	4-4-4	4.3	66.7%	1	20
HSI	S	3-2-2	1.6	100.0%	1	3
HSI	S	3-3-2	2.2	88.9%	1	6
HSI	S	3-3-3	2.2	88.9%	1	6
HSI	S	4-3-2	2.2	83.3%	1	7
HSI	S	4-3-3	2.2	83.3%	1	7
HSI	S	4-4-3	3.4	72.2%	1	18
HSI	S	4-4-4	3.4	72.2%	1	18
HSI	I	3-2-2	4.5	66.7%	1	19
HSI	I	3-3-2	4.2	66.7%	1	20
HSI	I	3-3-3	5.9	55.6%	1	20
HSI	I	4-3-2	4.1	66.7%	1	20
HSI	I	4-3-3	5.9	55.6%	1	20
HSI	I	4-4-3	4.4	55.6%	1	20
HSI	I	4-4-4	4.7	66.7%	1	20

Tabelle 4: Ergebnisse unter Verwendung von Farbhistogrammen mit Histogrammschnitt, HSI–Modell

FM	HKV	Quantisierung	$\overline{R}$	T_3	$R_{\min}$	$R_{\max}$
LAB	N	2-3-3	6.0	44.4%	1	16
LAB	N	3-3-3	5.3	44.4%	1	18
LAB	N	3-4-4	5.2	61.1%	1	19
LAB	N	4-4-4	4.6	66.7%	1	19
LAB	S	2-3-3	8.1	38.9%	1	16
LAB	S	3-3-3	7.3	38.9%	1	16
LAB	S	3-4-4	5.9	44.4%	1	13
LAB	S	4-4-4	5.9	44.4%	1	13
LAB	I	2-3-3	6.5	33.3%	1	16
LAB	I	3-3-3	5.6	44.4%	1	18
LAB	I	3-4-4	5.4	50.0%	1	18
LAB	I	4-4-4	5.3	44.4%	1	14

Tabelle 5: Ergebnisse unter Verwendung von Farbhistogrammen mit Histogrammschnitt, CIELAB–Modell

HKV S erreicht. Dies ist insofern verwunderlich, als daß bei diesem HKV die Intensität nicht berücksichtigt wird. Das HKV I ist für das RGB–Modell hier ungeeignet, da sich die Ergebnisse gegenüber dem Verzicht auf HK leicht verschlechtern. Insgesamt sind die Ergebnisse im RGB–Modell im betrachteten Bereich relativ unabhängig von der Quantisierung. Dies widerspricht den Erkenntnissen in [29], wurde aber auch in [46] beschrieben. Sowohl dieser Effekt als auch das gute Abschneiden des HKV S sind auf die geringe Größe der Datenbank zurückzuführen.
Das HSI–Modell (Tabelle 4) erreicht etwas schlechtere Werte als das RGB–Modell. Die Aussagen zur Eignung der HKV lassen sich jedoch übertragen. Der Einfluß der Quantisierung ist lediglich beim HKV S deutlich zu verzeichnen, wobei die besten Ergebnisse mit der gröbsten Quantisierung erreicht werden. Verantwortlich dafür sind Ungenauigkeiten bei der Bestimmung von Buntton und Sättigung. Gegenüber den Ergebnissen in [29] ist jedoch eine deutliche Verbesserung aufgrund der dynamischen Buntton–Quantisierung erkennbar.
Im Gegenfarbenmodell sind die Ergebnisse bei feinerer Quantisierung deutlich besser als bei gröberer, aber insgesamt etwas schlechter als im HSI–Modell. Das HKV hat nur einen geringen Einfluß. Lediglich beim HKV S und grober Quantisierung ergeben sich sehr schlechte Werte. Das CIELAB–Modell erzielt insgesamt sehr schlechte Ergebnisse, wobei die feine Quantisierung ohne HKV noch am besten abschneidet.
Bei Verzicht auf die Entzerrung der rekonstruierten Flächen verändern sich die Ergebnisse kaum (Tabellen 6, 7, 8 und 9). Vereinzelt ist sogar eine leichte Verbesserung erkennbar. Die größeren Änderungen in T_3 erklären sich dadurch, daß der Rang einiger zu erkennender Objekte von vier auf drei sinkt. Insgesamt läßt sich somit feststellen, daß der durch die Verzerrung eintretende Informationsverlust nicht sehr hoch ist. Die Objektoberflächen weisen größere homogen gefärbte Teile auf, auf die sich die Erkennung stützt. Somit betrifft die Reduktion der Pixelanzahl durch die Verzerrung alle wesentlichen Farben zu gleichen Teilen, und die Charakteristika der Farbverteilung ändern sich kaum.
Die Tabellen 10, 11 und 12 zeigen die Ergebnisse bei Verwendung der drei Standardmetriken anstelle der Histogrammschnitt–Methode; bei den Tabellen 13, 14 und 15 wurden die Metriken auf Summenhistogramme angewandt. Für die Umrechnung der Metriken in das einheitliche Intervall nach Abschnitt 6.2.1 wurde der Parameter $n = 2.0$ gewählt.
Die L_1–Metrik liefert — erwartungsgemäß — ungefähr die gleichen Ergebnisse wie der Histogrammschnitt, da beide mathematisch eng verwandt sind. Es ist jedoch deutlicher erkennbar, daß die jeweils feinste Quantisierung etwas schlechtere Ergebnisse bringt als die gröberen Quantisierungen, mit Ausnahme der Experimente ohne HKV.
Mit der L_2– und der L_∞–Metrik werden deutlich schlechtere Ergebnisse erzielt als mit der L_1–Metrik, wobei die L_∞–Metrik im Durchschnitt noch

FM	HKV	Quantisierung	$\overline{R}$	T_3	$R_{\min}$	$R_{\max}$
RGB	N	2-2-2	3.5	77.8%	1	19
RGB	N	2-3-2	4.0	72.2%	1	20
RGB	N	2-3-3	3.7	83.3%	1	21
RGB	N	3-2-3	4.3	66.7%	1	21
RGB	N	3-3-2	4.6	72.2%	1	20
RGB	N	3-3-3	4.1	77.8%	1	21
RGB	N	3-4-3	4.1	72.2%	1	21
RGB	N	3-4-4	4.2	66.7%	1	19
RGB	N	4-3-4	4.1	72.2%	1	21
RGB	N	4-4-3	4.1	72.2%	1	20
RGB	N	4-4-4	2.8	56.0%	1	19
RGB	S	2-2-2	1.8	88.9%	1	6
RGB	S	2-3-2	1.6	94.4%	1	6
RGB	S	2-3-3	1.7	94.4%	1	6
RGB	S	3-2-3	1.6	94.4%	1	4
RGB	S	3-3-2	1.5	94.4%	1	5
RGB	S	3-3-3	1.5	94.4%	1	4
RGB	S	3-4-3	1.4	100.0%	1	3
RGB	S	3-4-4	2.6	72.2%	1	6
RGB	S	4-3-4	2.2	72.2%	1	5
RGB	S	4-4-3	1.3	100.0%	1	3
RGB	S	4-4-4	2.3	72.2%	1	6
RGB	I	2-2-2	4.2	72.2%	1	20
RGB	I	2-3-2	4.3	72.2%	1	19
RGB	I	2-3-3	4.6	72.2%	1	20
RGB	I	3-2-3	4.2	72.2%	1	20
RGB	I	3-3-2	4.2	72.2%	1	21
RGB	I	3-3-3	4.4	72.2%	1	20
RGB	I	3-4-3	4.5	61.1%	1	20
RGB	I	3-4-4	4.6	61.1%	1	20
RGB	I	4-3-4	4.3	61.1%	1	20
RGB	I	4-4-3	4.4	61.1%	1	20
RGB	I	4-4-4	4.3	66.7%	1	20

Tabelle 6: Ergebnisse unter Verwendung von Farbhistogrammen mit Histogrammschnitt ohne Entzerrung der rekonstruierten Aufdrucke, RGB–Modell

FM	HKV	Quantisierung	$\overline{R}$	T_3	$R_{\min}$	$R_{\max}$
III	N	3-3-2	5.2	50.0%	1	20
III	N	3-3-3	5.0	50.0%	1	20
III	N	4-4-3	3.9	66.7%	1	20
III	N	4-4-4	3.9	72.2%	1	20
III	S	3-3-2	7.2	44.4%	1	16
III	S	3-3-3	7.2	44.4%	1	16
III	S	4-4-3	3.7	72.2%	1	12
III	S	4-4-4	3.7	72.2%	1	12
III	I	3-3-2	4.4	66.7%	1	20
III	I	3-3-3	4.4	55.6%	1	20
III	I	4-4-3	3.2	72.2%	1	19
III	I	4-4-4	3.9	66.7%	1	19

Tabelle 7: Ergebnisse unter Verwendung von Farbhistogrammen mit Histogrammschnitt ohne Entzerrung der rekonstruierten Aufdrucke, Gegenfarbenmodell

FM	HKV	Quantisierung	$\overline{R}$	T_3	$R_{\min}$	$R_{\max}$
HSI	N	3-2-2	4.4	66.7%	1	21
HSI	N	3-3-2	5.2	61.1%	1	20
HSI	N	3-3-3	5.3	55.6%	1	20
HSI	N	4-3-2	5.2	61.1%	1	20
HSI	N	4-3-3	5.3	55.6%	1	20
HSI	N	4-4-3	5.1	66.7%	1	20
HSI	N	4-4-4	4.3	72.2%	1	20
HSI	S	3-2-2	1.3	100.0%	1	3
HSI	S	3-3-2	2.0	88.9%	1	5
HSI	S	3-3-3	2.0	88.9%	1	5
HSI	S	4-3-2	2.0	88.9%	1	7
HSI	S	4-3-3	2.0	88.9%	1	7
HSI	S	4-4-3	3.3	72.2%	1	14
HSI	S	4-4-4	3.3	72.2%	1	14
HSI	I	3-2-2	4.1	66.7%	1	19
HSI	I	3-3-2	3.6	61.8%	1	20
HSI	I	3-3-3	4.8	57.1%	1	20
HSI	I	4-3-2	3.9	61.1%	1	20
HSI	I	4-3-3	4.9	55.6%	1	20
HSI	I	4-4-3	4.3	55.6%	1	20
HSI	I	4-4-4	4.7	66.7%	1	20

Tabelle 8: Ergebnisse unter Verwendung von Farbhistogrammen mit Histogrammschnitt ohne Entzerrung der rekonstruierten Aufdrucke, HSI–Modell

FM	HKV	Quantisierung	$\overline{R}$	T_3	$R_{\min}$	$R_{\max}$
LAB	N	2-3-3	5.9	44.4%	1	16
LAB	N	3-3-3	5.0	48.6%	1	18
LAB	N	3-4-4	5.2	55.6%	1	19
LAB	N	4-4-4	4.6	61.1%	1	18
LAB	N	2-3-3	5.9	44.4%	1	16
LAB	S	3-3-3	7.5	37.1%	1	16
LAB	S	3-4-4	5.9	44.4%	1	13
LAB	S	4-4-4	5.9	44.4%	1	13
LAB	I	2-3-3	6.7	33.3%	1	16
LAB	I	3-3-3	5.5	44.4%	1	18
LAB	I	3-4-4	5.4	50.0%	1	18
LAB	I	4-4-4	5.3	44.4%	1	15

Tabelle 9: Ergebnisse unter Verwendung von Farbhistogrammen mit Histogrammschnitt ohne Entzerrung der rekonstruierten Aufdrucke, CIELAB–Modell

FM	HKV	Quantisierung	$\overline{R}$	T_3	$R_{\min}$	$R_{\max}$
RGB	N	2-2-2	3.3	77.8%	1	19
RGB	N	3-3-3	4.2	77.8%	1	21
RGB	N	4-4-4	3.8	72.2%	1	19
RGB	S	2-2-2	1.8	88.9%	1	6
RGB	S	3-3-3	1.4	94.4%	1	4
RGB	S	4-4-4	2.2	77.8%	1	6
RGB	I	2-2-2	3.8	72.2%	1	20
RGB	I	3-3-3	4.2	72.2%	1	20
RGB	I	4-4-4	4.4	66.7%	1	20
III	N	4-4-3	3.7	72.2%	1	20
III	N	4-4-4	3.8	72.2%	1	20
III	S	4-4-3	3.6	72.2%	1	12
III	S	4-4-4	3.6	72.2%	1	12
III	I	4-4-3	3.2	72.2%	1	18
III	I	4-4-4	4.1	61.1%	1	19
HSI	N	4-3-2	5.1	61.1%	1	20
HSI	N	4-3-3	5.3	61.1%	1	20
HSI	N	4-4-3	4.9	66.7%	1	20
HSI	N	4-4-4	4.5	72.2%	1	20
HSI	S	4-3-2	2.1	88.9%	1	7
HSI	S	4-3-3	2.1	88.9%	1	7
HSI	S	4-4-3	3.3	72.2%	1	18
HSI	S	4-4-4	3.3	72.2%	1	18
HSI	I	4-3-2	4.1	72.2%	1	20
HSI	I	4-3-3	6.0	55.6%	1	20
HSI	I	4-4-3	4.3	50.0%	1	20
HSI	I	4-4-4	4.8	61.1%	1	20
LAB	N	3-4-4	5.2	55.6%	1	19
LAB	N	4-4-4	4.6	66.7%	1	19
LAB	S	3-4-4	6.1	44.4%	1	16
LAB	S	4-4-4	6.1	44.4%	1	16
LAB	I	3-4-4	5.3	50.0%	1	15
LAB	I	4-4-4	5.1	44.4%	1	15

Tabelle 10: Ergebnisse unter Verwendung von Farbhistogrammen mit L_1–Metrik

FM	HKV	Quantisierung	$\overline{R}$	T_3	$R_{\min}$	$R_{\max}$
RGB	N	2-2-2	5.2	72.2%	1	19
RGB	N	3-3-3	7.2	50.0%	1	20
RGB	N	4-4-4	10.7	38.9%	1	21
RGB	S	2-2-2	2.0	88.9%	1	7
RGB	S	3-3-3	1.7	88.9%	1	5
RGB	S	4-4-4	4.9	66.7%	1	17
RGB	I	2-2-2	4.3	72.2%	1	18
RGB	I	3-3-3	6.1	55.6%	1	21
RGB	I	4-4-4	6.8	44.4%	1	21
III	N	4-4-3	8.7	44.4%	1	21
III	N	4-4-4	9.1	44.4%	1	21
III	S	4-4-3	8.0	38.9%	1	19
III	S	4-4-4	8.0	38.9%	1	19
III	I	4-4-3	8.4	50.0%	1	21
III	I	4-4-4	10.1	33.3%	1	20
HSI	N	4-3-2	7.8	50.0%	1	20
HSI	N	4-3-3	8.7	44.4%	1	21
HSI	N	4-4-3	10.1	33.3%	1	21
HSI	N	4-4-4	10.3	33.3%	1	21
HSI	S	4-3-2	3.4	72.2%	1	14
HSI	S	4-3-3	3.4	72.2%	1	14
HSI	S	4-4-3	6.6	44.4%	1	20
HSI	S	4-4-4	6.6	44.4%	1	20
HSI	I	4-3-2	6.1	55.6%	1	19
HSI	I	4-3-3	7.8	50.0%	1	20
HSI	I	4-4-3	9.2	38.9%	1	19
HSI	I	4-4-4	10.5	38.9%	1	21
LAB	N	3-4-4	8.8	38.9%	1	20
LAB	N	4-4-4	9.3	38.9%	1	21
LAB	S	3-4-4	7.7	38.9%	1	18
LAB	S	4-4-4	7.8	38.9%	1	19
LAB	I	3-4-4	8.3	44.4%	1	18
LAB	I	4-4-4	9.6	33.3%	1	19

Tabelle 11: Ergebnisse unter Verwendung von Farbhistogrammen mit L_2–Metrik

FM	HKV	Quantisierung	$\overline{R}$	T_3	$R_{\min}$	$R_{\max}$
RGB	N	2-2-2	6.6	66.7%	1	19
RGB	N	3-3-3	8.8	38.9%	1	21
RGB	N	4-4-4	10.9	33.3%	1	21
RGB	S	2-2-2	2.0	83.3%	1	6
RGB	S	3-3-3	1.9	83.3%	1	5
RGB	S	4-4-4	5.2	50.0%	1	16
RGB	I	2-2-2	4.8	55.6%	1	19
RGB	I	3-3-3	6.7	38.9%	1	21
RGB	I	4-4-4	7.7	33.3%	1	21
III	N	4-4-3	8.7	44.4%	1	21
III	N	4-4-4	9.1	44.4%	1	21
III	S	4-4-3	8.2	38.9%	1	17
III	S	4-4-4	8.2	38.9%	1	17
III	I	4-4-3	9.0	38.9%	1	21
III	I	4-4-4	9.2	38.9%	1	19
HSI	N	4-3-2	7.4	55.6%	1	21
HSI	N	4-3-3	8.8	38.9%	1	21
HSI	N	4-4-3	8.7	44.4%	1	21
HSI	N	4-4-4	8.9	38.9%	1	21
HSI	S	4-3-2	2.8	83.3%	1	12
HSI	S	4-3-3	2.8	83.3%	1	12
HSI	S	4-4-3	6.5	44.4%	1	20
HSI	S	4-4-4	6.5	44.4%	1	20
HSI	I	4-3-2	6.2	50.0%	1	17
HSI	I	4-3-3	8.3	44.4%	1	21
HSI	I	4-4-3	8.7	50.0%	1	20
HSI	I	4-4-4	8.8	50.0%	1	21
LAB	N	3-4-4	8.8	44.4%	1	21
LAB	N	4-4-4	9.6	38.9%	1	21
LAB	S	3-4-4	8.2	38.9%	1	19
LAB	S	4-4-4	8.4	33.3%	1	19
LAB	I	3-4-4	7.6	38.9%	1	17
LAB	I	4-4-4	9.1	33.3%	1	18

