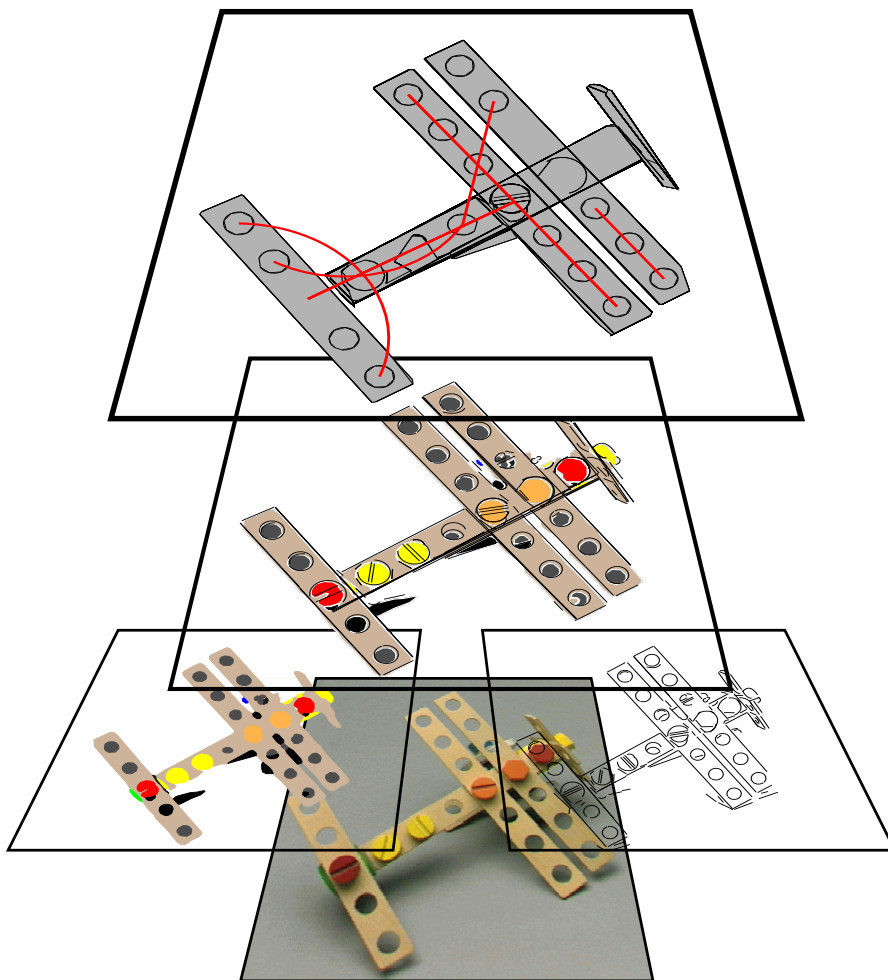


*Hierarchisches Perzeptives Gruppieren
mit Integration dualer
Bildbeschreibungen*



Daniel Schlüter

Dipl.-Inform. Daniel Schlüter
AG Angewandte Informatik
Technische Fakultät
Universität Bielefeld
email: dschluet@techfak.uni-bielefeld.de

Elektronische Version der genehmigten Dissertation
zur Erlangung des akademischen Grades
Doktor der Ingenieurwissenschaften (Dr.-Ing.).
Von Daniel Schlüter am 11. Juli 2001
der Technischen Fakultät an der Universität Bielefeld vorgelegt.
Am 29. Oktober 2001 verteidigt und genehmigt.

Prüfungsausschuß:

Prof. Dr.-Ing. Stefan Posch, Universität Halle-Wittenberg
Prof. Dr. Helge Ritter, Universität Bielefeld
HD Dr. Enno Ohlebusch, Universität Bielefeld
Dr.-Ing. Sven Wachsmuth, Universität Bielefeld

Hierarchisches Perzeptives Gruppieren mit Integration dualer Bildbeschreibungen

Dissertation zur Erlangung des akademischen Grades

Doktor der Ingenieurwissenschaften (Dr.-Ing.)

vorgelegt von

Daniel Schlüter

an der Technischen Fakultät
der Universität Bielefeld

11. Juli 2001

Danksagung

Die Anfertigung dieser Arbeit wäre ohne die Unterstützung von vielen Kollegen und Freunden nur schwer möglich gewesen. An dieser Stelle möchte ich mich dafür bedanken.