Tabelle 12: Ergebnisse unter Verwendung von Farbhistogrammen mit L_∞–Metrik

FM	HKV	Quantisierung	$\overline{R}$	T_3	$R_{\min}$	$R_{\max}$
RGB	N	2-2-2	2.9	83.3%	1	19
RGB	N	3-3-3	3.3	72.2%	1	20
RGB	N	4-4-4	3.1	77.8%	1	20
RGB	S	2-2-2	1.8	94.4%	1	6
RGB	S	3-3-3	1.5	94.4%	1	4
RGB	S	4-4-4	1.5	94.4%	1	4
RGB	I	2-2-2	4.8	66.7%	1	20
RGB	I	3-3-3	4.8	66.7%	1	20
RGB	I	4-4-4	5.1	66.7%	1	20
III	N	4-4-3	3.3	72.2%	1	20
III	N	4-4-4	3.3	83.3%	1	21
III	S	4-4-3	2.3	83.3%	1	5
III	S	4-4-4	2.3	83.3%	1	5
III	I	4-4-3	3.7	66.7%	1	21
III	I	4-4-4	4.1	61.1%	1	21
HSI	N	4-3-2	3.7	61.1%	1	16
HSI	N	4-3-3	3.4	66.7%	1	15
HSI	N	4-4-3	3.4	66.7%	1	16
HSI	N	4-4-4	3.4	66.7%	1	16
HSI	S	4-3-2	2.6	77.8%	1	9
HSI	S	4-3-3	2.6	77.8%	1	9
HSI	S	4-4-3	2.7	83.3%	1	11
HSI	S	4-4-4	2.7	83.3%	1	11
HSI	I	4-3-2	3.0	83.3%	1	13
HSI	I	4-3-3	3.7	66.7%	1	17
HSI	I	4-4-3	4.3	55.6%	1	18
HSI	I	4-4-4	4.2	61.1%	1	18
LAB	N	3-4-4	4.8	66.7%	1	20
LAB	N	4-4-4	4.7	61.1%	1	20
LAB	S	3-4-4	5.4	38.9%	1	12
LAB	S	4-4-4	5.6	44.4%	1	12
LAB	I	3-4-4	4.6	61.1%	1	19
LAB	I	4-4-4	4.8	55.6%	1	18

Tabelle 13: Ergebnisse unter Verwendung von Farbsummenhistogrammen mit L_1–Metrik

FM	HKV	Quantisierung	$\overline{R}$	T_3	$R_{\min}$	$R_{\max}$
RGB	N	2-2-2	4.2	77.8%	1	21
RGB	N	3-3-3	3.9	66.7%	1	21
RGB	N	4-4-4	3.9	72.2%	1	21
RGB	S	2-2-2	2.1	94.4%	1	7
RGB	S	3-3-3	1.5	94.4%	1	4
RGB	S	4-4-4	1.9	83.3%	1	6
RGB	I	2-2-2	4.8	66.7%	1	20
RGB	I	3-3-3	4.7	72.2%	1	20
RGB	I	4-4-4	4.8	66.7%	1	20
III	N	4-4-3	4.4	72.2%	1	20
III	N	4-4-4	4.3	72.2%	1	21
III	S	4-4-3	5.6	44.4%	1	13
III	S	4-4-4	5.6	44.4%	1	13
III	I	4-4-3	4.2	66.7%	1	21
III	I	4-4-4	4.2	61.1%	1	21
HSI	N	4-3-2	4.5	55.6%	1	17
HSI	N	4-3-3	4.5	61.1%	1	18
HSI	N	4-4-3	4.4	55.6%	1	18
HSI	N	4-4-4	4.2	61.1%	1	18
HSI	S	4-3-2	2.7	77.8%	1	9
HSI	S	4-3-3	2.7	77.8%	1	9
HSI	S	4-4-3	3.2	72.2%	1	15
HSI	S	4-4-4	3.2	72.2%	1	15
HSI	I	4-3-2	3.4	77.8%	1	18
HSI	I	4-3-3	4.2	61.1%	1	18
HSI	I	4-4-3	4.5	55.6%	1	18
HSI	I	4-4-4	4.4	55.6%	1	18
LAB	N	3-4-4	5.3	55.6%	1	20
LAB	N	4-4-4	5.3	55.6%	1	20
LAB	S	3-4-4	7.2	38.9%	1	15
LAB	S	4-4-4	7.2	44.4%	1	15
LAB	I	3-4-4	5.1	55.6%	1	18
LAB	I	4-4-4	5.1	50.0%	1	18

Tabelle 14: Ergebnisse unter Verwendung von Farbsummenhistogrammen mit L_2–Metrik

FM	HKV	Quantisierung	$\overline{R}$	T_3	$R_{\min}$	$R_{\max}$
RGB	N	2-2-2	5.2	61.1%	1	21
RGB	N	3-3-3	5.5	44.4%	1	21
RGB	N	4-4-4	5.2	55.6%	1	21
RGB	S	2-2-2	2.0	88.9%	1	7
RGB	S	3-3-3	1.4	100.0%	1	3
RGB	S	4-4-4	2.4	72.2%	1	8
RGB	I	2-2-2	5.6	61.1%	1	20
RGB	I	3-3-3	6.1	61.1%	1	20
RGB	I	4-4-4	6.1	61.1%	1	20
III	N	4-4-3	6.9	50.0%	1	20
III	N	4-4-4	6.9	50.0%	1	20
III	S	4-4-3	7.4	38.9%	1	18
III	S	4-4-4	7.4	38.9%	1	18
III	I	4-4-3	5.8	55.6%	1	20
III	I	4-4-4	5.9	44.4%	1	19
HSI	N	4-3-2	4.8	61.1%	1	19
HSI	N	4-3-3	4.6	55.6%	1	20
HSI	N	4-4-3	5.4	44.4%	1	20
HSI	N	4-4-4	5.2	50.0%	1	20
HSI	S	4-3-2	2.9	83.3%	1	12
HSI	S	4-3-3	2.9	83.3%	1	12
HSI	S	4-4-3	4.9	50.0%	1	20
HSI	S	4-4-4	4.9	50.0%	1	20
HSI	I	4-3-2	4.3	66.7%	1	21
HSI	I	4-3-3	4.9	66.7%	1	21
HSI	I	4-4-3	5.4	50.0%	1	21
HSI	I	4-4-4	4.7	55.6%	1	19
LAB	N	3-4-4	7.3	44.4%	1	17
LAB	N	4-4-4	7.2	44.4%	1	17
LAB	S	3-4-4	8.6	33.3%	1	18
LAB	S	4-4-4	8.4	38.9%	1	18
LAB	I	3-4-4	7.1	38.9%	1	17
LAB	I	4-4-4	6.8	38.9%	1	16

Tabelle 15: Ergebnisse unter Verwendung von Farbsummenhistogrammen mit L_∞–Metrik

etwas schlechter ist. Sie erreicht nur beim HKV S und grober Quantisierung in den Modellen RGB und HSI akzeptable Werte.
Bei der L_2–Metrik ist die Ursache darin zu suchen, daß große Differenzen in einzelnen Histogramm–Bins aufgrund der Quadrierung überbewertet werden. Diese Differenzen treten oft dann auf, wenn sich korrespondierende Pixel in den zu vergleichenden Histogrammen in benachbarten Spalten (Bins) befinden. Dies ist auf kleine Farbabweichungen zwischen der Aufnahme des zu erkennenden Objekts und dem Bild in der Datenbank zurückzuführen. Die L_∞–Metrik führt mit ihrer Maximumsbildung offenbar zu einem zu großen Informationsverlust, der sie für den Vergleich von Farbhistogrammen ungeeignet macht.
Durch die Verwendung von Summenhistogrammen lassen sich bei der L_1–Metrik die Ergebnisse teilweise deutlich verbessern; die Ergebnisse im RGB–Modell beim HKV I und im HSI–Modell beim HKV S und I verschlechtern sich jedoch. Das Gegenfarbenmodell erreicht bei den Summenhistogrammen mit L_1–Metrik ungefähr die gleichen Ergebnisse wie das HSI–Modell. Das RGB–Modell liefert jedoch auch hier die besten Ergebnisse, wobei das HKV S wieder am besten abschneidet und das HKV I schlechtere Ergebnisse liefert als der Verzicht auf HKV.
Auch bei der L_2–Metrik erzielt man mit der Verwendung von Summenhistogrammen eine merkliche Verbesserung der Ergebnisse. Die Ergebnisse der L_1–Metrik werden jedoch nicht erreicht. Bei der L_∞–Metrik wird ebenfalls eine deutliche Verbesserung erzielt. Im RGB–Modell beim HKV S und einer mittleren Quantisierung (3-3-3) wird sogar ein sehr guter Wert erzielt. Insgesamt sind die Ergebnisse aber relativ schlecht.

7.2.2 Momente von Farbverteilungen

FM	HKV	$\overline{R}$	T_3	$R_{\min}$	$R_{\max}$
RGB	N	5.4	61.1%	1	21
RGB	S	2.9	83.3%	1	11
RGB	I	5.8	55.6%	1	21
III	N	3.6	66.7%	1	19
III	S	2.9	77.8%	1	10
III	I	4.5	61.1%	1	20
HSI	N	3.6	72.2%	1	13
HSI	S	3.4	66.7%	1	12
HSI	I	3.9	66.7%	1	16
LAB	N	4.2	66.7%	1	19
LAB	S	3.1	72.2%	1	7
LAB	I	3.6	66.7%	1	17

Tabelle 16: Ergebnisse unter Verwendung der ersten drei Momente der Farbverteilung

Tabelle 16 zeigt die Ergebnisse bei der Verwendung der ersten drei Momente der Farbverteilung ohne Einbeziehung räumlicher Information. Die Gewichtung des Moments erster Ordnung betrug für alle Kanäle 1.0; die Momente zweiter und dritter Ordnung wurden in allen Kanälen mit 2.0 gewichtet. Die leichte Untergewichtung des Mittelwertes führt zu einer gewissen Unempfindlichkeit gegenüber Verschiebungen im Farbenraum und somit zu einer gewissen Unabhängigkeit von Schwankungen in der Farbe und Intensität der Beleuchtung.
Als Parameter für die Umrechnung des Abstandes zweier Momentenvektoren in das einheitliche Intervall nach Abschnitt 6.2.1 wurde $n = 3.0$ gewählt, da sich damit in den Experimenten zur Ermittlung geeigneter Parameter (Abschnitt 7.1) die beste Trennschärfe erreichen ließ.
Die Ergebnisse sind insgesamt deutlich schlechter als die bei der Verwendung von Farbhistogrammen erzielbaren. Die besten Ergebnisse werden im RGB– und im Gegenfarbenmodell, jeweils mit dem HKV S, erreicht. Auch beim CIELAB–Modell liefert das HKV S die besten Ergebnisse. Beim HSI–Modell sind das HKV S und der Verzicht auf HKV ungefähr gleich. Beim CIELAB–Modell liefert das HKV I noch relativ zu den anderen HKV gute Ergebnisse. Das HKV I und der Verzicht auf HKV sind für das RGB–Modell hier ungeeignet.
Eine Verbesserung der Ergebnisse läßt sich sehr wahrscheinlich durch eine geeignetere Wahl der Gewichte, insbesondere mit einer Anpassung an das Farbmodell, erzielen.

7.2.3 Invarianten von Farbhistogrammen

Die Tabellen 17 auf der nächsten Seite und 18 auf der nächsten Seite zeigen die mit den Invarianten von Farbhistogrammen erzielten Ergebnisse. Für Tabelle 17 wurde als Maß für die Übereinstimmung der euklidische Abstand mit der Umrechnung nach Abschnitt 6.2.1 ($n = 2.0$) gewählt, für Tabelle 18 wurde die in Abschnitt 5.2.5 definierte Ähnlichkeitsfunktion verwendet.
Insgesamt fallen die sehr schlechten Werte im durchschnittlichen Rang $\overline{R}$ und im maximalen Rang $R_{\max}$ auf, wobei — wie am minimalen Rang gut zu sehen ist — das Verfahren durchaus geeignet ist. Es ist zu vermuten, daß die Quantisierung des Farbenraums vor der Berechnung der Histogramme dafür verantwortlich ist. In den Vortests ist dieses Problem nicht zu Tage getreten. Es ist daher zu untersuchen, wie sich das Verfahren bei Verwendung von Histogrammen mit voller Farbauflösung verhält. D. Slater hat eine entsprechende speicherplatzsparende Datenstruktur (hashbasiertes Histogramm) entwickelt.
Die besten Ergebnisse werden im HSI–Modell ohne HK mit dem Verhältnis–Maß erreicht. Mit dem euklidischen Abstand können diese Ergebnisse nicht erreicht werden. Hier erzielt das RGB–Modell ohne HK noch die besten Ergebnisse.

FM	HKV	Quantisierung	$\overline{R}$	T_3	$R_{\min}$	$R_{\max}$
RGB	N	4	9.4	16.7%	1	21
RGB	N	5	4.7	55.6%	1	14
RGB	S	4	6.1	55.6%	1	21
RGB	S	5	5.7	55.6%	1	14
RGB	I	4	8.8	38.9%	1	20
RGB	I	5	8.0	38.9%	1	19
III	N	4	9.0	16.7%	1	21
III	N	5	7.0	55.6%	1	18
III	S	4	12.2	5.6%	3	21
III	S	5	12.2	5.6%	3	21
III	I	4	7.1	27.8%	1	21
III	I	5	5.5	55.6%	1	15
HSI	N	4	6.8	50.0%	1	19
HSI	N	5	5.9	61.1%	1	19
HSI	S	4	12.2	5.6%	3	21
HSI	S	5	12.2	5.6%	3	21
HSI	I	4	7.3	44.4%	1	19
HSI	I	5	5.8	61.1%	1	19
LAB	N	4	8.5	11.1%	2	20
LAB	N	5	7.8	16.7%	1	20
LAB	S	4	9.5	22.2%	1	21
LAB	S	5	8.7	22.2%	1	16
LAB	I	4	9.2	27.8%	1	21
LAB	I	5	10.6	11.1%	2	21

Tabelle 17: Ergebnisse unter Verwendung der Invarianten von Farbhistogrammen mit euklidischem Abstand

FM	HKV	Quantisierung	$\overline{R}$	T_3	$R_{\min}$	$R_{\max}$
RGB	N	4	7.6	22.2%	1	16
RGB	S	4	8.7	38.9%	1	20
RGB	I	4	6.8	38.9%	1	18
III	N	4	6.9	50.0%	1	19
III	S	4	12.2	5.6%	3	21
III	I	4	7.2	27.8%	1	16
HSI	N	4	3.2	72.2%	1	14
HSI	S	4	12.2	5.6%	3	21
HSI	I	4	4.7	55.6%	1	17
LAB	N	4	7.7	22.2%	1	17
LAB	I	4	9.4	11.1%	1	20

Tabelle 18: Ergebnisse unter Verwendung der Invarianten von Farbhistogrammen mit Verhältnis–Maß nach Abschnitt 5.2.5

Die Quantisierung hat nur bei einigen Farbmodellen und HKV Einfluß auf die Ergebnisse (RGB ohne HKV, Gegenfarbenmodell, HSI und CIELAB ohne HKV und mit dem HKV I). Mit Ausnahme von CIELAB mit dem HKV I erreicht die feinere Quantisierung von fünf Bit pro Kanal die besseren Ergebnisse. Dies stützt die obige These für die Ursache der schlechten Ergebnisse.
Bei der Verwendung des Verhältnis–Maßes liefert der Verzicht auf die Intensitätsinformation (HKV S), mit Ausnahme des RGB–Modells, die deutlich schlechteren Ergebnisse. Bei der Verwendung des euklidischen Abstands ist das HKV S nur beim Gegenfarben- und HSI–Modell völlig unbrauchbar.
Die Fähigkeit, größere Farbschwankungen auszugleichen, war bei den durchgeführten Experimenten nicht erforderlich. Sie hat im Gegenteil sogar zur Verschlechterung einiger Ergebnisse beigetragen, da sie dazu führen kann, daß Objekte mit sehr ähnlichen Aufdrucken, die sich nur in der Farbe unterscheiden, nicht gut unterschieden werden können. Ein Beispiel sind die Objekte `Valensina_Saft_Orange_0.75l` bzw. `Valensina_Saft_Orange_0.2l` und `Valensina_Saft_Multi_0.75l`. Beim ersteren ist der Hintergrund des Aufdrucks rot, beim zweiten rotbraun. Die andersfarbigen Teile des Aufdrucks haben einen relativ geringen Anteil an der Gesamtoberfläche und sind im wesentlichen sehr ähnlich. Eine Verwechslung der Objekte trat bei insgesamt zwölf Parametersätzen[49] (davon zehn mal das Objekt `Valensina_Saft_Multi_0.75l`) auf, wobei das jeweils andere Objekt oft einen Rang über drei erhielt.

7.2.4 Integration von räumlicher Information bei Farbhistogrammen und Momenten

Kachelung

Die Tabellen 19, 20, 21, 22 und 23 zeigen die Ergebnisse der Integration von räumlicher Information in Form einer gleichmäßigen Kachelung in die Farbhistogramme. Zum Vergleich der wurde Histogramme wurde die Histogrammschnitt–Methode verwendet.
Gegenüber den Ergebnissen ohne räumliche Information (Tabellen 2, 3, 4 und 5) werden , entgegen der Erwartung, in etwa gleiche Ergebnisse erzielt. Bei einigen Parametern ergeben sich leichte Verbesserungen, bei anderen geringe Verschlechterungen.
Es sei zunächst die gröbere 2×2–Kachelung betrachtet. Im RGB–Modell verbessern sich hauptsächlich die Werte ohne HKV bei den mittleren Quantisierungen sowie einige feinere Quantisierungen (3-4-4, 4-3-4 und 4-4-4) beim HKV S und eine gröbere und eine mittlere Quantisierung (2-3-3 und 3-4-3) beim HKV I. Die deutlichsten Verschlechterungen im RGB–Modell ergeben sich beim HKV S und den groben Quantisierungen ohne HKV.

[49] davon sechs im RGB–Modell, vier im Gegenfarbenmodell und zwei im CIELAB–Modell

FM	HKV	Quantisierung	$\overline{R}$	T_3	$R_{\min}$	$R_{\max}$
RGB	N	2-2-2	4.1	72.2%	1	19
RGB	N	2-3-2	4.5	72.2%	1	20
RGB	N	2-3-3	3.9	77.8%	1	21
RGB	N	3-2-3	4.4	66.7%	1	21
RGB	N	3-3-2	4.2	72.2%	1	20
RGB	N	3-3-3	3.9	72.2%	1	21
RGB	N	3-4-3	3.8	72.2%	1	21
RGB	N	3-4-4	3.8	72.2%	1	20
RGB	N	4-3-4	4.1	77.8%	1	21
RGB	N	4-4-3	4.2	72.2%	1	21
RGB	N	4-4-4	3.9	77.8%	1	20
RGB	S	2-2-2	1.9	94.4%	1	10
RGB	S	2-3-2	1.8	94.4%	1	9
RGB	S	2-3-3	1.8	94.4%	1	9
RGB	S	3-2-3	2.0	94.4%	1	13
RGB	S	3-3-2	1.9	94.4%	1	13
RGB	S	3-3-3	2.0	94.4%	1	13
RGB	S	3-4-3	1.9	94.4%	1	12
RGB	S	3-4-4	2.2	83.3%	1	6
RGB	S	4-3-4	2.0	83.3%	1	6
RGB	S	4-4-3	1.9	94.4%	1	12
RGB	S	4-4-4	2.1	83.3%	1	6
RGB	I	2-2-2	4.7	66.7%	1	20
RGB	I	2-3-2	4.6	61.1%	1	19
RGB	I	2-3-3	4.7	66.7%	1	20
RGB	I	3-2-3	4.7	66.7%	1	21
RGB	I	3-3-2	4.5	66.7%	1	21
RGB	I	3-3-3	4.5	66.7%	1	21
RGB	I	3-4-3	4.8	61.1%	1	20
RGB	I	3-4-4	5.1	61.1%	1	20
RGB	I	4-3-4	5.0	55.6%	1	21
RGB	I	4-4-3	4.6	66.7%	1	21
RGB	I	4-4-4	4.8	61.1%	1	21

Tabelle 19: Ergebnisse unter Verwendung von Farbhistogrammen mit Kachelung 2×2, RGB–Modell

FM	HKV	Quantisierung	$\overline{R}$	T_3	$R_{\min}$	$R_{\max}$
III	N	3-3-2	5.5	61.1%	1	20
III	N	3-3-3	5.4	61.1%	1	20
III	N	4-4-3	4.6	66.7%	1	20
III	N	4-4-4	4.6	72.2%	1	19
III	S	3-3-2	6.6	44.4%	1	16
III	S	3-3-3	6.6	44.4%	1	16
III	S	4-4-3	4.0	72.2%	1	17
III	S	4-4-4	4.0	72.2%	1	17
III	I	3-3-2	5.0	66.7%	1	20
III	I	3-3-3	4.9	61.1%	1	20
III	I	4-4-3	3.9	66.7%	1	20
III	I	4-4-4	4.8	66.7%	1	20

Tabelle 20: Ergebnisse unter Verwendung von Farbhistogrammen mit Kachelung 2×2, Gegenfarbenmodell

FM	HKV	Quantisierung	$\overline{R}$	T_3	$R_{\min}$	$R_{\max}$
HSI	N	3-2-2	4.4	61.1%	1	19
HSI	N	3-3-2	5.0	61.1%	1	20
HSI	N	3-3-3	5.4	61.1%	1	20
HSI	N	4-3-2	5.1	61.1%	1	20
HSI	N	4-3-3	5.4	61.1%	1	20
HSI	N	4-4-3	4.7	55.6%	1	20
HSI	N	4-4-4	4.4	55.6%	1	21
HSI	S	3-2-2	1.9	94.4%	1	13
HSI	S	3-3-2	2.1	94.4%	1	13
HSI	S	3-3-3	2.1	94.4%	1	13
HSI	S	4-3-2	2.1	94.4%	1	13
HSI	S	4-3-3	2.1	94.4%	1	13
HSI	S	4-4-3	3.1	72.2%	1	14
HSI	S	4-4-4	3.1	72.2%	1	14
HSI	I	3-2-2	4.5	72.2%	1	19
HSI	I	3-3-2	4.6	66.7%	1	20
HSI	I	3-3-3	6.3	50.0%	1	20
HSI	I	4-3-2	4.6	66.7%	1	20
HSI	I	4-3-3	6.4	50.0%	1	20
HSI	I	4-4-3	4.6	61.1%	1	20
HSI	I	4-4-4	4.8	55.6%	1	20

Tabelle 21: Ergebnisse unter Verwendung von Farbhistogrammen mit Kachelung 2×2, HSI–Modell

FM	HKV	Quantisierung	$\overline{R}$	T_3	$R_{\min}$	$R_{\max}$
LAB	N	2-3-3	6.1	44.4%	1	15
LAB	N	3-3-3	6.2	44.4%	1	18
LAB	N	3-4-4	5.5	55.6%	1	20
LAB	N	4-4-4	5.2	66.7%	1	20
LAB	S	2-3-3	8.0	38.9%	1	16
LAB	S	3-3-3	7.3	38.9%	1	16
LAB	S	3-4-4	5.9	44.4%	1	15
LAB	S	4-4-4	5.9	44.4%	1	15
LAB	I	2-3-3	6.4	38.9%	1	16
LAB	I	3-3-3	6.2	44.4%	1	18
LAB	I	3-4-4	5.7	50.0%	1	17
LAB	I	4-4-4	5.1	55.6%	1	16

Tabelle 22: Ergebnisse unter Verwendung von Farbhistogrammen mit Kachelung 2×2, CIELAB–Modell

FM	HKV	Quantisierung	$\overline{R}$	T_3	$R_{\min}$	$R_{\max}$
RGB	N	2-2-2	5.0	70.6%	1	21
RGB	N	3-3-3	4.3	77.8%	1	21
RGB	N	4-4-4	3.9	72.2%	1	21
RGB	S	2-2-2	2.2	88.9%	1	13
RGB	S	3-3-3	1.9	94.4%	1	11
RGB	S	4-4-4	2.1	83.3%	1	6
RGB	I	2-2-2	4.6	72.2%	1	20
RGB	I	3-3-3	4.3	66.7%	1	20
RGB	I	4-4-4	4.3	72.2%	1	21
III	N	4-4-3	4.4	77.8%	1	20
III	N	4-4-4	4.7	77.8%	1	20
III	S	4-4-3	4.2	66.7%	1	18
III	S	4-4-4	4.2	66.7%	1	18
III	I	4-4-3	4.2	72.2%	1	20
III	I	4-4-4	4.2	72.2%	1	20
HSI	N	4-3-2	4.5	50.0%	1	20
HSI	N	4-3-3	5.1	61.1%	1	21
HSI	N	4-4-3	4.8	72.2%	1	21
HSI	N	4-4-4	4.6	66.7%	1	21
HSI	S	4-3-2	2.0	94.4%	1	12
HSI	S	4-3-3	2.0	94.4%	1	12
HSI	S	4-4-3	3.4	55.6%	1	10
HSI	S	4-4-4	3.4	55.6%	1	10
HSI	I	4-3-2	4.6	66.7%	1	20
HSI	I	4-3-3	5.3	55.6%	1	20
HSI	I	4-4-3	4.4	66.7%	1	20
HSI	I	4-4-4	4.3	66.7%	1	20
LAB	N	3-4-4	5.1	66.7%	1	20
LAB	N	4-4-4	5.0	72.2%	1	20
LAB	S	3-4-4	6.1	44.4%	1	16
LAB	S	4-4-4	6.1	44.4%	1	16
LAB	I	3-4-4	5.3	44.4%	1	17
LAB	I	4-4-4	4.9	50.0%	1	16

Tabelle 23: Ergebnisse unter Verwendung von Farbhistogrammen mit Kachelung 4×4

Im Gegenfarbenmodell ergeben sich nur für das HKV S in den groben Quantisierungen (3-3-2 und 3-3-3) Verbesserungen. Es handelt sich dabei jedoch um die schlechtesten Ergebnisse im Gegenfarbenmodell. Für alle anderen Parameter werden im Gegenfarbenmodell durch die Kachelung nur deutliche Verschlechterungen erreicht. Beim HSI–Modell hingegen erzielt die Kachelung ohne HKV deutliche Verbesserungen (mit Ausnahme der feinsten Quantisierung, 4-4-4) und beim HKV S leichte Verbesserungen (Ausnahme Quantisierung 3-2-2). Beim HKV I ergeben sich jedoch deutliche Verschlechterungen.
Bei Verwendung des CIELAB–Modells bleiben die Werte für das HKV S nahezu unverändert. Beim HKV I verbessern sich die Ergebnisse für die Quantisierungen 2-3-3 und 4-4-4 leicht, für die mittleren Quantisierungen tritt jedoch eine merkliche Verschlechterung ein. Die Ergebnisse ohne HKV sind alle schlechter als die Ergebnisse ohne Kachelung.
Bei einer feineren Kachelung (4×4) ändert sich das Bild etwas. Beim RGB–Modell gibt es gegenüber den Ergebnissen ohne Kachelung zwei leichte und eine deutliche Verbesserung (HKV S 4-4-4 und HKV I 3-3-3 sowie HKV I 4-4-4). Dem gegenüber stehen viele deutliche Verschlechterungen, wobei der erste Parametersatz (ohne HKV, 2-2-2) das Extrem darstellt.
Im Gegenfarbenmodell ergeben sich ausschließlich Verschlechterungen, wohingegen im HSI–Modell sich insbesondere ohne HKV Verbesserungen ergeben (Ausnahme: Quantisierung 4-4-4). Bei den HKV S und I ergeben sich beim HSI–Modell teilweise Verbesserungen, teilweise bleiben die Ergebnisse annähernd gleich. Eine Verschlechterung ergibt sich nur für das HKV I mit der Quantisierung 4-3-2.
Beim CIELAB–Modell erreicht die feinere Kachelung im wesentlichen Verschlechterungen. Lediglich beim HKV I verbessern sich die Ergebnisse, bei der feinen Quantisierung sogar deutlich.

FM	HKV	$\overline{R}$	T_3	$R_{\min}$	$R_{\max}$
RGB	N	5.4	61.1%	1	21
RGB	S	2.9	83.3%	1	11
RGB	I	5.8	61.1%	1	21
III	N	3.6	66.7%	1	19
III	S	2.8	77.8%	1	9
III	I	4.4	61.1%	1	20
HSI	N	3.5	72.2%	1	13
HSI	S	3.4	66.7%	1	12
HSI	I	3.9	66.7%	1	16
LAB	N	4.0	72.2%	1	19
LAB	S	3.1	72.2%	1	7
LAB	I	3.6	72.2%	1	17

Tabelle 24: Ergebnisse unter Verwendung der ersten drei Momente der Farbverteilung mit Kachelung 2×2

FM	HKV	$\overline{R}$	T_3	$R_{\min}$	$R_{\max}$
RGB	N	6.3	44.4%	1	20
RGB	S	2.4	77.8%	1	9
RGB	I	5.6	66.7%	1	21
III	N	5.0	61.1%	1	20
III	S	2.7	77.8%	1	10
III	I	5.6	50.0%	1	20
HSI	N	2.6	77.8%	1	8
HSI	S	2.2	88.9%	1	8
HSI	I	3.1	66.7%	1	11
LAB	N	4.3	66.7%	1	19
LAB	S	2.6	77.8%	1	9
LAB	I	4.1	72.2%	1	18

Tabelle 25: Ergebnisse unter Verwendung der ersten drei Momente der Farbverteilung mit Kachelung 4×4

Die Tabellen 24 und 25 zeigen die Ergebnisse bei der Verwendung der Kachelung im Zusammenhang mit den Momenten. Bei der groben Kachelung (2×2) bleiben die Ergebnisse weitgehend unverändert. Verbesserungen ergeben sich beim Gegenfarbenmodell bei den HKV S und I, beim HSI–Modell ohne HKV und beim CIELAB–Modell ohne HKV. Mit der feineren 4×4–Kachelung verbessern sich einige Ergebnisse deutlich: beim RGB–Modell mit den HKV S und I, beim Gegenfarben- und CIELAB–Modell jeweils mit dem HKV S und generell beim HSI–Modell. Die anderen Ergebnisse verschlechtern sich, teilweise sogar sehr deutlich.

Insgesamt erreicht die Kachelung nicht die erwartete Verbesserung. Die besten erzielbaren Ergebnisse sind deutlich schlechter als die Ergebnisse, die ohne Kachelung erzielt werden können. Hier ist jedoch der Einfluß der geringen Größe der Datenbank zu untersuchen. Ein weiteres Problem für die Integration räumlicher Information ist auch die schlechte Qualität der Entzerrung der rekonstruierten Flächen aufgrund von Schwächen der verwendeten geometrischen Rekonstruktion, insbesondere die ungenaue Extraktion der Flächen.

Unscharfe Regionen

Die Tabellen 26, 27, 28, 29, 30 und 31 zeigen die Ergebnisse der Kombination von unscharfen Regionen mit Farbhistogrammen und Histogrammschnitt. Gegenüber Farbhistogrammen ohne räumliche Information ergeben sich größtenteils deutliche Verschlechterungen. Beim HSI–Modell ohne HKV ergeben sich dagegen Verbesserungen. Allerdings sind die Ergebnisse merklich schlechter als die im RGB–Modell mit dem HKV S erzielbaren. Mit den kleineren Zentralregionen (0.3 und 0.4, Tabellen 31 und 30) sind die Ergebnisse insgesamt etwas besser als mit der großen Zentralregion. Die

FM	HKV	Quantisierung	$\overline{R}$	T_3	$R_{\min}$	$R_{\max}$
RGB	N	2-2-2	4.6	72.2%	1	21
RGB	N	2-3-2	4.4	72.2%	1	21
RGB	N	2-3-3	3.8	77.8%	1	21
RGB	N	3-2-3	4.2	72.2%	1	21
RGB	N	3-3-2	4.2	72.2%	1	20
RGB	N	3-3-3	4.3	77.8%	1	21
RGB	N	3-4-3	4.2	77.8%	1	21
RGB	N	3-4-4	4.1	77.8%	1	21
RGB	N	4-3-4	4.3	72.2%	1	21
RGB	N	4-4-3	4.2	77.8%	1	21
RGB	N	4-4-4	4.2	77.8%	1	21
RGB	S	2-2-2	2.2	83.3%	1	9
RGB	S	2-3-2	2.1	88.9%	1	9
RGB	S	2-3-3	1.6	100.0%	1	3
RGB	S	3-2-3	1.9	88.9%	1	7
RGB	S	3-3-2	2.2	88.9%	1	8
RGB	S	3-3-3	1.8	94.4%	1	7
RGB	S	3-4-3	1.6	94.4%	1	6
RGB	S	3-4-4	2.7	77.8%	1	11
RGB	S	4-3-4	2.6	72.2%	1	8
RGB	S	4-4-3	1.8	88.9%	1	7
RGB	S	4-4-4	2.7	72.2%	1	9
RGB	I	2-2-2	5.4	66.7%	1	21
RGB	I	2-3-2	5.1	61.1%	1	20
RGB	I	2-3-3	4.9	61.1%	1	20
RGB	I	3-2-3	5.0	61.1%	1	20
RGB	I	3-3-2	5.0	66.7%	1	21
RGB	I	3-3-3	4.8	61.1%	1	20
RGB	I	3-4-3	4.9	61.1%	1	20
RGB	I	3-4-4	4.8	61.1%	1	20
RGB	I	4-3-4	4.6	66.7%	1	20
RGB	I	4-4-3	4.6	66.7%	1	20
RGB	I	4-4-4	4.7	66.7%	1	20

Tabelle 26: Ergebnisse unter Verwendung von Farbhistogrammen mit unscharfen Regionen, beide Halbachsen 0.5, Abklingbereich 20 Pixel, RGB–Modell