Mein besonderer Dank gilt Prof. Stefan Posch, der mich im Laufe unserer Zusammenarbeit nicht nur in wissenschaftlichen Belangen stets hervorragend unterstützt und inspiriert hat. Trotz der Belastung durch den Aufbau einer eigenen Arbeitsgruppe an der Universität Halle-Wittenberg war er stets erreichbar und fand immer Zeit für konstruktive Diskussionen, die letztendlich zum Gelingen der vorliegenden Arbeit entscheidend beigetragen haben. Prof. Helge Ritter danke ich für seine hervorragende Arbeit als Leiter des Projektes, in dessen Rahmen diese Dissertation entstanden ist, und für seine Bereitschaft die vorliegende Niederschrift zu begutachten.

Bedanken möchte ich mich auch bei Birgit Möller, die mich in ihrer Funktion als Hilfskraft bei der Realisierung verschiedener Module in hohem Maße unterstützt hat. Darüber hinaus wären ohne die Diplomanden Björn Cordes und Sebastian Lang einige wesentliche Bestandteile der vorliegenden Arbeit nicht in der entsprechenden Ausführlichkeit zu bearbeiten gewesen. Auch Ihnen gilt daher mein Dank für die Zusammenarbeit.

Im Übrigen bleibt die gute Zusammenarbeit mit allen Kollegen aus der Arbeitsgruppe Angewandte Informatik der Technischen Fakultät an der Universität Bielefeld in sehr guter Erinnerung. Insbesondere möchte ich mich hier bei Franz Kummert bedanken, der über die gemeinsame Zeit mehr als nur ein Kollege war. Er stand stets für kritische Diskussionen zur Verfügung und hat mich bei der Erstellung dieser Arbeit mit seinen Vorschlägen konstruktiv unterstützt. Gleiches gilt für Prof. Gerhard Sagerer, der als Leiter der Arbeitsgruppe für hervorragende Rahmenbedingungen gesorgt hat und neben der Förderung der wissenschaftlichen Entwicklung auch das Menschliche nie vernachlässigt hat. Bedanken möchte ich mich ebenfalls bei meinen Bürokollegen Elke Braun und Markus Wienecke für die vielen konstruktiven Gespräche und Diskussionen. Für die Unterstützung in technischen Fragen danke ich Gernot A. Fink, sowie Lisabeth van Iersel für ihren Beistand nicht nur in formalen Angelegenheiten.

Als Projektpartner und Freund danke ich Robert Rae und allen weiteren Mitgliedern des SFBs, die mehr oder weniger zur Realisierung dieser Arbeit beigetragen haben.

Abschließend möchte ich noch Familie Lars Remmerbach und Birgit Horstmann erwähnen, die mir häufig den nötigen Rückhalt gegeben haben. Ganz besonders danke ich meinen Eltern für ihre permanente Unterstützung und Ermutigung während meines Werdegangs.

Inhaltsverzeichnis

1. Einleitung	1
1.1. Motivation	2
1.2. Zielsetzung und Gliederung der Arbeit	5
2. Bildsegmentierung	7
2.1. Kontursegmentierung	8
2.1.1. Detektion	10
2.1.2. Approximation	14
2.1.3. Approximation nach Leonardis	16
2.1.4. Ergebnisse der Kontursegmentierung	22
2.2. Regionensegmentierung	25
2.2.1. Pixelweise Klassifikation	27
2.2.2. Island Growing	28
2.2.3. Ergebnisse der Regionensegmentierung	29
3. Perzeptives Gruppieren	33
3.1. Gesetzmäßigkeiten zur Gruppenbildung	33
3.1.1. Gestaltgesetze	34
3.1.2. Wechselwirkungen und Kausalität	37
3.2. Gruppierung in der Bildanalyse	39
3.2.1. Verfahren in der Dimension 2D	42
3.2.2. Verfahren für Bildsequenzen	48
4. Konturbasiertes hierarchisches Gruppieren	51
4.1. Gruppierungshierarchie	51
4.2. Gruppierungen mit linearem Charakter	55
4.2.1. Bereiche perzeptiver Aufmerksamkeit	56
4.2.2. Hypothesisierung	61
4.3. Geschlossene Figuren	66
4.4. Gruppierung von geschlossenen Figuren	69
4.4.1. Nähe	71
4.4.2. Ähnlichkeit	80
4.4.3. Klassifikation und Kombination	87
4.4.4. Gute Fortsetzung	89