FM	HKV	Quantisierung	$\overline{R}$	T_3	$R_{\min}$	$R_{\max}$
III	N	3-3-2	6.2	61.1%	1	20
III	N	3-3-3	5.7	61.1%	1	20
III	N	4-4-3	5.1	72.2%	1	20
III	N	4-4-4	4.9	72.2%	1	20
III	S	3-3-2	7.5	50.0%	1	19
III	S	3-3-3	7.5	50.0%	1	19
III	S	4-4-3	5.2	61.1%	1	16
III	S	4-4-4	5.2	61.1%	1	16
III	I	3-3-2	5.4	72.2%	1	20
III	I	3-3-3	5.5	61.1%	1	20
III	I	4-4-3	4.6	66.7%	1	20
III	I	4-4-4	4.6	66.7%	1	20

Tabelle 27: Ergebnisse unter Verwendung von Farbhistogrammen mit unscharfen Regionen, beide Halbachsen 0.5, Abklingbereich 20 Pixel, Gegenfarbenmodell

FM	HKV	Quantisierung	$\overline{R}$	T_3	$R_{\min}$	$R_{\max}$
HSI	N	3-2-2	4.4	66.7%	1	20
HSI	N	3-3-2	5.1	66.7%	1	20
HSI	N	3-3-3	5.0	61.1%	1	21
HSI	N	4-3-2	5.1	66.7%	1	20
HSI	N	4-3-3	5.1	61.1%	1	21
HSI	N	4-4-3	3.9	88.9%	1	21
HSI	N	4-4-4	4.0	72.2%	1	21
HSI	S	3-2-2	2.4	88.9%	1	12
HSI	S	3-3-2	2.4	88.9%	1	11
HSI	S	3-3-3	2.3	88.9%	1	11
HSI	S	4-3-2	2.3	88.9%	1	11
HSI	S	4-3-3	2.3	88.9%	1	11
HSI	S	4-4-3	4.1	61.1%	1	18
HSI	S	4-4-4	4.1	61.1%	1	18
HSI	I	3-2-2	5.2	66.7%	1	20
HSI	I	3-3-2	4.9	61.1%	1	20
HSI	I	3-3-3	6.2	55.6%	1	21
HSI	I	4-3-2	5.1	61.1%	1	20
HSI	I	4-3-3	6.2	55.6%	1	21
HSI	I	4-4-3	4.8	55.6%	1	21
HSI	I	4-4-4	4.4	66.7%	1	20

Tabelle 28: Ergebnisse unter Verwendung von Farbhistogrammen mit unscharfen Regionen, beide Halbachsen 0.5, Abklingbereich 20 Pixel, HSI–Modell

FM	HKV	Quantisierung	$\overline{R}$	T_3	$R_{\min}$	$R_{\max}$
LAB	N	2-3-3	6.8	50.0%	1	19
LAB	N	3-3-3	6.3	50.0%	1	20
LAB	N	3-4-4	5.1	61.1%	1	20
LAB	N	4-4-4	4.6	66.7%	1	20
LAB	S	2-3-3	8.5	38.9%	1	18
LAB	S	3-3-3	7.7	38.9%	1	18
LAB	S	3-4-4	6.3	44.4%	1	18
LAB	S	4-4-4	6.3	44.4%	1	18
LAB	I	2-3-3	7.4	38.9%	1	20
LAB	I	3-3-3	6.9	44.4%	1	20
LAB	I	3-4-4	6.2	50.0%	1	20
LAB	I	4-4-4	6.6	55.6%	1	20

Tabelle 29: Ergebnisse unter Verwendung von Farbhistogrammen mit unscharfen Regionen, beide Halbachsen 0.5, Abklingbereich 20 Pixel, CIELAB–Modell

FM	HKV	Quantisierung	$\overline{R}$	T_3	$R_{\min}$	$R_{\max}$
RGB	N	2-2-2	4.6	66.7%	1	21
RGB	N	3-3-3	4.1	77.8%	1	21
RGB	N	4-4-4	4.1	83.3%	1	21
RGB	S	2-2-2	2.4	88.9%	1	13
RGB	S	3-3-3	1.8	88.9%	1	7
RGB	S	4-4-4	2.8	72.2%	1	9
RGB	I	2-2-2	5.3	66.7%	1	20
RGB	I	3-3-3	4.8	61.1%	1	20
RGB	I	4-4-4	4.7	66.7%	1	20
III	N	4-4-3	5.1	72.2%	1	20
III	N	4-4-4	4.9	72.2%	1	20
III	S	4-4-3	5.1	66.7%	1	17
III	S	4-4-4	5.1	66.7%	1	17
III	I	4-4-3	4.6	66.7%	1	20
III	I	4-4-4	4.7	61.1%	1	20
HSI	N	4-3-2	4.8	72.2%	1	20
HSI	N	4-3-3	4.8	61.1%	1	21
HSI	N	4-4-3	4.0	77.8%	1	21
HSI	N	4-4-4	3.9	88.9%	1	21
HSI	S	4-3-2	2.3	88.9%	1	9
HSI	S	4-3-3	2.4	88.9%	1	9
HSI	S	4-4-3	4.0	66.7%	1	16
HSI	S	4-4-4	4.0	66.7%	1	16
HSI	I	4-3-2	5.2	72.2%	1	20
HSI	I	4-3-3	6.2	55.6%	1	21
HSI	I	4-4-3	4.7	61.1%	1	21
HSI	I	4-4-4	4.7	61.1%	1	20
LAB	N	3-4-4	5.2	66.7%	1	20
LAB	N	4-4-4	4.9	66.7%	1	20
LAB	S	3-4-4	6.4	44.4%	1	18
LAB	S	4-4-4	6.4	44.4%	1	18
LAB	I	3-4-4	6.6	50.0%	1	20
LAB	I	4-4-4	6.7	55.6%	1	20

Tabelle 30: Ergebnisse unter Verwendung von Farbhistogrammen mit unscharfen Regionen, beide Halbachsen 0.4, Abklingbereich 20 Pixel

FM	HKV	Quantisierung	$\overline{R}$	T_3	$R_{\min}$	$R_{\max}$
RGB	N	2-2-2	4.7	66.7%	1	21
RGB	N	3-3-3	4.2	77.8%	1	21
RGB	N	4-4-4	4.2	83.3%	1	21
RGB	S	2-2-2	2.5	83.3%	1	16
RGB	S	3-3-3	1.7	88.9%	1	5
RGB	S	4-4-4	2.8	72.2%	1	9
RGB	I	2-2-2	5.3	66.7%	1	20
RGB	I	3-3-3	4.8	61.1%	1	20
RGB	I	4-4-4	4.7	66.7%	1	21
III	N	4-4-3	5.1	66.7%	1	20
III	N	4-4-4	4.8	66.7%	1	20
III	S	4-4-3	5.2	66.7%	1	17
III	S	4-4-4	5.2	66.7%	1	17
III	I	4-4-3	4.7	66.7%	1	20
III	I	4-4-4	4.6	61.1%	1	20
HSI	N	4-3-2	4.7	72.2%	1	20
HSI	N	4-3-3	4.8	66.7%	1	21
HSI	N	4-4-3	4.2	83.3%	1	21
HSI	N	4-4-4	4.2	77.8%	1	21
HSI	S	4-3-2	2.2	83.3%	1	7
HSI	S	4-3-3	2.1	83.3%	1	5
HSI	S	4-4-3	3.9	55.6%	1	15
HSI	S	4-4-4	3.9	55.6%	1	15
HSI	I	4-3-2	5.1	72.2%	1	20
HSI	I	4-3-3	6.2	55.6%	1	21
HSI	I	4-4-3	5.1	61.1%	1	21
HSI	I	4-4-4	4.9	61.1%	1	20
LAB	N	3-4-4	5.5	66.7%	1	20
LAB	N	4-4-4	5.4	66.7%	1	20
LAB	S	3-4-4	6.4	44.4%	1	18
LAB	S	4-4-4	6.4	44.4%	1	18
LAB	I	3-4-4	6.2	55.6%	1	20
LAB	I	4-4-4	6.6	55.6%	1	20

Tabelle 31: Ergebnisse unter Verwendung von Farbhistogrammen mit unscharfen Regionen, beide Halbachsen 0.3, Abklingbereich 20 Pixel

schlechten Ergebnisse sind auf Fehler bei der geometrischen Rekonstruktion zurückzuführen. Der Abklingbereich von 20 Pixeln ist für den Ausgleich der auftretenden Verschiebungen zwischen rekonstruiertem Aufdruck und korrespondierender Datenbankfläche vermutlich zu klein.
Für die Farbsummenhistogramme mit unscharfen Regionen und Metriken wurde für beide Halbachsen der Wert 0.5 sowie ein Abklingbereich von 20 Pixeln verwendet.
Die Tabellen 32, 33 und 34 zeigen die Ergebnisse mit Farbsummenhistogrammen und Metriken in Kombination mit unscharfen Regionen. Gegenüber den entsprechenden Verfahren ohne Integration räumlicher Information (Tabellen 13, 14 bzw. 15) verschlechtert sich eine Reihe von Ergebnissen. Verbesserungen werden hauptsächlich bei der L_1–Metrik beim HSI–Modell (ohne HKV und mit dem HKV S), bei der L_2–Metrik beim HSI–Modell ohne HKV in den Quantisierungen 4-3-2 und 4-3-3 und beim Gegenfarben- und CIELAB–Modell (jeweils mit dem HKV S) sowie bei der L_∞–Metrik beim Gegenfarbenmodell (ohne HKV und mit dem HKV S), beim CIELAB–Modell (HKV S) und beim HSI–Modell mit HKV S in den Quantisierungen 4-3-2 und 4-3-3 erreicht.
Insgesamt sind die Ergebnisse jedoch relativ schlecht. Es gilt das bereits oben bezüglich der Metriken gesagte. Die durch die Kombination mit den unscharfen Regionen erzielten Verbesserungen betreffen nur Farbmodelle und HK, deren Ergebnisse ohne Einbeziehung räumlicher Information lediglich durchschnittlich sind.
Die Tabelle 35 zeigt die Ergebnisse bei Verwendung unscharfer Regionen mit den Momenten. Gegenüber den Ergebnissen ohne Einbeziehung räumlicher Information ergibt sich größtenteils eine Verschlechterung. Nur die Ergebnisse im RGB–Modell ohne HKV und im HSI–Modell ohne HKV und mit dem HKV I verbessern sich, beim HSI–Modell sogar deutlich. Die Ergebnisse bleiben insgesamt meist hinter den durch die Kachelung erzielbaren Werten. Im RGB– und HSI–Modell ohne HKV und im HSI–Modell mit dem HKV S werden jedoch die besten Werte aller auf Momenten basierender Verfahren erzielt. Die Ergebnisse im HSI–Modell sind auch die besten mit Momenten erreichten, wobei das HKV S sich davon am besten eignet. Eventuell läßt sich eine Verbesserung der Ergebnisse durch eine Gewichtung der Regionen erreichen.

FM	HKV	Quantisierung	$\overline{R}$	T_3	$R_{\min}$	$R_{\max}$
RGB	N	2-2-2	4.9	77.8%	1	21
RGB	N	3-3-3	4.8	77.8%	1	21
RGB	N	4-4-4	5.0	77.8%	1	21
RGB	S	2-2-2	2.8	77.8%	1	9
RGB	S	3-3-3	2.1	88.9%	1	8
RGB	S	4-4-4	2.3	83.3%	1	8
RGB	I	2-2-2	5.1	66.7%	1	21
RGB	I	3-3-3	4.9	72.2%	1	21
RGB	I	4-4-4	5.1	72.2%	1	21
III	N	4-4-3	4.6	77.8%	1	21
III	N	4-4-4	4.7	77.8%	1	21
III	S	4-4-3	3.1	77.8%	1	12
III	S	4-4-4	3.1	77.8%	1	12
III	I	4-4-3	5.2	72.2%	1	21
III	I	4-4-4	5.2	72.2%	1	21
HSI	N	4-3-2	3.3	77.8%	1	16
HSI	N	4-3-3	3.2	83.3%	1	17
HSI	N	4-4-3	3.4	72.2%	1	18
HSI	N	4-4-4	3.4	77.8%	1	18
HSI	S	4-3-2	2.4	83.3%	1	9
HSI	S	4-3-3	2.3	83.3%	1	9
HSI	S	4-4-3	2.3	83.3%	1	9
HSI	S	4-4-4	2.3	83.3%	1	9
HSI	I	4-3-2	3.4	72.2%	1	15
HSI	I	4-3-3	3.9	72.2%	1	17
HSI	I	4-4-3	4.1	66.7%	1	18
HSI	I	4-4-4	3.9	72.2%	1	19
LAB	N	3-4-4	5.1	72.2%	1	21
LAB	N	4-4-4	5.3	77.8%	1	21
LAB	S	3-4-4	5.4	44.4%	1	16
LAB	S	4-4-4	5.8	50.0%	1	16
LAB	I	3-4-4	5.2	66.7%	1	20
LAB	I	4-4-4	4.7	61.1%	1	20

Tabelle 32: Ergebnisse unter Verwendung von Farbsummenhistogrammen mit unscharfen Regionen und L_1–Metrik

FM	HKV	Quantisierung	$\overline{R}$	T_3	$R_{\min}$	$R_{\max}$
RGB	N	2-2-2	5.2	72.2%	1	20
RGB	N	3-3-3	4.7	77.8%	1	20
RGB	N	4-4-4	4.6	72.2%	1	20
RGB	S	2-2-2	2.6	83.3%	1	12
RGB	S	3-3-3	2.1	94.4%	1	13
RGB	S	4-4-4	2.8	77.8%	1	14
RGB	I	2-2-2	5.1	72.2%	1	20
RGB	I	3-3-3	5.3	72.2%	1	20
RGB	I	4-4-4	5.2	72.2%	1	20
III	N	4-4-3	4.3	72.2%	1	18
III	N	4-4-4	4.4	72.2%	1	19
III	S	4-4-3	4.1	61.1%	1	13
III	S	4-4-4	4.2	61.1%	1	13
III	I	4-4-3	4.6	66.7%	1	21
III	I	4-4-4	4.8	66.7%	1	21
HSI	N	4-3-2	4.7	61.1%	1	18
HSI	N	4-3-3	3.9	55.6%	1	18
HSI	N	4-4-3	4.5	61.1%	1	20
HSI	N	4-4-4	4.4	61.1%	1	20
HSI	S	4-3-2	2.7	88.9%	1	11
HSI	S	4-3-3	2.8	88.9%	1	11
HSI	S	4-4-3	3.5	72.2%	1	13
HSI	S	4-4-4	3.5	66.7%	1	13
HSI	I	4-3-2	4.0	66.7%	1	18
HSI	I	4-3-3	5.3	50.0%	1	21
HSI	I	4-4-3	5.6	50.0%	1	21
HSI	I	4-4-4	5.3	50.0%	1	20
LAB	N	3-4-4	5.3	72.2%	1	19
LAB	N	4-4-4	5.2	72.2%	1	19
LAB	S	3-4-4	6.7	44.4%	1	16
LAB	S	4-4-4	6.7	44.4%	1	15
LAB	I	3-4-4	5.6	61.1%	1	19
LAB	I	4-4-4	5.1	66.7%	1	18

Tabelle 33: Ergebnisse unter Verwendung von Farbsummenhistogrammen mit unscharfen Regionen und L_2–Metrik

FM	HKV	Quantisierung	$\overline{R}$	T_3	$R_{\min}$	$R_{\max}$
RGB	N	2-2-2	6.2	61.1%	1	21
RGB	N	3-3-3	6.4	61.1%	1	21
RGB	N	4-4-4	6.4	50.0%	1	21
RGB	S	2-2-2	2.1	88.9%	1	11
RGB	S	3-3-3	2.0	88.9%	1	10
RGB	S	4-4-4	3.3	72.2%	1	12
RGB	I	2-2-2	5.3	66.7%	1	21
RGB	I	3-3-3	6.1	61.1%	1	21
RGB	I	4-4-4	5.9	55.6%	1	21
III	N	4-4-3	6.3	50.0%	1	19
III	N	4-4-4	6.6	50.0%	1	19
III	S	4-4-3	7.0	50.0%	1	19
III	S	4-4-4	7.0	50.0%	1	19
III	I	4-4-3	6.2	50.0%	1	21
III	I	4-4-4	6.4	50.0%	1	21
HSI	N	4-3-2	4.1	72.2%	1	20
HSI	N	4-3-3	5.6	61.1%	1	21
HSI	N	4-4-3	5.4	55.6%	1	21
HSI	N	4-4-4	5.3	61.1%	1	21
HSI	S	4-3-2	2.7	83.3%	1	10
HSI	S	4-3-3	2.7	83.3%	1	10
HSI	S	4-4-3	5.4	50.0%	1	19
HSI	S	4-4-4	5.4	50.0%	1	19
HSI	I	4-3-2	5.6	61.1%	1	21
HSI	I	4-3-3	5.9	55.6%	1	21
HSI	I	4-4-3	6.2	38.9%	1	21
HSI	I	4-4-4	6.1	50.0%	1	21
LAB	N	3-4-4	7.4	55.6%	1	19
LAB	N	4-4-4	7.3	55.6%	1	19
LAB	S	3-4-4	7.9	44.4%	1	20
LAB	S	4-4-4	7.9	44.4%	1	20
LAB	I	3-4-4	7.1	50.0%	1	20
LAB	I	4-4-4	7.1	55.6%	1	20

Tabelle 34: Ergebnisse unter Verwendung von Farbsummenhistogrammen mit unscharfen Regionen und L_∞–Metrik

FM	HKV	$\overline{R}$	T_3	$R_{\min}$	$R_{\max}$
RGB	N	5.2	61.1%	1	19
RGB	S	3.2	72.2%	1	13
RGB	I	6.1	61.1%	1	21
III	N	4.7	61.1%	1	19
III	S	3.2	72.2%	1	11
III	I	4.7	61.1%	1	20
HSI	N	2.2	77.8%	1	6
HSI	S	2.1	88.9%	1	7
HSI	I	3.2	66.7%	1	13
LAB	N	4.4	83.3%	1	18
LAB	S	3.2	72.2%	1	14
LAB	I	5.1	66.7%	1	18

Tabelle 35: Ergebnisse unter Verwendung der ersten drei Momente der Farbverteilung mit unscharfen Regionen, beide Halbachsen 0.5, Abklingbereich 20 Pixel

7.2.5 Tupel–Histogramme

FM	HKV	Quantisierung	$\overline{R}$	T_3	$R_{\min}$	$R_{\max}$
RGB	N	2-2-2	3.1	83.3%	1	19
RGB	S	2-2-2	1.9	88.9%	1	7
RGB	I	2-2-2	4.4	66.7%	1	20
III	N	2-2-2	5.9	50.0%	1	16
III	S	2-2-2	8.5	38.9%	1	18
III	I	2-2-2	5.5	50.0%	1	17
LAB	N	2-2-2	6.3	44.4%	1	16
LAB	S	2-2-2	8.1	38.9%	1	16
LAB	I	2-2-2	6.8	33.3%	1	16

Tabelle 36: Ergebnisse unter Verwendung von Tupel–Histogrammen mit Histogrammschnitt. Tupelgeometrie „1.0,0.0; 0.0,1.0; 1.0,1.0“

FM	HKV	Quantisierung	$\overline{R}$	T_3	$R_{\min}$	$R_{\max}$
RGB	N	2-2-2	3.3	83.3%	1	19
RGB	S	2-2-2	1.9	88.9%	1	7
RGB	I	2-2-2	4.4	66.7%	1	20
III	N	2-2-2	5.9	50.0%	1	16
III	S	2-2-2	8.5	38.9%	1	18
III	I	2-2-2	5.6	50.0%	1	17
HSI	N	3-2-2	4.6	61.1%	1	20
HSI	S	3-2-2	1.6	100.0%	1	3
HSI	I	3-2-2	4.5	66.7%	1	19
LAB	N	2-2-2	6.3	44.4%	1	16
LAB	S	2-2-2	8.1	38.9%	1	16
LAB	I	2-2-2	6.8	33.3%	1	16

Tabelle 37: Ergebnisse unter Verwendung von Tupel–Histogrammen mit Histogrammschnitt. Tupelgeometrie „1.0,0.0; 0.0,1.0“

Die Tabellen 36, 37 und 38 zeigen die mit den Tupel–Histogrammen erhaltenen Ergebnisse. Auffällig ist die große Schwankungsbreite der Ergebnisse in Abhängigkeit vom Farbmodell und dem HKV. Durch Vergleich der Tabellen 36 und 37 läßt sich der Einfluß der Tupelgeometrie beurteilen. Tabelle 38 zeigt die Veränderungen in den Ergebnissen mit der Schwellenwert–Methode und den Einfluß des Schwellenwertes.
Die besten Ergebnisse werden in den Farbmodellen RGB und HSI mit dem HKV S erreicht. Beide Farbmodelle haben auch ohne HKV und mit dem HKV I noch bessere Werte als das Gegenfarben- und das CIELAB–Modell. Ein Einfluß der Tupelgeometrie ist nicht zu verzeichnen. Das insgesamt beste Ergebnis wird im HSI–Modell mit dem HKV S erzielt.
Die Verwendung der Schwellenwert–Methode verbessert beim kleineren Schwellenwert von 1% die Ergebnisse im HSI–Modell nochmals sehr. Im Ge-

FM	Quantisierung	Schwellenwert	$\overline{R}$	T_3	$R_{\min}$	$R_{\max}$
RGB	2-2-2	1%	4.6	44.4%	1	12
III	2-2-2	1%	5.7	38.9%	1	18
HSI	3-2-2	1%	3.1	72.2%	1	14
LAB	2-2-2	1%	6.7	27.8%	1	19
RGB	2-2-2	5%	3.0	77.8%	1	14
III	2-2-2	5%	6.0	55.6%	1	17
HSI	3-2-2	5%	5.6	44.4%	1	14
LAB	2-2-2	5%	7.5	27.8%	2	17

Tabelle 38: Ergebnisse unter Verwendung von Tupel–Histogrammen mit Schwellenwert–Methode. 2–Tupel, Tupelgeometrie „1.0,0.0; 0.0,1.0“, ohne Verfahren zur Helligkeitskonstanz

genfarbenmodell wird ebenfalls eine leichte Verbesserung erzielt. Ganz anders fallen die Ergebnisse bei einem Schwellenwert von 5% aus. Hier verbessern sich die Ergebnisse im RGB–Modell gegenüber der Histogrammschnitt–Methode. In den anderen Farbmodellen verschlechtern sich die Ergebnisse. Das bedeutet, daß auch seltener vorkommende Farbkombinationen für den Zuordnungsprozeß von Bedeutung sind. Bei einem hohen Schwellenwert werden nur besonders häufige Merkmalsausprägungen berücksichtigt, die für eine eindeutige Zuordnung von rekonstruierten Aufdrucken zu Bildern der Datenbank zu unspezifisch sind.

Das CIELAB–Modell erscheint für die Tupel–Histogramme ungeeignet; es liefert die schlechtesten Ergebnisse. Auch das Gegenfarbenmodell erzielt nur schlechte Ergebnisse. Sowohl RGB– als auch HSI–Modell erreichen gute Ergebnisse, wobei das HSI–Modell am besten abschneidet.

Inwieweit die Anzahl der Proben das Ergebnis beeinflußt, ist noch zu untersuchen. Ein wichtiger Faktor hierbei ist auch die Güte des Zufallszahlengenerators, mit dem die Koordinaten der Proben im Bild bestimmt werden.

7.2.6 Deformierte und verschmutzte Objekte

Ergänzend wurden einige Verfahren mit zwei Datensätzen deformierter[50] und vier Datensätzen verschmutzter[51] Objekte getestet.

Bei den deformierten Objekten wird ein durchschnittlicher Rang von 3.1 ($T_3 = 30.6\%$) und bei den verschmutzten Objekten von 2.0 ($T_3 = 80.0\%$) erreicht. Die geringe Toleranz gegenüber Deformationen wird im wesentlichen von den Verfahren zur geometrischen Rekonstruktion verursacht, die bei diesen Objekten die einzelnen Flächen nicht mehr korrekt erkennen.

Die Ergebnisse liegen im Bereich der Ergebnisse der anderen Objekte. Bei den Farbhistogrammen mit Kachelung werden sogar merklich bessere Ergeb-

[50] `Milsani_Kondensmilch_0.312l` und `Valensina_Saft_Orange_0.2l`

[51] zweimal `Milsani_Vollmilch_1l`, `Kirberg_Saft_TraubenRot_1l` und `Milsani_Kondensmilch_0.312l`

Verfahren	FM	HKV	$\overline{R}$	T_3	$R_{\min}$	$R_{\max}$
HIST	RGB	N	2.5	83.3%	1	8
HIST	RGB	S	1.5	83.3%	1	4
HIST	RGB	I	2.3	83.3%	1	8
HIST	III	N	2.3	83.3%	1	8
HIST	III	S	2.2	83.3%	1	4
HIST	III	I	1.8	83.3%	1	5
HIST	HSI	N	2.8	83.3%	1	9
HIST	HSI	S	2.7	83.3%	1	10
HIST	HSI	I	2.0	83.3%	1	4
HIST	LAB	N	2.7	66.7%	1	7
HIST	LAB	S	4.2	50.0%	1	12
HIST	LAB	I	3.0	66.7%	1	7
HISTx	RGB	N	2.5	83.3%	1	8
HISTx	RGB	S	1.3	100.0%	1	3
HISTx	RGB	I	1.8	83.3%	1	5
HISTx	III	N	2.0	83.3%	1	6
HISTx	III	S	2.0	100.0%	1	3
HISTx	III	I	1.8	83.3%	1	5
HISTx	HSI	N	2.3	83.3%	1	5
HISTx	HSI	S	2.7	83.3%	1	10
HISTx	HSI	I	1.8	100.0%	1	3
HISTx	LAB	N	2.8	66.7%	1	7
HISTx	LAB	S	3.8	50.0%	1	10
HISTx	LAB	I	2.8	66.7%	1	6
HISTsp	RGB	N	2.5	83.3%	1	8
HISTsp	RGB	S	1.7	83.3%	1	5
HISTsp	RGB	I	2.7	66.7%	1	7
HISTsp	III	N	2.3	83.3%	1	9
HISTsp	III	S	2.2	83.3%	1	4
HISTsp	III	I	2.5	83.3%	1	9
HISTsp	HSI	N	3.0	66.7%	1	10
HISTsp	HSI	S	3.0	83.3%	1	12
HISTsp	HSI	I	2.3	83.3%	1	7
HISTsp	LAB	N	3.0	83.3%	1	10
HISTsp	LAB	S	3.7	66.7%	1	10
HISTsp	LAB	I	3.7	66.7%	1	10

Tabelle 39: Ergebnisse mit deformierten und verschmutzten Objekten. Verfahren: HIST: Farbhistogramme mit Histogrammschnitt, HISTx: wie HIST, jedoch ohne Entzerrung der rekonstruierten Flächen, HISTsp: wie HIST, jedoch mit Kachelung 2×2, Quantisierung einheitlich 4-4-4

nisse erzielt. Hier macht sich der Einfluß der geringen Anzahl von Objekten bemerkbar, insbesondere das Fehlen der „problematischen“ Datensätze, siehe Abschnitt 7.2.

7.2.7 Vergleich des Rechenzeitbedarfs

Verfahren	Zeit	Systemanteil	
Farbhistogramme m. Histogrammschnitt	39.7s	28.0%	11.1s
dto., ohne Entzerrung	40.0s	28.5%	11.4s
Farbhist. m. H.schnitt u. Kachelung	71.2s	16.3%	11.6s
Farbhist. m. H.schnitt u. unscharfen R.	152.7s	8.6%	13.1s
Farbhist. m. L_1–Metrik	39.9s	28.6%	11.4s
Farbhist. m. L_2–Metrik	39.9s	28.1%	11.2s
Farbhist. m. L_∞–Metrik	40.1s	27.9%	11.2s
Farbsummenhist. m. L_1–M.	158.9s	7.2%	11.5s
Farbsummenhist. m. L_2–M.	158.8s	7.1%	11.2s
Farbsummenhist. m. L_∞–M.	158.8s	7.1%	11.2s
Farbsum.h. m. L_1–M. u. unscharfen R.	705.2s	2.1%	14.9s
Farbsum.h. m. L_2–M. u. unscharfen R.	704.7s	2.2%	15.2s
Farbsum.h. m. L_∞–M. u. unscharfen R.	707.7s	2.1%	15.1s
Momente	52.9s	21.4%	11.3s
Momente m. Kachelung	119.9s	9.6%	11.5s
Momente m. unscharfen Regionen	217.8s	5.5%	11.9s
Invarianten v. Farbhist.	80.5s	14.0%	11.3s
dto., mit Verhältnismaß	80.7s	13.6%	11.0s
Tupel–Histogramme	57.6s	20.0%	11.5s
dto., m. Schwellenwert–Methode	58.1s	20.1%	11.7s

Tabelle 40: Rechenzeit für die Erkennung eines Objekts (drei rekonstruierte Flächen nach der Vorauswahl) mit den einzelnen Verfahren inklusive der Zeiten für Programmstart und Einlesen der Datenbank jedoch ohne die Rechenzeit für die 2.5D–Rekonstruktion. Gemessen wurde auf einer SPARCStation 20–502. Die Meßwerte sind das arithmetische Mittel aus drei Messungen. Es wurde jeweils das RGB–Modell ohne Verfahren zur Helligkeitskonstanz verwendet. Histogrammauflösung 4-4-4 (für Tupel 1-1-1), Kachelung mit 4×4 Regionen, unscharfe Regionen: beide Halbachsen 0.5, Abklingbereich 20 Pixel, Tupel: Proben: 20% der Pixel, 4–Tupel, Schwellenwert 1% der Proben.

Tabelle 40 zeigt die Zeiten, die für die Erkennung eines Objekts bei den jeweiligen Verfahren benötigt wird. Der Systemanteil ist der Anteil an Rechenzeit, der in Routinen des Betriebssystems benötigt wird. Er ist, wenn überhaupt, nur durch eine veränderte Benutzung der Funktionen des Betriebssystems zu senken.

Die Zeit, die die allgemeinen Routinen benötigen, beträgt ohne Ausrichtung der rekonstruierten Flächen 34.6s (davon 10.9s Systemanteil). Das Einlesen der Datenbank hat daran einen Anteil von 1.5s (davon 0.7s Systemanteil). Die reale Zeit, die während des Einlesens der Datenbank vergeht, beträgt je nach Netzbelastung ca. 4s. Diese Zeit entfällt bei einer wiederholten Erkennung von Objekten mit derselben Datenbank und denselben Parametern. Das Einlesen der Bilddaten ist in den letztgenannten Zeiten nicht enthalten, da die Bilder während der Erkennung bei Bedarf geladen werden. Dies ist erforderlich, weil der Bedarf an virtuellem Speicher zu groß wäre, wenn alle Bilder der Datenbank ständig im Speicher gehalten würden [29]. In einer weiterentwickelten Implementierung könnten die zum Vergleich der Bilder verwendeten Merkmale selbst gespeichert werden. Ihre Größe ist (in der Regel) gering genug, um die Merkmale aller Bilder der Datenbank im Speicher zu halten. Aufgrund der Vielzahl der in der vorliegenden Implementierung verwendeten Farbmerkmale wurde bislang auf deren Integration in die abgespeicherte Datenbank verzichtet.
Die starke Parametrisierung der Verfahren führt zu einer merklichen Verlängerung der Rechenzeit, da keine spezifischen Optimierungen vorgenommen werden, können. Insbesondere die Berechnung der Farbhistogramme ließe sich dann merklich in der Geschwindigkeit steigern. Bedarf für Optimierung besteht auch bei der sehr langwierigen Berechnung der Summenhistogramme.
Die Routinen zur Berechnung der Regionenzugehörigkeit bei der Verwendung unscharfer Regionen nach [42] können ebenfalls noch optimiert werden. In der vorliegenden Implementierung wird die Regionenzugehörigkeit zwar im voraus berechnet und in einer Tabelle abgelegt, jedoch muß diese Tabelle bei jedem Bild neu berechnet werden, da sie nur für jeweils eine bestimmte Bildgröße berechnet wird[52]. Es wäre zu untersuchen, ob eine einmalige Berechnung der Tabelle in hoher Auflösung und ihre anschließende Skalierung entsprechend der Größe des zu bearbeitenden Bildes effizienter ist.
Es sei an dieser Stelle auf die theoretischen Überlegungen zum Zeitaufwand in Abschnitt 5.3.2 verwiesen.
Tabell 41 zeigt die für die Konvertierung der Farbmodelle und die Verfahren zur Helligkeitskonstanz zusätzlich benötigte Rechenzeit. Sie wurden mit den gleichen Daten ermittelt wie die Werte aus Tabelle 40, so daß Zeiten einfach addiert werden können.
Es ist zu beachten, daß die Routinen zur Konvertierung der Farbmodelle nicht optimiert sind. Durch Verwendung von Lookup–Tabellen bzw. Caches für häufig vorkommende RGB–Tripel ließe sich eine deutliche Rechenzeitein-

[52]Die Bilder der Datenbank sind nicht auf eine einheitliche Größe skaliert, und die rekonstruierten Aufdrucke werden ebenfalls nicht skaliert, da die räumliche Information nur relativ zur Bildgröße berücksichtigt wird, siehe Abschnitt 6.1 auf Seite 53. Eine einheitliche Skalierung wäre in bezug auf den Verlust an Auflösung bzw. die Erhöhung des Speicherplatzbedarfs ungeeignet.

FM	HKV	Zeit	Systemanteil
RGB	N	0.0s	0.0s
RGB	S	14.7s	0.1s
RGB	I	70.5s	0.5s
HSI	N	95.5s	0.6s
HSI	S	111.1s	0.8s
HSI	I	130.6s	0.8s
III	N	12.1s	0.4s
III	S	26.7s	0.5s
III	I	47.3s	0.6s
LAB	N	242.5s	0.3s
LAB	S	259.5s	0.5s
LAB	I	275.0s	0.8s

Tabelle 41: Zusätzliche Rechenzeit für die Konvertierung der Farbmodelle und die Verfahren zur Helligkeitskonstanz bei der Erkennung eines Objekts mit drei rekonstruierten Flächen

sparung erzielen. Eine weitere Verbesserung ergäbe sich mit der Verwendung von Spezialhardware.