4.4.5. Rekursive Gruppierung geschlossener Figuren	93
5. Integration regionen- und konturorientierter Beschreibung	97
5.1. Motivation	97
5.2. Kombination von Konturgruppen und Regionen	99
5.2.1. Allgemeine Konzeption einer Zuordnung	100
5.2.2. Einheitliche Repräsentation	103
5.2.3. Merkmale und Klassifikation einer Zuordnung	104
5.3. Konturbasiertes Gruppieren mit Regionenintegration	108
5.3.1. Bearbeitungsreihenfolge gruppierender Konturmodelle	108
5.3.2. Inkrementelles perzeptives Gruppieren	110
5.3.3. Regionenbasierte Proximitäten	112
5.4. Objekterkennung durch Integration	116
5.4.1. Hybride Objekterkennung	116
5.4.2. Objektfokussierte Kontur-Integration	120
6. Bewertung und Selektion mit Markov Random Fields	125
6.1. Markov Random Fields	126
6.1.1. Aufbau und Eigenschaften	126
6.1.2. Energieminimierung	129
6.1.3. Einsatzgebiete in der Bildanalyse	132
6.2. MRF zur Bewertung von Gruppierungshypothesen	133
6.2.1. Gruppierungsgraph und Labelmenge	134
6.2.2. Kontextbezüge zwischen Gruppierungen	139
6.2.3. Datenabhängigkeit	143
6.2.4. Energieminimierung mit HCF	148
6.3. MRF zur Kontursegmentierung in Bildfolgen	149
6.3.1. Konturgraph und Labelmenge	151
6.3.2. Modellierung des a priori und a posteriori Wissens	152
6.3.3. Parameterbestimmung	155
6.3.4. Detektion korrespondierender Konturmodelle	157
7. Ergebnisse und Diskussion	161
7.1. Generierung von Gruppierungshypothesen	162
7.2. Bewertung von Gruppierungshypothesen	173
7.3. Integrierte Objekterkennung	178
7.4. Resümee und Ausblick	179
8. Zusammenfassung	181
Literaturverzeichnis	185
A. Bilder und Bildsegmentierung	197

B. Gruppierungshypothesen und Kombination von Regionen und Konturgruppen	201
C. Signifikante Gruppierungshypothesen	215
D. Integrierte Objekterkennung	221

1. Einleitung

Betrachtet ein Mensch die Szene in Abbildung 1.1, so kann er spontan die abgebildeten Objekte auf dem homogenen Untergrund erkennen. Ohne spezielles Wissen über die Objekte zu besitzen, können wir beispielsweise die Anordnung von zwei holzfarbenen Leisten zu einem Kreuz erkennen. Im oberen Teil des Bildes erkennen wir verschiedene farbige Objekte, die zum Teil verdeckt sind und scheinbar auf irgendeine Art miteinander verbunden sind. Implizit erzeugen wir auch Ergänzungen, die Schlußfolgerungen über die reale Form und Struktur der weniger verdeckten Objekte ermöglichen. Wir nehmen also räumliche Strukturen und Relationen zwischen den Objekten oder Objektteilen wahr und ergänzen fehlende visuelle Information weit möglichst, um letztendlich zu einer Interpretation des Wahrgenommenen zu kommen.

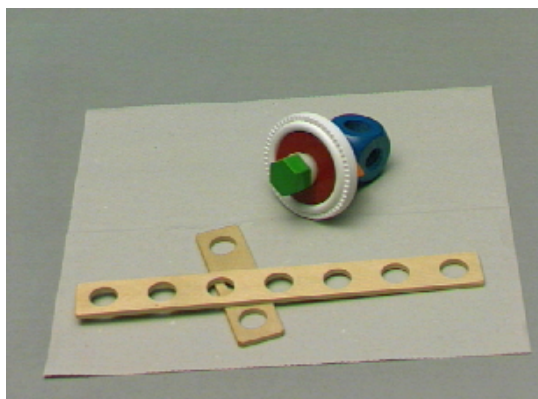


Abbildung 1.1.: Szene mit Objekten aus der Spielzeugwelt von baufix®¹.

Das Ziel eines automatischen Bildanalyse-Systems besteht darin, die relevanten Aspekte dieser visuellen perzeptiven Fähigkeiten algorithmisch zu modellieren, und so automatisch eine Beschreibung der abgebildeten Szene zu erlangen. In welcher Form diese Beschreibung vorliegen soll, hängt dabei stark von der Aufgabenstellung ab. Häufig wird eine Erkennung und Benennung bekannter Objekte angestrebt und deren Beschreibung beispielsweise durch Angabe ihrer Position, Lage und Farbe gewünscht. Dabei ist die Detektion der inhärenten Strukturen von Objekten von starkem Interesse. Gerade bei künstlichen Objekten erleichtert die Regelmäßigkeit und Einfachheit der Objekte und

¹baufix® ist ein eingetragenes Warenzeichen der Heinz H. Lorenz KG, D-82538 Geretsried.

die Integration verschiedener Objekteigenschaften deren Identifikation und Erkennung. Darüber hinaus führt eine regelmäßige Anordnung von Objekten oder Objektbestandteilen zu einer weiteren Strukturierung der betrachteten Szene und unterstützt eine Beschreibung derselben.