Für die im vorherigen Abschnitt aufgeführten Ergebnisse haben zwei SPARC-Station 10–40 einen Monat gerechnet. Dabei ist jedoch zu berücksichtigen, daß beide Computer während dieser Zeit auch anderweitig benutzt werden mußten und somit zeitweilig weniger als 50% der CPU–Leistung für die Tests zur Verfügung stand. Ein weiteres Problem waren die über das Netzwerk anzusprechenden Massenspeicher (NFS). Die dadurch entstehenden Wartezeiten waren zum Teil erheblich hoch[53].

7.3 Zusammenfassung

Von den auf Farbhistogrammen basierenden Verfahren werden die besten Ergebnisse mit dem Basisverfahren (mit Histogrammschnitt, ohne räumliche Information) erzielt. Beim Verzicht auf die Entzerrung der rekonstruierten Flächen verändern sich die Ergebnisse kaum. Die schlechten Ergebnisse bei der Einbeziehung von räumlicher Information sind in Fehlern bei der geometrischen Rekonstruktion zu suchen. Werden die Flächen des zu erkennenden Objekts nur partiell rekonstruiert, so verschieben sich die Regionengrenzen entsprechend. Dies führt zu einer ungünstigen Überdeckung mit den Regionen im korrespondierenden Bild der Datenbank. Im Rahmen von [29] wurden diese Beobachtungen bereits bezüglich eines pixelweisen Vergleichs der Bilder gemacht. Untersuchungen zur Behebung dieser Probleme wurden zeitgleich mit dieser Arbeit von K. Oppermann vorgenommen [28].

[53] Genaue Werte konnten jedoch nicht ermittelt werden.

Eine weitere Ursache für die schlechten Ergebnisse bei Berücksichtigung räumlicher Information sind die Bilder der Datenbank. Die Glanzlichter, die bei einigen Objekten nicht zu vermeiden waren (siehe Abschnitt 3.4.1 auf Seite 28), treten in den für die Rekonstruktion der Aufdrucke verwendeten Aufnahmen nicht auf, da dort weniger deformierte Objekte verwendet werden konnten und die Objekte manuell so positioniert wurden, daß keine Glanzlichter auftraten. Dadurch kommt es zu einem deutlichen Unterschied in der Farbverteilung zwischen dem zu erkennenden Objekt und den korrespondierenden Bildern der Datenbank. Da der Einfluß der Glanzlichter regional begrenzt ist, sind besonders die Verfahren unter Einbeziehung von räumlicher Information betroffen. Die bereits erwähnten schwarzen Streifen am Rand einiger weniger Bilder der Datenbank[54] resultieren in einer Verschiebung des Aufdrucks, die sich ebenfalls negativ auf die Verfahren mit räumlicher Information auswirkt
Bei den Verfahren unter Verwendung der Momente der Farbverteilung können durch die Integration räumlicher Information trotzdem Verbesserungen erzielt werden. Insgesamt werden jedoch die mit den Farbhistogrammen erreichbaren Ergebnisse nicht erreicht.
Die Ursachen für die schlechten Ergebnisse mit den Invarianten von Farbhistogrammen wurden bereits oben dargelegt. Die Stärke des Verfahrens, große Farbschwankungen auszugleichen, war bei den durchgeführten Experimenten, anders als bei den Experimenten zur Vorauswahl der Parameter, nicht relevant. Sie kann jedoch dazu führen, daß Objekte mit sehr ähnlichen Aufdrucken, die sich nur in der Farbe unterscheiden, nicht gut unterschieden werden können, siehe Diskussion der Ergenisse des Verfahrens.
Mit den Tupel–Histogrammen können ähnlich gute Ergebnisse wie mit den Farbhistogrammen erzielt werden. Dies trifft sowohl auf die Histogrammschnitt–Methode als auch auf die mit weniger Speicherplatzaufwand verbundene Schwellenwert–Methode zu.
Die Tupel–Histogramme sind das mit deutlichem Abstand schnellste Verfahren, das räumliche Information mit einbezieht. Bei ihnen besteht jedoch ein ernsthaftes Problem bezüglich des Speicherbedarfs, das eine entsprechend grobe Quantisierung des Farbenraumes erforderlich macht. Dadurch wird die maximale Anzahl unterscheidbarer Merkmalsausprägungen stark eingeschränkt. Das kann bei großen Datenbanken dazu führen, daß die Anzahl der Bilder von Aufdrucken in den Größenordnungsbereich der Anzahl unterscheidbarer Merkmalsausprägungen gelangt und so die Unterscheidbarkeit der Bilder stark verringert.

Insgesamt werden mit den Farbhistogrammen die besten Ergebnisse erreicht. Die Tupel–Histogramme erzielen nur etwas schlechtere Werte. Sie können den Vorteil der Einbeziehung räumlicher Information erst bei größeren Da-

[54]aufgrund von Meßfehlern bei der manuellen Erstellung der 3D–Modelle, siehe Abschnitt 3.4.1

tenbanken zur Geltung bringen. Als besonders geeignet erwiesen sich das RGB–Modell und daneben das HSI–Modell. Bei den Tupel–Histogrammen erreicht das HSI–Modell die besseren Ergebnisse, bei den Farbhistogrammen das RGB–Modell. Als bestes Verfahren zur Helligkeitskonstanz erwies sich die Transformation in das rgb–Modell vor der Konvertierung in das Zielfarbmodell, also der Verzicht auf Intensitätsinformation. Dies ist hauptsächlich darin begründet, daß die meisten Farben der verwendeten Objekte sich deutlich im Buntton unterscheiden. Bei einer größeren Anzahl von Objekten ist das eventuell nicht der Fall. Das Verfahren zur Helligkeitskonstanz mit Normierung der Intensität hat sich insgesamt als ungeeignet erwiesen. Es ist zu vermuten, daß die Modellannahme (Skalierbarkeit der Intensität mit einem im gesamten Bild konstanten Faktor) für das verwendet Bildmaterial nicht gilt. Die Ergebnisse mit synthetischen Bildern, die dem Modell genügten, waren deutlich besser.
Die in den Experimenten erhaltenen Ergebnisse sind allerdings mit einer gewissen Vorsicht zu betrachten, da die Experimente aus Zeitgründen[55] mit einer relativ kleinen Datenbank erfolgen mußten. Eine Übertragung der Ergebnisse auf Objekte mit völlig anderen Aufdrucken ist nicht ohne weiteres möglich.

Die für die Erkennung eines Objekts benötigte Rechenzeit ist für einen Praxiseinsatz des Objekterkennungssystems noch zu hoch. Zwar entspricht die verwendete Rechentechnik nicht mehr dem Stand der Technik, jedoch ist in der Praxis mit einer wesentlich größeren Datenbank zu rechnen. Einige Optimierungen bei den einzelnen Verfahren wurden bereits oben (Abschnitt 7.2.7) erwähnt. Weitere Optimierungen sind bei der Festlegung auf ein bestimmtes Verfahren zum Vergleich der Bilder im Bereich der allgemeinen Routinen möglich.

[55] Je nach Verfahren und Belastung des Netzes erfordert ein Durchlauf mit allen 18 Datensätzen der geometrischen Rekonstruktion 20 Minuten bis 4.5 Stunden (SPARCStation 10–40).

8 Ausblick

8.1 3D–Texture–Scanning

Beim 3D–Texture–Scanning besteht im Bereich der Segmentierung noch Bedarf für weitergehende Untersuchungen, insbesondere ist an eine automatische Überprüfung der Segmentierungsergebnisse und eine adaptive Einstellung der Parameter (Schwellenwert der Hintergrundsegmentierung, Wahl des Verfahrens zur Überlagerung der Linienzeichnung) zu denken.
Alternative Verfahren zur Bereitstellung der 3D–Daten wurden bereits in Abschnitt 3.5 diskutiert. Für gekrümmte Flächen ist eine geeignete Erweiterung der ESNI–Repräsentation, wie sie in Abschnitt 2.2.2 bereits angesprochen wurde, zu untersuchen.

8.2 Verfahren zum Vergleich von Bildern

Ein Schwerpunkt weiterer Untersuchungen sollten die Verfahren zum Vergleich von Bildern (Matchingverfahren) sein. Hier sind sowohl weitergehende Experimenete mit veränderten Parametern sowie Modifikationen der vorgestellten Verfahren als auch die Analyse weiterer Verfahren von Interesse.

Variationen der vorgestellten Verfahren

Eine vollständige Rotationsinvarianz bei den in Abschnitt 5.1.5 beschrieben unscharfen Regionen (*fuzzy regions*) ließe sich erzielen, indem man die vier Randregionen zusammenfaßt. Damit entstünde eine annähernde Rotationsinvarianz für beliebige Winkel, und die in Abschnitt 6.1 beschriebene (relativ aufwendige) Ausrichtung der rekonstruierten Fläche könnte entfallen. Das Problem dabei ist die nicht unbeding rotationssymmetrische Zentralregion, da sie elliptisch und von dem Seitenlängenverhältnis des Bildes abhängig ist. Eine Lösung wäre, beide Halbachsen der Zentralregion gleichlang zu wählen und die Bilder vor dem Vergleich auf eine quadratische Größe[56] „umzuskalieren“.

Eine Kachelung mit unscharfer Trennung der Regionen ist vermutlich besser geeignet als die einfache Kachelung, da sie weniger empfindlich auf Verschiebungen reagiert. Die Untersuchung einer linear abklingenden Zugehörigkeit an der Regionengrenze wäre der cosinusförmigen nach [42] gegenüberzustellen. Bei einem linearen Abklingen verringert sich der Rechenaufwand für die Bestimmung der Regionenzugehörigkeit.

[56] Die tatsächliche Bildgröße ist für das Verfahren nur sekundär von Bedeutung, da sie nur die Detailauflösung bestimmt.

Weitere Verfahren

Trotz der Vielfalt an Verfahren, die berücksichtigt wurden, handelt es sich nur um eine kleine Auswahl. Einige Matchingverfahren konnten aufgrund der Beschränkung des Umfangs dieser Arbeit nicht untersucht werden. Bei anderen Verfahren wurde bereits in der Literatur auf ihre Unterlegenheit gegenüber den im Objekterkennungssystem verwendeten Verfahren hingewiesen. Ihre Untersuchung ist jedoch — wenn auch in verkürzter Form — für eine vollständige Beurteilung der zur Verfügung stehenden Verfahren zum Vergleich von Bildern durchaus wünschenswert.

Das von B. Funt and G. Finlayson in [15] vorgestellte *color constant color indexing* ist eine Variation des von M. Swain und D. Ballard vorgestellten Verfahrens [46] mit dem Versuch, Beleuchtungsunabhängigkeit zu erzielen. Die starken Einschränkungen des Verfahrens und die in der Literatur [40] berichteten, im Vergleich mit anderen Verfahren schlechten Ergebnisse haben dazu geführt dieses Verfahren nicht in die Untersuchung einzubeziehen.
Die von M. Stricker in [40] vorgestellten *Boundary Histograms*, die Histogramme der Farbkanten zum Vergleich der Bilder heranziehen, wurden ebenfalls nicht untersucht, da sie nach Aussagen von M. Stricker den Verfahren unter Verwendung der Momente zusammen mit Fuzzy Regions unterlegen sind.

T. Tan und J. Kittler beschreiben in [47] ein Verfahren, das getrennt sowohl räumliche Texturinformation als auch Farbinformation verwendet. Um eine Reduktion des Speicherplatzbedarfs und einer Verringerung der Anzahl der zu vergleichenden Merkmale zu erreichen, schlagen sie eine Approximation des dreidimensionalen Farbhistogramms durch das Produkt der drei Histogramme erster Ordnung, die für die Hauptachsen der dreidimensionalen Farbverteilung berechnet wurden, vor.
Eine Untersuchung dieses Verfahren, auch im Vergleich zum Verfahren von G. Healey und D. Slater, scheint sehr aussichtsreich, da es ebenfalls beleuchtungsinvariante Merkmale verwendet.

Einige Verfahren verwenden gefilterte Bilder zur Berechnung der Merkmale. Das in [32] beschriebene Verfahren vergleicht die Filterantworten von Ableitungen der Gaussfunktion. Die Filterantworten werden dabei wieder als Bilder repräsentiert, so daß das Verfahren eng mit den von K. Oppermann in [28] untersuchten Template–Matching–Verfahren verwandt ist. Eine Integration in die dortigen Verfahren erscheint daher sinvoller als eine getrennte Untersuchung.

Weitere Farbmodelle

Die Untersuchung weiterer, spezieller Farbenräume bzw. Farbmodelle ist ein Ansatz für weitergehende Untersuchungen, da spezielle Farbmodelle für

einige Anwendungen von Vorteil sind. Insbesondere sind Farbmodelle mit möglichst unkorrellierten Kanälen von Interesse[57]. Hierbei sollte jedoch darauf geachtet werden, daß der Aufwand zur Transformation in das Zielfarbmodell in einem realistischen Verhältnis zur erzielbaren Verbesserung der Ergebnisse steht.
Von besonderem Interesse ist das HSV–Modell. Es läßt sich einfacher berechnen als das HSI–Modell, ist diesem aber sehr ähnlich. Nachteil dieses Farbmodelles ist, daß die Werte für V nicht die Helligkeit der Farbe widerspiegeln, da sie auch vom Buntton abhängig sind. Es ist zu überprüfen, ob sich die guten im HSI–Modell erzielten Ergebnisse auch im HSV–Modell erreichen lassen. Das mit dem HSV–Modell verwandte HLS–Modell wäre ebenfalls in die Untersuchung einzubeziehen.

Theoretische Untersuchung der Matching–Verfahren

Wie sich gezeigt hat, ist die Suche nach möglichst guten Verfahren und optimalen Parametern schon bei einer sehr kleinen Datenbank und relativ wenigen Testobjekten sehr aufwendig. Daher ist eine eingehende theoretische Untersuchung der Matching–Verfahren von Interesse, wie sie z.B. M. Stricker et al. in [45, 44, 41] für Farbindizierung (*color indexing*) mit Farbhistogrammen und der Histogrammschnitt–Methode durchgeführt haben.
Ein wichtiger Punkt dabei ist — im Hinblick auf große Datenbanken — die Mächtigkeit der Repräsentation der Farbmerkmale, d.h. die maximal mögliche Anzahl verschiedener Merkmalsausprägungen.

Hierarchische Kombination von Matchingverfahren

Eine hierarchische Kombination von Matchingverfahren benutzt eine schnelles, aber ungenaues Verfahren zur Vorauswahl von Kandidaten bei der Zuordnung der Bilder. Mit Hilfe eines langsameren, dafür aber genaueren Verfahrens wird dann die bestmögliche Zuordnung unter den Kandidaten ermittelt. Im Rahmen von [4] wurden bereits solche Untersuchungen vorgenommen. Dabei wurde die Farbindizierung zur Vorauswahl und ein Template–Matching für die endgültige Zuordnung verwendet. Insgesamt ergab sich so eine merkliche Verbesserung der Ergebnisse. Zeitgleich mit dieser Arbeit wurden von K. Oppermann weitere Untersuchungen durchgeführt [28].

8.3 Erweiterungsmöglichkeiten des Objekterkennungssystems

Neben veränderten Verfahren zum Vergleich der Bilder sind auch Erweiterungen des gesamten Objekterkennungssystems erfolgversprechend für eine Steigerung der Erkennungsrate.

[57] Diese erhält man durch die Karhunen–Loeve–Transformation

Geometrische Rekonstruktion

Wie sich in den Experimenten gezeigt hat, sind Schwächen in der geometrischen Rekonstruktion die Hauptursache für die teilweise schlechten Ergebnisse der Objekterkennung. K. Oppermann hat zeitgleich mit dieser Arbeit [28] alternative Verfahren untersucht und damit eine deutliche Steigerung der Erkennungsrate erzielt. Erste Untersuchungen lassen eine Übertragbarkeit auf die in dieser Arbeit verwendeten Matchingverfahren vermuten. Speziell für die in Abschnitt 7.2 genannten, besonders schlecht erkannten Objekte ist eine Verbesserung der Ergebnisse erreichbar.

Rückkopplung

PolyRec enthält in der implementierten Form keine Rückkopplungen. Es wäre jedoch wünschenswert, die Parameter der Merkmalsextraktion (z.B. Quantisierung des Farbenraums, Schwellenwerte etc.) automatisch den spezifischen Merkmalsausprägungen der zu erkennenden Objekte anzupassen. Ein guter Startwert ließe sich anhand der zur Erstellung der Datenbank verwendeten Objekte ermitteln.

Konsistenzprüfung

Beim Voting könnten die Ergebnisse noch durch Konsistenzprüfungen verbessert werden. Es könnte z.B. eine Überprüfung anhand der Nachbarschaften und der Winkel zwischen benachbarten Flächen erfolgen: Benachbarte rekonstruierte Flächen sollten bei einem korrekt erkannten Objekt auch benachbarten Flächen des Modells zugeordnet werden, und die Winkel zwischen den rekonstruierten Flächen sollten annähernd gleich den Winkeln zwischen den Flächen des Modells sein. Dabei sollte sich die zulässige Abweichung der Winkel an der Genauigkeit des Verfahrens zur geometrischen Rekonstruktion orientieren. Bei dem verwendeten Photometric–Stereo–Verfahren sind das ca. 5–10 Grad.

Affine Transformation der Flächen der Datenbank

Bereits in Abschnitt 6.1 wurde die Transformation der Flächen der Datenbank analog zur perspektivischen Verzerrung der Aufnahme der rekonstruierten Fläche angesprochen. Dadurch wird der durch die in der perspektivischen Aufnahme gröbere Diskretisierung des Orts und der damit einhergehende Informationsverlust simuliert. Eine Verbesserung der Ergebnisse ist jedoch nur zu erwarten, wenn durch die perspektivische Verzerrung und anschließende Rücktransformation der rekonstruierten Fläche signifikante Farbmerkmale gegenüber der Fläche der Datenbank fehlen.
In den Experimenten hat sich jedoch ergeben, daß der Einfluß der Verzerrung und des dadurch bedingten Informationsverlustes sehr gering ist. Erste Experimente mit der Verzerrung der Bilder der Datenbank ergaben keine

Verbesserung der Ergebnisse. Von Bedeutung ist hierbei aber auch die Qualität der bei der geometrischen Rekonstruktion ermittelten Orientierungen. Es sind daher weitere Untersuchungen notwendig.

8.4 Relevanz der Ergebnisse bezüglich großer Datenbanken

Die Ergebnisse wurden — wie bereits dargelegt — mit einer relativ kleinen Datenbank ermittelt. Für eine Beurteilung der Praxistauglichkeit der einzelnen Verfahren sind jedoch Tests mit sehr großen Datenbank (über 1000 Objekte) erforderlich. Es ist zu erwarten, daß sich Ergebnisse bezüglich der Eignung der Verfahren nicht vollständig mit den gemachten Feststellungen decken. Bei einer großen Datenbank tritt auch die Frage, ob die Objektauswahl repräsentativ ist, etwas in den Hintergrund.

8.5 Anwendung für die Suche in Bilddatenbanken

Der Vergleich der rekonstruierten Flächen mit den Flächen der Datenbank entspricht einer Suche in einer Bilddatenbank (*query by example*). Die untersuchten Matching–Verfahren sind daher direkt für eine inhaltsbasierte Suche in Bilddatenbanken verwendbar. Einige der vorgestellten Matching–Verfahren sind auch im Hinblick auf die Suche in Bilddatenbanken entworfen worden [42, 34]. Allerdings sind rein farbbasierte Verfahren nach [43] allein nicht mächtig genug. Sie eignen sich jedoch zur Vorauswahl, da sie im allgemeinen schneller sind als textur- oder geometriebasierte Verfahren. Auch neigen farbbasierte Verfahren zu „false positives“ (fälschlicherweise als ähnlich bewertete Bilder), da völlig verschiedene Bilder durchaus sehr ähnliche Farbverteilungen aufweisen können. Durch nachgeschaltete Verfahren lassen sich diese „false positives“ ausschließen.

Die Abfrage der Datenbank soll in der Regel mehrere Bilder liefern. Bei einem Objekterkennungssystem hingegen soll das zu erkennende Objekt möglichst nur einem Objekt aus der Datenbank zugeordnet werden. Das bedeutet, daß pro rekonstruierter Fläche auch nur maximal ein Bild aus der Datenbank als übereinstimmend betrachtet wird. Daher ist für die Anwendung zur Suche in Bilddatenbanken noch zu untersuchen, wie stark sich die Matching–Werte von besser und schlechter übereinstimmenden Paarungen von Bildern unterscheiden. Ist der Unterschied gering, wird es schwer fallen, einen geeigneten Schwellenwert für die Auswahl der Bilder, die auf die Abfrage zurückgeliefert werden, zu finden. Andererseits sind die Ergebnisse, aufgrund dessen, daß mehrere Bilder extrahiert werden sollen, auch etwas weniger kritisch zu betrachten: Die Reihenfolge im oberen Bereich der Rangliste ist für diese Anwendung weniger von Bedeutung, da die Menge der extrahierten Bilder abschließend durch die abfragende Person bewertet wird.

Bei Bilddatenbanken entfällt die beim Objekterkennungssystem PolyRec notwendige Ausrichtung, da die Bilder sich in der Regel nur durch Rotationen um Vielfache von 90° um den Bildmittelpunkt unterscheiden. Da z.B. das in [42] beschriebene Verfahren speziell für die Suche in Bilddatenbanken entwickelt wurde, ist dort genau diese eingeschränkte Rotationsinvarianz (integriert im Matchingverfahren) realisiert.
Die Testumgebung aus Abschnitt 7.1 ist direkt zum Test von Verfahren zur Suche in Bilddatenbanken verwendbar. Diese Arbeit kann auch als ein Ansatz zur Untersuchung von Verfahren zum Vergleich von Bildern aufgefaßt werden, wobei jedoch andere Bewertungskriterien verwendet werden müßten.

Für die Suche in Bilddatenbanken ist noch eine andere Frage von Interesse: Eignen sich die Verfahren dazu, die Ähnlichkeit von Teilbildern zu bestimmen. Insbesondere wäre eine Abfrage der Art „Welche Bilder enthalten einen Gegenstand wie im Beispielbild?“ von Bedeutung. Dabei müssen die anderen Bereiche der zu findenden Bilder nicht mit dem Abfragebild ähnlich sein.

A Grundlegende mathematische Definitionen

A.1 Koordinatensysteme

Dreidimensionale Koordinaten werden mit Großbuchstaben und zweidimensionale Koordinaten mit Kleinbuchstaben bezeichnet. Das dreidimensionale Koordinatensystem, das für die CAD–Daten und die PTU verwendet wird, ist rechtshändig. Die Z–Achse wird mit der Höhe identifiziert. Das zweidimensionale Koordinatensystem der Linienzeichnungen entspricht dem in der Mathematik üblichen: Der Koordinatenursprung befindet sich links unten. Davon abweichend befindet sich der Koordinatenursprung bei Bildern links oben. Siehe auch Abb. 18.

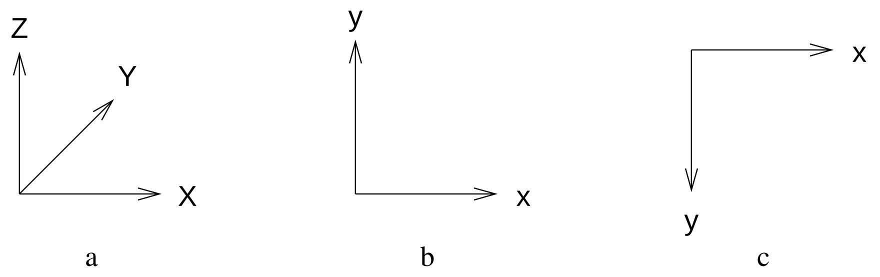

Abbildung 18: Koordinatensysteme: (a) dreidimensionales Objektkoordinatensystem, (b) zweidimensionales K. für Linienzeichnungen, (c) zweidimensionales K. für Bilder

Beim ebenfalls dreidimensionalen Kamerakoordinatensystem, das bei der 2.5D–Rekonstruktion verwendet wird (siehe auch Abschnitt 6.1), sind gegenüber dem obigen dreidimensionalen Koordinatensystem Y– und Z–Achse vertauscht. Es ist also linkshändig.
Als Norm wird immer die euklidische Norm (L2–Norm) verwendet.

A.2 Definitionen

Die *Orientierung* eines planaren Oberflächenstücks bezeichnet die partiellen Ableitungen (im Kamerakoordinatensystem)

$$p = \frac{\partial Z}{\partial X} \qquad \text{und} \qquad q = \frac{\partial Z}{\partial Y} \tag{53}$$

Die Normale eines sichtbaren Oberflächenstückes ist definiert durch den Vektor

$$\vec{n} = \frac{1}{\sqrt{p^2 + q^2 + 1}} \begin{pmatrix} p \\ q \\ -1 \end{pmatrix} \tag{54}$$

B Abbildungen der verwendeten Objekte

Die folgenden Abbildungen zeigen einige der verwendeten Objekte. Es handelt sich dabei um die zur Rekonstruktion der Texturen verwendeten Farbaufnahmen. Die Größenverhältnisse zwischen den Objekten sind nicht maßstabsgetreu.

Abbildung 19: Objekte `Uelzena_Vollmilch_0.5l`, `Reichelt_Schlagsahne_0.5l`, `Reichelt_FettarmeMilch_1l`, `Reichelt_FettarmeMilch_2l`, `Brandenburg_Vollmilch_2l` (von links nach rechts)

Abbildung 20: Objekte `Emzett_FettarmeMilch_1l`, `Erlenhof_Vollmilch_1l`, `Milsani_Vollmilch_1l`, `Penny_Vollmilch_1l`, `Penny_FettarmeMilch_1l` (von links nach rechts)

Abbildung 21: Objekte `Pfanner_EisTee_Ananas_1l`, `Kirberg_Saft_TraubenRot_1l`, `Valensina_Saft_Orange_0.75l`, `Valensina_Saft_Multi_0.75l`, `Valensina_Saft_Orange_0.2l` (von links nach rechts)

Abbildung 22: Objekte `Milsani_Kondensmilch_0.312l`, `Toblerone_400g`, `Milram_Quark_Pfeffer_200g`, `Pril_Supra_0.5l`, `Odol_75ml` (von links nach rechts)

Die nächste Abbildung (Abb. 23) zeigt die in der Datenbank gespeicherten Bilder (im Original farbig) für das Objekt `Mueritz_Buttermilch_1l`.

Abbildung 23: Bilder der Datenbank am Beispiel des Objekts `Mueritz_Buttermilch_1l`

C Aufbau der Datenbank

Die Struktur der Datenbank wurde im wesentlichen unverändert aus den Arbeiten zu [29] übernommen, so daß hier nur eine kurze Darstellung erfolgt. Der Aufbau der Datenbank orientiert sich prinzipiell an der in Abschnitt 2.2.2 beschriebenen ESNI–Struktur. Die Datenbank besteht aus einer Reihe von Dateien, die hierarchisch aufgebaut sind und aufeinander verweisen. Die „Wurzel" ist die Objektdatei. Sie enthält pro Objekt einen Eintrag, in dem der Objektname und Verweise auf eine pol–, rel– sowie sim–Datei angegeben sind (siehe Abb. 24). Durch Verwendung mehrerer Objektdateien können verschiedene Datenbanken angelegt werden, die gemeinsame Objekte besitzen, ohne daß die Daten der Objekte mehrfach abgelegt werden müssen.

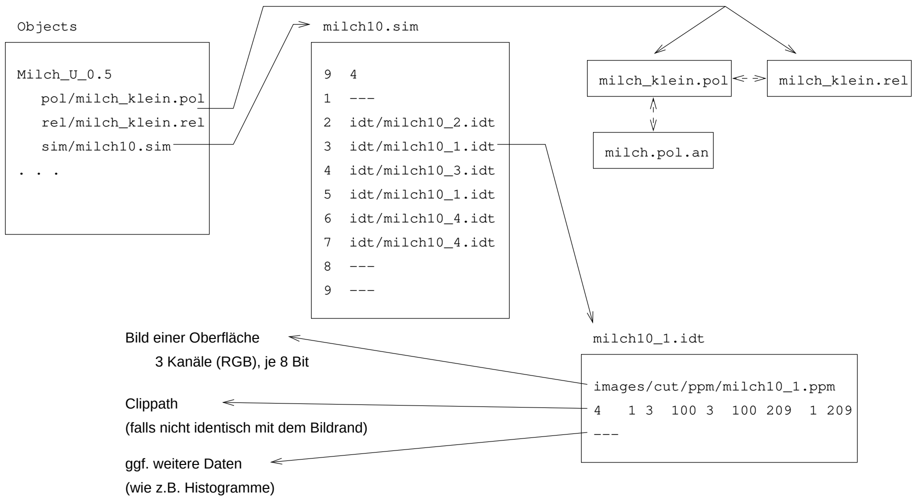

Abbildung 24: Struktur der Datenbank

In den pol–Dateien ist das dreidimensionale, in der Regel vereinfachte Modell des Objekts als CAD–Datensatz gespeichert. Die Objekte wurden größtenteils mit dem Programm `Polygraph` modelliert, dessen Format übernommen wurde.
Aus jeder pol–Datei wird eine rel–Datei generiert. Sie enthält für jede Fläche des Objektmodells die Nachbarflächen und die dazugehörigen Winkel. Diese Daten sind für eine Konsistenzprüfung nach dem Matching vorgesehen (siehe Abschnitt 8.3).
Die sim–Dateien[58] geben die Zuordnung von Objektflächen zu den Bildern in der Datenbank an. Für jedes Bild in der Datenbank existiert eine Beschreibung, seine idt–Datei[59]. Sie verweist auf die Bild–Datei und soll in

[58] sim steht für Surface IMages
[59] idt steht für Image DaTa

einer späteren Ausbaustufe des Objekterkennungssytems auch auf die abgespeicherten Bildmerkmale (Histogramme etc.) verweisen oder diese direkt enthalten. Ist die Objektfläche nicht rechteckig, so muß noch angegeben werden, welcher Teil des Bildes in der Bild–Datei zur Fläche gehört. Dafür wird der sogenannte Clip–Path verwendet, der dann auch in der idt–Datei vermerkt ist. Der Clip–Path ist eine Liste von Koordinaten des Polygons, das die Fläche umschließt.
Die Aufteilung der Datenbank in mehrere Dateien dient nicht nur der Übersichtlichkeit sondern auch der Wiederverwendbarkeit. Objekte mit gleicher Form und Größe aber mit verschiedenen Aufdrucken haben die gleichen pol– und rel–Dateien und nur verschiedene sim–Dateien. Einzelne Bilder können innerhalb einer sim–Datei und auch in verschiedenen sim–Dateien mehrfach verwendet werden. So z.B., wenn verschiedene Flächen eines Objekts den gleichen Aufdruck haben oder wenn mehrere Objekte gleich bedruckte Flächen aufweisen. Ein typisches Beispiel ist der weiße Boden gleich großer Objekte.
Während der Erzeugung der Datenbank sind noch zwei weitere Gruppen von Dateien von Bedeutung: die pol.an–Dateien und die Linedraw–Dateien. Jede pol.an–Datei gehört zu genau einer pol–Datei und gibt an, welche Flächen relevante Aufdrucke besitzen. Die Linedraw–Dateien enthalten die Koordinaten der synthetischen Linienzeichnung, die bei der Segmentierung der Aufnahmen für die Datenbank erforderlich ist.

D Dateiformate

Ein Großteil der vom realisierten Objekterkennungssystem verwendeten und hier beschriebenen Dateiformate ist bereits in [29] verwendet worden und wird hier nur der Vollständigkeit halber erwähnt.

D.1 Objektdatei

Die Objektdatei enthält pro Objekt einen einzeiligen Eintrag. Dieser Eintrag besteht aus dem Objektnamen und den Namen und Pfaden der pol–, rel– und sim–Dateien, jeweils durch Leerraum (White–Space, d.h. Leerzeichen, Tabulatoren etc.) getrennt. Alle Pfade sind relativ zur Wurzel der Datenbank. Dort sollte sich auch die Objektdatei befinden.
Da in der momentanen Programmversion pol– und rel–Dateien nicht benutzt werden, können sie in der Objektdatei durch jeweils drei Minuszeichen („`---`“) ersetzt werden. Ohne rel–Dateien steht jedoch in der eingelesenen Datenbank keine Information über die Beziehungen der Flächen (Nachbarschaft, Winkel etc.) zur Verfügung.

D.2 pol–Datei

Pol–Dateien enthalten die CAD–Datensätze der Objekte. Die Z–Koordinaten sind der Höhe zugeordnet.
Die pol–Dateien liegen im ASCII–Format vor. Alle Einträge sind Dezimalzahlen der Form *nnn.nnnnnn*. Einträge innerhalb einer Zeile werden durch Leerraum getrennt.
In der ersten Zeile werden die Anzahl der Eckpunkte und die Anzahl der Flächen angegeben. In den folgenden Zeilen werden die X–, Y– und Z–Koordinaten der Eckpunkte abgelegt — jeder Punkt in einer eigenen Zeile. Danach folgt die Liste der Flächen mit jeweils zwei Zeilen pro Fläche. In der jeweils ersten Zeile steht die Anzahl der Eckpunkte der Fläche. Danach folgen die Nummern der Eckpunkte in der Eckenliste. Die Zählung beginnt bei 1. In der jeweils zweiten Zeile stehen die Kantenlängen in Millimetern.

D.3 pol.an–Datei

Die pol.an–Datei ist eine Annotation zur pol–Datei für das 3D–Texture–Scanning zur automatischen Erzeugung der Datenbank. Sie gibt an, welche Flächen sinnvoll verwertbare Aufdrucke haben und welche Flächen gleichartige Aufdrucke aufweisen.
Die erste Zeile enthält die Anzahl der Flächen. In den folgenden Zeilen stehen jeweils die Nummern der Flächen (aufsteigend, von 1 beginnend) gefolgt von Leerraum und einer Zahl bzw. einem Plus- oder Minuszeichen. Ein Zahl gibt eine Referenzfläche an, die einen gleichartigen Aufdruck besitzt. Die Verknüpfung erfolgt nur in einer Richtung, d.h. bei der Referenzfläche

selbst steht keine Zahl (sondern ein Pluszeichen). Ein Pluszeichen bedeutet, daß diese Fläche einen verwertbaren Aufdruck besitzt und es entweder keine Fläche mit gleichartigem Aufdruck gibt oder es sich um eine Referenzfläche handelt. Ein Minuszeichen kennzeichnet Flächen ohne verwertbaren Aufdruck, z.B. die Bodenflächen vieler Tetrapaks.