Die vorliegende Arbeit befaßt sich mit der Modellierung und algorithmischen Realisierung wesentlicher Aspekte dieser Strukturierungs- oder auch Gruppierungsaufgaben. Sie ist jedoch nicht nur auf die Erkennung von Objektstrukturen fokussiert, wenn auch in dem gegebenen Rahmen vornehmlich diese Strukturen vorherrschen. Die Arbeit ist Bestandteil eines Gesamtsystems zur Bildanalyse, das wiederum Bestandteil eines größeren Forschungsprojekts ist, dem Sonderforschungsbereich 360 „Situierete künstliche Kommunikatoren“ an der Universität Bielefeld. Das Ziel dieses Projektes ist die Entwicklung von künstlichen Systemen, die in gegebenen Situationen eine Kommunikation und Interaktion ermöglichen, die durch das Muster natürlicher Kommunikation beim Menschen motiviert ist. Als Entwicklungsziel wurde hierzu die Bewältigung einer Konstruktionsaufgabe gewählt, in der das System mit den Elementen eines Holzbaukastens **baufix**[®] (vergleiche Abbildung 1.1) und durch Instruierung eines Menschen ein Spielzeugflugzeug bauen soll. Hierzu werden die in dieser Arbeit vorgestellten Techniken in den visuellen Interpretationsteil des künstlichen Systems integriert [Sag01]. Dieser visuelle Interpretationsteil konzentriert sich dabei primär auf die Erkennung der **baufix**[®] Objekte und in weiteren Modulen auf die Detektion gestikbasierter Anweisungen. Die Anwendbarkeit der hier vorgestellten Techniken ist dabei auch auf andere Anwendungsfelder ausgelegt.

1.1. Motivation

Bereits in den zwanziger Jahren des letzten Jahrhunderts untersuchten Gestaltpsychologen die Wahrnehmungsmechanismen des Menschen und postulierten, daß der Mensch verschiedene wahrgenommene Bildelemente unter bestimmten Beziehungen zu „figuralen Einheiten“ oder *Gestalten* zusammenfaßt. So können Figuren beispielsweise ihre Farbe oder Größe ändern, ohne daß sich die Wahrnehmung ihrer Gestalt ändert. Ein Quadrat wird nach einer Vergrößerung immer noch mit der Gestalt eines Quadrates wahrgenommen. Die Psychologen formulierten aus diesen Beobachtungen Gesetzmäßigkeiten, unter denen der Mensch scheinbar Teile zu Ganzheiten gruppiert. Diese Gesetze beschreiben, wie Beziehungen zwischen Bildelementen, beispielsweise aufgrund von Nähe oder Ähnlichkeit, zur Strukturbildung ausgenutzt werden. Dieser Strukturbildungsprozeß trägt dabei entscheidend zur Objektidentifizierung und -erkennung beim Menschen bei und ist allgemein unter dem Begriff des *perzeptiven Gruppierens* oder *perzeptiver Organisation* bekannt.

Für die automatische Beschreibung einer Szene durch ein Bildanalyzesystem besteht zunächst die Problematik, daß die Repräsentation eines Bildes im Computerspeicher nur durch eine Zahlenmatrix erfolgt, in der die Werte die Helligkeit oder die Farbe eines diskreten Bildpunktes bzw. Sensorsignals beschreiben. Daraus ergibt sich die Aufgabe, aus

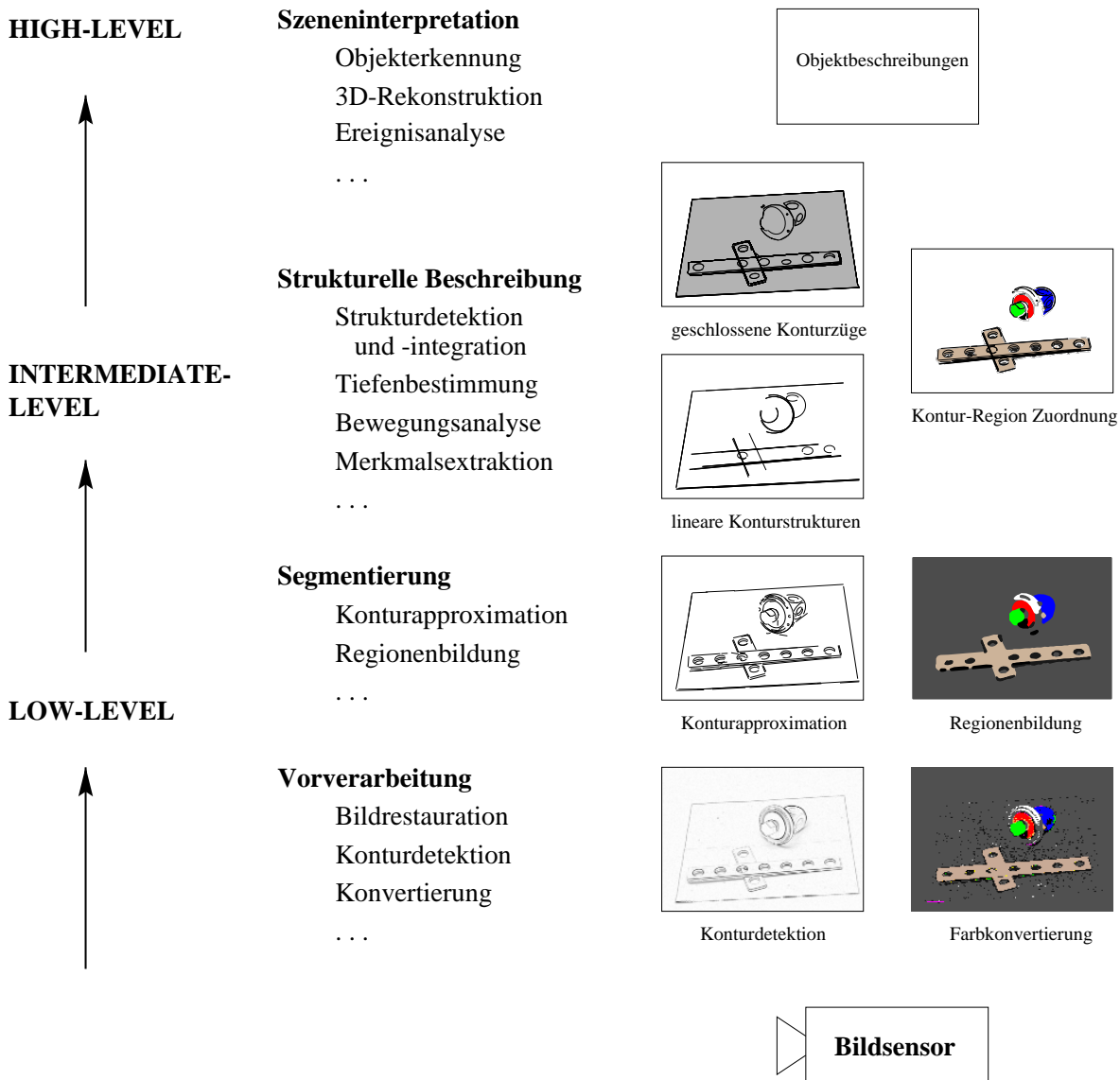


Abbildung 1.2.: Exemplarischer Aufbau eines Bildanalyse-systems. Ausgehend von einem Sensor zur Bildaufnahme gliedert sich ein System in mehrere Ebenen mit Arbeitsschritten, die eine zunehmend abstraktere Beschreibung der abgebildeten Szene erzeugen. Im linken Teil sind die einzelnen Ebenen aufgeführt, wobei je nach Anwendung die Trennung zwischen LOW-LEVEL und INTERMEDIATE-LEVEL fließend ist. In der Mitte sind prinzipielle Arbeitsschritte und exemplarisch einige Verarbeitungsprozesse angegeben. Im rechten Teil sind für das Bild aus Abbildung 1.1 die Ergebnisse einiger Verarbeitungsprozesse visualisiert. Dabei werden mit der vorliegenden Arbeit primär die Prozesse der Strukturdetektion und -integration für eine strukturelle Beschreibung realisiert.

dieser Darstellung eine für die Interpretationsaufgabe adäquate Beschreibung und Strukturierung der Szene zu gewinnen. Zur Lösung dieser Aufgabe werden Bildanalyzesysteme zumeist in mehrere Ebenen mit wachsendem Abstraktionsgrad der Bildbeschreibung unterteilt. Diese Einteilung wird im folgenden an dem Bildanalyzesystem, das dieser Arbeit zugrundeliegt, exemplarisch erläutert (siehe Abbildung 1.2).

In der untersten Ebene (*low-level*) werden in einem ersten Schritt die digitalisierten Sensordaten **vorverarbeitet**. So wird hier beispielsweise eine optionale Bildrestauration vorgenommen, um gestörte Bildsignale zu rekonstruieren, die z.B. durch Unzulänglichkeiten der Aufnahmeapparatur entstanden sind. Aber auch die Ermittlung von starken Helligkeitsunterschieden in den Bilddaten (Konturdetektion) oder eine Konvertierung des Farbraumes sind in dieser Ebene anzusetzen.