D.4 rel–Datei

Die rel–Dateien liegen ebenfalls im ASCII–Format vor. Alle Einträge sind natürliche Zahlen. Einträge innerhalb einer Zeile werden durch Leerraum getrennt.
In der ersten Zeile steht die Anzahl der Flächen. Sie muß identisch sein mit der Anzahl der Flächen in der zugehörigen pol–Datei. Ab der zweiten Zeile beginnt die Liste der Flächen mit jeweils zwei Zeilen pro Fläche. In der jeweils ersten Zeile stehen die Nummer der Fläche laut pol–Datei[60], die Anzahl der angrenzenden Flächen (gleich der Anzahl der Eckpunkte in der pol–Datei) und die Nummern (laut pol–Datei) der angrenzenden Fläche. In der jeweils zweiten Zeile steht zu jeder angrenzenden Fläche der Winkel, den die Oberflächennormalen der Fläche und der angrenzenden Fläche bilden.

D.5 Linedraw–Datei

Die Linedraw–Datei ähnelt im Aufbau sehr einer pol–Datei, mit dem Unterschied, daß sie nur 2D–Koordinaten enthält. In der ersten Zeile werden ebenfalls die Anzahl der Eckpunkte und die Anzahl der Flächen angegeben. In den folgenden Zeilen werden die x– und y–Koordinaten der Eckpunkte abgelegt — jeder Punkt in einer eigenen Zeile. Danach folgt die Liste der Flächen mit jeweils einer Zeile pro Fläche: Zuerst steht die Anzahl der Eckpunkte der Fläche. Danach folgen die Nummern der Eckpunkte in der Eckenliste. Die Zählung beginnt bei 1.
Das Koordinatensystem ist die XZ–Ebene des entsprechend der Normale der Basisfläche gedrehten Koordinatensystems der entsprechenden pol–Datei.

D.6 sim–Datei

Auch die sim–Dateien liegen im ASCII–Format vor. Einträge innerhalb einer Zeile werden durch Leerraum getrennt.
In der ersten Zeile einer sim–Datei stehen die Anzahl der Flächen und die Anzahl verschiedener Bilder. (Da bei gleichartig bedruckten Flächen nur ein Bild für beide Flächen gespeichert wird, wird in der sim–Datei ein Bild mehrfach eingetragen. Es wird jedoch nur das erste Auftreten eines Bildes in der sim–Datei gezählt.)

[60] Die Numerierung beginnt mit 1. Die Flächen müssen in aufsteigender Reihenfolge angegeben werden.

In den folgenden Zeilen erfolgt eine Zuordnung von Flächen zu Bildern, jeweils eine Fläche pro Zeile. Die Zeile wird mit der Nummer der Fläche laut pol–Datei eingeleitet. (Die Flächen müssen in aufsteigender Reihenfolge angegeben werden.) Danach wird das Bild durch Angabe des Namens und Pfads einer idt–Datei zugeordnet. Existiert zu einer Fläche kein Bild in der Datenbank, so werden stattdessen drei Minuszeichen (ohne Leerzeichen o.ä. dazwischen) eingetragen.

D.7 Clip–Path–Datei

Die Clip–Path–Datei enthält nur eine Zeile mit der Angabe der Punkte des sogenannten Clip–Paths. Die Angaben bestehen aus der Anzahl der Punkte gefolgt von den durch Leerzeichen getrennten Koordinatenangaben. Dabei sind x– und y–Anteil wiederum durch Leerzeichen getrennte natürliche Zahlen.
Die Clip–Path–Datei dient nur als Zwischenstufe bei der Erzeugung einer idt–Datei.

D.8 idt–Datei

Eine idt–Datei ist die Beschreibung zu einem Datenbankbild. Sie besteht aus vier Zeilen. In der ersten Zeile sind der Name und der Pfad der Bild–Datei und in der dritten Zeile der Name und der Pfad der Histogramm–Datei angegeben. Die Pfade sind, wie bereits erwähnt, relativ zum Pfad der Objektdatei. Ist zu dem Bild kein Histogramm gespeichert, so werden in der dritten Zeile drei Minuszeichen (ohne Leerzeichen o.ä. davor, dazwischen oder danach) angegeben.
In der zweiten Zeile wird der sogennante Clip–Path angegeben. Der Aufbau entspricht dem der Clip–Path–Datei. Ist der Clip–Path identisch mit dem Rand des Bildes in der Bild–Datei, so muß er nicht angegeben werden. Stattdessen können drei Minuszeichen (wie oben) eingetragen werden.
In der vierten Zeile werden durch Leerzeichen getrennt „Main Hue“ und „Average Hue“ angegeben. „Main Hue“ ist eine natürliche Zahl im Bereich von 0 bis 255 und „Average Hue“ eine reelle Zahl im Bereich von 0.0 bis 255.0. Die beiden Werte können zur Vorauswahl beim Matching verwendet werden (siehe dazu Abschnitt 6.2.2).
Hinweis: In einer früheren Programmversion war die vierte Zeile der idt–Datei noch nicht vorhanden. Daher kann zur Verwendung älterer Datenbanken auch eine Programmversion für dreizeilige idt–Dateien generiert werden.

D.9 Bild

Als Format für die Bild–Dateien wurde PPM (für Farbbilder) bzw. PGM (für Grauwertbilder) gewählt. Farbbilder werden dabei als Bilder mit drei

Grauwert–Kanälen betrachtet, die den Rot–, Grün– und Blau–Anteil repräsentieren. Pro Kanal wird eine maximale Auflösung von 8 Bit zugelassen.
Für beide Formate ist eine binäre und eine ASCII–Variante definiert. Von den erstellten Programmen wird aber nur die binäre Variante unterstützt, da das Laden und Speichern so schneller möglich ist. Da das Binärformat byteorientiert ist, ist es auch (weitgehend) maschinenunabhängig.
Eine Bild–Datei beginnt mit einem aus einem ASCII–Text bestehenden Header. Alle mit `#` beginnenden Zeilen sind dabei Kommentare und werden ignoriert. Die Angabe der Zeilennummern im nachfolgenden Text bezieht daher Kommentarzeilen *nicht* mit ein.
In der ersten Zeile steht der sogenannte Magic Code. Er ist für Grauwertbilder `P5` und für Farbbilder `P6`.
In der zweiten Zeile geben zwei natürliche Zahlen die Bildgröße an, zuerst die Breite (x) und dann die Höhe (y). Die beiden Zahlen sind durch Leerraum voneinander getrennt.
In der dritten Zeile gibt eine natürliche Zahl den maximalen Wert an, der in einem Kanal auftreten kann. Hier darf dieser Wert 255 (aufgrund der maximalen Auflösung von 8 Bit) nicht überschreiten.

Literatur

[1] R. Bajcsy, S.W. Lee und A. Leonardis. Color image segmentation and color constancy. In *Perceiving, Measuring and Using Color* (Februar 1990), Bd. 1250 von *Proceedings of the SPIE/SPSE Symposium on Electronic Image Science and Technology*, S. 245–255.

[2] Gunter Bellaire. Komponenten der 3D–Objekterkennung. Technischer Bericht 94–33, TU Berlin, 1994.

[3] Gunter Bellaire. *3D–Objektidentifizierung. Betrachterabhängige Modelle für Repräsentation und Erkennung.* Dissertation, TU Berlin, Mai 1996.

[4] Gunter Bellaire, Karsten Schlüns, Kjell Oppermann und Wolfram Schimke. Matching Using Object Models Generated From Photometric Stereo Images. In *Machine Vision Applications in Industrial Inspection IV* (Januar/Februar 1996), Bd. 2665 von *SPIE Proceedings Series*, S. 70–81.

[5] A.E. Cawkell. Selected aspects of image processing and management: review and future prospects. *Journal of Information Science 18*, 3 (1992), S. 179–192.

[6] Michael D'Zmura. Color constancy: surface color from changing illumination. *Journal of the Optical Society of America A*, 9 (1992), S. 490–493.

[7] Micheal D'Zmura und Geoffrey Iverson. Color constancy. I. Basic theory of twostage linear recovery of spectral descriptions for lights and surfaces. *Journal of the Optical Society of America A*, 10 (1993), S. 2148–2165.

[8] Michael D'Zmura und Geoffrey Iverson. Color constancy. II. Results for two-stage linear recovery of spectral descriptions for lights and surfaces. *Journal of the Optical Society of America A*, 10 (1993), S. 2166–2180.

[9] Michael D'Zmura und Geoffrey Iverson. Color constancy. III. General linear recovery of spectral descriptions for lights and surfaces. *Journal of the Optical Society of America A*, 11 (1994), S. 2389–2400.

[10] Michael D'Zmura und Geoffrey Iverson. A formal approach to color constancy: the recovery of surface and light source spectral properties using bilinear models. In *Recent Progress in Mathematical Psychology*, C. Dowling, F. Roberts, und P. Theuns, Eds. Hillsdale, NJ: Lawrence Erlbaum Associates, 1995.

[11] MICHAEL D'ZMURA UND P. LENNIE. Mechanisms of color constancy. *Journal of the Optical Society of America A*, 3 (1986), S. 1662–1672.

[12] MICHAEL D'ZMURA UND ANOOP MANGALICK. Detection of contrary chromatic change. *Journal of the Optical Society of America A*, 11 (1994), S. 543–546.

[13] JR. FRED W. BILLMEYER UND MAX SALTZMAN. *Principles of Color Technology*. Interscience Publishers (a division of John Wiley & Sons), New York, 1966.

[14] BRIAN V. FUNT UND MARK S. DREW. Color Constancy Computation in Near–Mondrian Scenes using a Finite Dimensional Linear Model. In *Proc. CVPR* (1988), S. 544–549.

[15] BRIAN V. FUNT UND GRAHAM D. FINLAYSON. Color Constant Color Indexing. Technischer Bericht 91–09, Simon Fraser University, Vancouver, School of Computing Science, 1991.

[16] RON GERSHON, ALLAN D. JEPSON UND JOHN K. TSOTSOS. From [R,G,B] to Surface Reflectance: Computing Color Constant Descriptors in Images. In *Proc. International Joint Conference on Artifical Intelligence* (1987), S. 755–758.

[17] KLAUS GWINNER UND ARNO MITRITZ. Objekterkennung auf der Grundlage von Surface Normal Images. Projektbericht: Modellbasierte Objekterkennung, TU Berlin, Oktober 1994.

[18] GLENN E. HEALEY UND DAVID ALAN SLATER. Global color constancy: recognition of objects by use of illumination-invariant properties of color distributions. *Journal of the Optical Society of America*, 11 (1994), S. 3003–3010.

[19] R.W.G. HUNT. *Measuring colour*. Ellis Horwood Limited, Chichester, 1987.

[20] GEOFFREY IVERSON UND MICHAEL D'ZMURA. Color constancy: spectral recovery using trichromatic bilinear models. In *Geometric Representations of Perceptual Phenomena: Articles in Honor of Tarow Indow's 70th Birthday*, R.D. Luce, M. D'Zmura, D.D. Hoffman, G. Iverson, und K. Romney, Eds. Hillsdale, NJ: Lawrence Erlbaum Associates, 1994.

[21] GEOFFREY IVERSON UND MICHAEL D'ZMURA. Criteria for color constancy in trichromatic bilinear models. *Journal of the Optical Society of America A*, 11 (1994), S. 1970–1975.

[22] JEONG-YEOP KIM UND YEONG-HO HA. Color image enhancement using color constancy based on modified IHS coordinate system. In *Intelligent Robots and Computer Vision XII: Algorithms and Techniques* (August 1993), David P. Casasent, Ed., Bd. 2055 von *SPIE Proceedings Series*, S. 359–368.

[23] REINHARD KLETTE, ANDREAS KOSCHAN UND KARSTEN SCHLÜNS. *Computer Vision*. Vieweg, Braunschweig/Wiesbaden, 1996.

[24] ANDREAS KOSCHAN UND KARSTEN SCHLÜNS. Grundlagen und Voraussetzungen für die Digitale Farbbildverarbeitung. Technischer Bericht 94–14, TU Berlin, März 1994.

[25] J. MATAS, R. MARIK UND J. KITTLER. On Representation and Matching of Multi–Coloured Objects. In *Fifth International Conference on Computer Vision* (Juni 1995), S. 726–732.

[26] KENJI NAGAO. Recognizing 3D Objects Using Photometric Invariant. In *Fifth International Conference on Computer Vision* (Juni 1995), S. 480–487.

[27] SHREE K. NAYAR, SAMEER A. NENE UND HIROSHI MURASE. Real–Time 100 Object Recognition System. In *Image Understanding Workshop* (1996), S. 1223–1227.

[28] KJELL OPPERMANN. Segmentierung und Matching bildhafter 2.5D–Daten. Diplomarbeit, TU Berlin, September 1996.

[29] KJELL OPPERMANN UND WOLFRAM J. SCHIMKE. Automatische Erkennung von Objekten mit dem „Grünen Punkt“. Projektbericht: Modellierung und Extraktion von Objekten aus Entfernungs- und Schichtdaten, TU Berlin, Januar 1996.

[30] JONG HOON PARK, TAE GYU CHANG UND JONG SOO CHOI. Three-Dimensional Object Representation and Recognition based on Surface Normal Images. *Pattern Recognition 26*, 6 (1993), S. 913–921.

[31] WILLIAM K. PRATT. *Digital Image Processing*, 2. Aufl. John Wiley & Sons, New York, 1991.

[32] RAJESH P.N. RAO UND DANA H. BALLARD. Object Indexing using an Iconic Sparse Distributed Memory. In *Fifth International Conference on Computer Vision* (Juni 1995), S. 24–31.

[33] MANFRED RICHTER. *Einführung in die Farbmetrik*, 2. Aufl. de Gruyter, Berlin, 1980.

[34] Rick Rickman und John Stonham. Content–based image retrieval using colour tuple histograms. Technischer Bericht, Dept. Electronic Engineering, Brunel University, Uxbridge, Middx, UK, 1996.

[35] J. Rubner und K. Schulten. A Regularized Approach to Color Constancy. *Biological Cybernetics 61* (1989), S. 29–36.

[36] Rafał Sałustowicz, Christian Krüger und Alexander Bachem. Analyse von Farbhistogrammen zur 3D–Gestaltserkennung. Projektbericht, TU Berlin, Fachbereich 13 (Informatik), 1993.

[37] Marian Sänger. Rekonstruktion von 3D–Objekten aus lasergenerierten Entfernungsbildern. Studienarbeit, CV–Bericht 25, TU Berlin, Fachbereich 13 (Informatik), Gruppe Computer Vision, August 1994.

[38] Benjamin Singer und Michael D'Zmura. Contrast gain control. A bilinear model for chromatic selectivity. *Journal of the Optical Society of America A*, 12 (1995), S. 667–685.

[39] David Slater und Glenn Healey. Combining Color and Geometric Information for the Illumination Invariant Recognition of 3-D Objects. In *Fifth International Conference on Computer Vision* (Juni 1995), S. 563–567.

[40] Markus A. Stricker. Color and Geometry as Cues for Indexing. Technischer Bericht 92–22, Department of Computer Science, The University of Chicago, November 1992.

[41] Markus A. Stricker. Bounds for the discrimination power of color indexing techniques. In *Storage and Retrieval for Image and Video Databases II* (Februar 1994), Bd. 2185 von *SPIE Proceedings Series*, S. 15–24.

[42] Markus A. Stricker und Alexander Dimai. Color Indexing with Weak Spatial Constraints. In *Storage and Retrieval for Image and Video Databases IV* (Februar 1996), Bd. 2670 von *SPIE Proceedings Series*.

[43] Markus A. Stricker und Markus Orengo. Similarity of Color Images. In *Storage and Retrieval for Image and Video Databases III* (Februar 1995), Bd. 2420 von *SPIE Proceedings Series*, S. 381–392.

[44] Markus A. Stricker und Michael J. Swain. The Capacity and the Sensivity of Color Histogram Indexing. Technischer Bericht 94–05, Department of Computer Science, The University of Chicago, März 1994.

[45] Markus A. Stricker und Michael J. Swain. The Capacity of Color Histogram Indexing. In *Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR)* (Juni 1994), S. 704–708.

[46] Michael J. Swain und Dana H. Ballard. Color Indexing. *International Journal of Computer Vision 7*, 1 (1991), S. 11–31.

[47] T.S.C. Tan und J. Kittler. Colour texture analysis using colour histogram. In *Vis. Image Signal Process.* (Dezember 1994), Bd. 141 von *IEE Proceedings*, S. 403–412.

[48] Gabriel Taubin und David B. Cooper. Object recognition based on moment (or algebraic) invariants. In *Geometric Invariance in Computer Vision*, Joseph L. Mundy und Andrew Zisserman, Eds. MIT Press, Cambridge, Mass., 1992, S. 375–397.

[49] G. Wyszecki. *Farbsysteme.* Musterschmidt–Verlag, Göttingen, 1960.

[50] G. Wyszecki und W.S. Stiles. *Color Science: Concepts and Methods, Quantitative Data and Formulae*, 2. Aufl. John Wiley & Sons, New York, 1982.