Auf Basis dieser vorverarbeiteten Bildsignale wird in einem nächsten Schritt eine **Segmentierung** durchgeführt. Dieser Arbeitsschritt repräsentiert den Übergang vom *low-level* Bereich in die nächst höhere Ebene, dem *intermediate-level*. Bei der Segmentierung wird eine erste Strukturierung der Bilddaten in ‚bedeutungsvolle‘ Teilbereiche angestrebt, und damit häufig der Übergang von der bildpunktbasierten (*ikonischen*) Beschreibung zu einer abstrakteren, symbolischen Form realisiert. So werden die ermittelten Helligkeitsunterschiede beispielsweise durch Geradenstücke und Ellipsenbögen beschrieben (Konturapproximation) oder Regionen mit ähnlichen Farbwerten ermittelt. Gerade diese beiden Beschreibungen nehmen eine *duale* Interpretation der Bilddaten vor, da Helligkeitsunterschiede häufig auf Objektkanten hindeuten, während die Regionen zumeist auf Objektoberflächen schließen lassen.

Motiviert durch die Gruppierungsmechanismen beim Menschen werden bei der Generierung einer **strukturellen Beschreibung** die Ergebnisse der Segmentierung in zunehmend globalere und abstraktere Strukturen organisiert. Geradenstücke werden beispielsweise aufgrund ihrer linearen Anordnung in lineare Konturgruppen organisiert, die wiederum in geschlossene Züge von Konturen gruppiert werden, und damit potentielle Beschreibungen von Objektoberflächen darstellen. Auch die regelmäßigen Anordnung von Regionen beinhaltet einen stark strukturierenden Charakter, und deren Detektion kann auf markante Objekteigenschaften hinweisen. Darüber hinaus kann auch die Integration dualer Bildbeschreibungen von entscheidender Bedeutung sein, um beispielsweise fehlerhafte Beschreibungen aufzulösen (siehe Regionenbildung in Abbildung 1.2) oder Informationen zu ergänzen.

Mit solchen strukturellen Beschreibungen kann dann für nachfolgende Prozesse im *high-level* Bereich eine Reduktion der Mehrdeutigkeiten und damit der Komplexität bei der Berechnung einer **Szeneninterpretation** erreicht werden. Während die Arbeitsschritte in den Ebenen *low-level* und *intermediate-level* zumeist problemunabhängig angelegt sind, also nur sehr allgemeines Wissen über die abgebildeten Szenen ausnutzen, sind Verarbeitungsprozesse im *high-level* sehr stark durch das Wissen über den betrachteten Problemkreis geprägt. So ist es beispielsweise für eine Objekterkennung von essentieller Bedeutung, bestimmte Objekteigenschaften und deren strukturellen Aufbau zu kennen, um aus den strukturellen Beschreibung oder Segmentierungsergebnissen eine

Hypothese für ein bestimmtes Objekt aufzustellen.

Mit dieser exemplarischen Darstellung für den Aufbau eines Bildanalyse-Systems wird die Komplexität der Aufgabe deutlich, eine automatische Beschreibung und Interpretation einer abgebildeten realen Szene zu generieren. Dabei ist die hier angedeutete, rein datengetriebene (*bottom-up*) Verarbeitung der Bilddaten nicht immer tragfähig. Eine Interaktion zwischen den einzelnen Arbeitsschritten steigert häufig die Leistungsfähigkeit des Gesamtsystems in hohem Maße. So kann beispielsweise eine Objekterkennung aus der Interpretation der Szene heraus die Aktivierung zusätzlicher Verarbeitungsmodulare aus dem Bereich *low-, intermediate-level* bewirken, um ergänzende Beschreibungen in problematischen Bildbereichen zu erlangen. Zudem ist unsere Umwelt nicht statisch. Die Integration einer zeitlichen Korrespondenz von Analyseergebnissen über eine Sequenz von Bildern unterstützt die Robustheit einzelner Module, und damit auch die des Gesamtsystems.

Mit der vorliegenden Arbeit wird neben der Strukturierung primitiver Bildelemente auch ein Schritt in die Richtung eines in sich rückgekoppelten Gesamtsystems, sowie eine Integration zeitlich korrespondierender Segmentierungsergebnisse vorgeschlagen.

1.2. Zielsetzung und Gliederung der Arbeit

Bei der Betrachtung des Bildes in Abbildung 1.1 wird bereits deutlich, daß ein menschlicher Beobachter trotz teilweise verdeckter Objekte offenbar eine Strukturierung des Bildes vornehmen kann, die ihm die Identifizierung einzelner Objekte ermöglicht. Motiviert durch die Erkenntnisse der Gestaltpsychologen werden in dieser Arbeit Methoden für eine Gruppierung von primitiven Konturbeschreibungen vorgestellt, die das Ziel einer automatisch generierten, konsistenten Organisation derselben verfolgen und damit eine ähnlich komplexe Objekterkennungsleistung unterstützen. Darüber hinaus wird eine Integration der konturbasierten Gruppierungen und der parallel erzeugten Regionen ähnlicher Farbe vorgeschlagen, um zum einen den Gruppierungsprozeß der Konturen zu unterstützen und zum anderen die Erkennungsleistung einer regionenbasierten Objekterkennung zu verbessern. Dabei ist von entscheidender Bedeutung, daß bei den entwickelten Methoden weitestgehend problemunabhängige Ansätze verfolgt werden, auch wenn der Entwicklungsrahmen stark von dem Problemkreis des Situiereten Künstlichen Kommunikators geprägt ist. Im folgenden wird ein Überblick über die Gliederung der vorliegenden Arbeit gegeben.

In Kapitel 2 wird zunächst auf die initiale Segmentierung der Bilddaten eingegangen. Dabei wird die vollständige Verarbeitung von statischen Bildern in eine initiale konturbasierte Beschreibung durch parametrische Geradenstücke und Ellipsenbögen detailliert vorgestellt. In dem zweiten Teil dieses Kapitels werden Methoden zur Generierung einer regionenorientierten Beschreibung diskutiert. Die dargestellten Verfahren sind zum Teil im Rahmen des Situiereten Künstlichen Kommunikators entwickelt worden und zum Teil aus der Literatur übernommen und angepaßt worden. Sie bilden die Basis für die

im weiteren beschriebenen Methoden zum perzeptiven Gruppieren und der Integration von Konturgruppierungen und Regionen. Die entsprechenden Ergebnisse werden bereits innerhalb dieses Kapitels diskutiert.

Die Mechanismen des perzeptiven Gruppierens werden, anhand der von den Gestaltpsychologen formulierten Gestaltgesetze, in Kapitel 3 anschaulich dargestellt. Darüber hinaus wird auf die Motivation und Ideen eingegangen, perzeptive Gruppierungsmechanismen in der Bildanalyse einzusetzen. Abschließend erfolgt eine Darstellung über verschiedene Gruppierungssysteme und deren Verfahrensweise.

Der zum Teil in der vorliegenden Arbeit entwickelte Ansatz eines hierarchisch orientierten, perzeptiven Gruppierens von initialen Konturbeschreibungen wird in Kapitel 4 vorgestellt. Dabei wird in diesem Kapitel auf die Generierung von Gruppierungshypothesen aufgrund lokaler Evidenzen eingegangen und deren potentielle datengetriebene Redundanz diskutiert. Im letzten Abschnitt dieses Kapitels wird ein Verfahren vorgestellt, das über die Gruppierungsleistung vieler bisher bekannter Systeme hinausgeht. Es realisiert das Gruppieren polygonal beschriebener Bildprimitiva mit ähnlicher Form aufgrund deren linearer Anordnung im Bild. Mit diesem Ansatz eröffnet sich somit die Möglichkeit, unter Verwendung der zu Beginn des Kapitels eingeführten Gruppierungshierarchie, zunehmend globalere Strukturen in Bildern zu detektieren.

In Kapitel 5 wird die Integration konturbasierter Gruppierungen und regionenbasierter Bildbeschreibungen vorgestellt. Nach einer Motivation und Diskussion verwandter Ansätze wird die Methodik des Verfahrens ausführlich entwickelt. Anschließend erfolgt eine Darstellung, wie eine Kombination aus Konturen und Regionen den Prozeß des konturbasierten Gruppierens sowohl in der Auswahl potentieller Gruppierungspartner unterstützt, als auch die zusätzliche Generierung von Konturgruppen aufgrund der Regioneninformation hervorruft. Im letzten Abschnitt dieses Kapitels wird ein Ansatz vorgeschlagen, der ein in sich rückgekoppeltes Gesamtsystem zur Objekterkennung durch die Integration von Konturbeschreibungen und Regionenbeschreibungen realisiert.

Die Generierung der konturbasierten Gruppierungshypothesen in Kapitel 4 beruht auf einer rein datengetriebenen Auswertung lokaler Evidenzen für eine Gruppierung. Zur Reduzierung von Mehrdeutigkeiten und Bewertung der Hypothesen in einem globaleren Kontext wird in Kapitel 6, nach einer allgemeinen Einführung in den Formalismus von Markov Random Fields, ein solches Bewertungskonzept vorgestellt. Hierbei wird detailliert auf die wesentlichen Aspekte zum Design eines entsprechenden Markov Random Fields eingegangen. Im zweiten Teil dieses Kapitels wird dann ein Markov Random Field vorgeschlagen, das für ein Konzept zur Kontursegmentierung in Bildfolgen verwendet wird und gleichzeitig die Etablierung einer zeitlichen Korrespondenz von Konturen realisiert.

In Kapitel 7 erfolgt schließlich die Präsentation und Diskussion der verschiedenen Parametrisierungen und der erzielten Ergebnisse auf realen Bildern für alle vorgestellten Methoden. Darüberhinaus erfolgt ein Ausblick auf weiterführende Entwicklungen. Den Abschluß bildet ein Kapitel mit einer Zusammenfassung der gesamten Arbeit.

2. Bildsegmentierung

Die Grundlage der meisten Bildanalyse-Systeme ist die Bildsegmentierung. Das Ziel einer Segmentierung ist die Überführung der rohen Bilddaten eines Aufnahmesystems in eine dem Kontext einer Bildanalyse angepaßten Beschreibungsform des Bildinhaltes. Diese Beschreibungsform kann beispielsweise durch eine Konvertierung der ikonischen Bilddaten aus einem Farbraum in einen anderen Farbraum erreicht werden oder aber einer ersten abstrakten, symbolischen Beschreibung des Bildinhaltes entsprechen. Welche Art von Segmentierung für eine bestimmte Analyse zu bevorzugen ist, kann nicht allgemein beantwortet werden. Diese Frage muß von Fall zu Fall unter Betrachtung des verfolgten Analyseziels diskutiert werden. Grundsätzlich sind Segmentierungstechniken zunächst jedoch als problemunabhängig anzusehen, und verschiedene Segmentierungen liefern nur verschiedene Beschreibungen *eines* Ereignisses innerhalb einer projizierten Szene.

Ein häufig auftretendes Problem bei der Segmentierung eines Bildes ist, daß die Beschreibung einiger Ereignisse gegebenenfalls entfällt bzw. von benachbarten Beschreibungen subsumiert wird. So kann beispielsweise eine farbbasierte Bereichsegmentierung häufig eine Trennung von gleich gefärbten, eng benachbarten Objekten nicht leisten und faßt alle Objekte zu einem Bereich gleicher Farbe zusammen. Solche oder ähnliche *Fehlsegmentierungen* treten in vielen Segmentierungstechniken auf und erfordern eine für das Problem adäquate Parametrisierung zur Reduzierung dieser fehlerhaften Interpretationen. Mit dieser Parametrisierung wird somit jede Segmentierung der Charakteristik des Analyseziels angepaßt. Damit eng verbunden ist die Frage nach der Evaluierung der Segmentierungsergebnisse in Abhängigkeit des Analyseziels. Diese gestaltet sich in der Regel schwierig, da für eine Evaluierung eine Segmentierung von Hand durchgeführt werden muß, die bei entsprechend aussagekräftigem Stichprobenumfang schnell unüberschaubaren Aufwand erzeugt. In vielen Fällen wird deshalb auf eine visuelle Inspektion der Ergebnisse zurückgegriffen, da die Ergebnisse genau im Kontext des Analyseziels betrachtet werden müssen. Diese visuelle Inspektion wird auch in der vorliegenden Arbeit eingesetzt.

In dieser Arbeit kommen zwei sich gegenseitig ergänzende Segmentierungstechniken zum Einsatz. Zum einen ist dies die *Kontursegmentierung*, mit der die in einem Bild vorhandenen Intensitätsübergänge zwischen homogenen Bildbereichen hervorgehoben werden. Diese sogenannten *Diskontinuitäten* beschreiben im Idealfall die Begrenzung von Objektoberflächen und sind damit sehr wichtig für eine genaue Lagebeschreibung und Rekonstruktion von Objekten in einer projizierten Szene. Sie können aber auch Tex-

turen von Objektoberflächen beschreiben oder die Schattenlinien des Objektschattens auf dem Hintergrund markieren. Darüber hinaus liefern Verfahren zur Kontursegmentierung häufig nur fragmentierte Beschreibungen der realen Konturen. Ursächlich für diese Fragmentierung sind gegenseitige Verdeckungen von Objekten oder aber schwach ausgeprägte Helligkeitsunterschiede zwischen dem Objekt und seinem Hintergrund.

Eine zur Kontursegmentierung duale Segmentierungstechnik ist die Bereichs- oder *Regionensegmentierung*. Allgemein formuliert, werden mit diesen Verfahren *homogene* Bereiche im Bild identifiziert, optimalerweise gerade die zu idealen Konturen komplementären Bildinhalte. Der Begriff der Homogenität kann hierbei aber auf unterschiedlichste Weise interpretiert werden. So kann ein *Homogenitätskriterium* direkt auf die Bilddaten von Grau- oder Farbwerten abgestimmt sein, um Bereiche gleicher Helligkeit oder Farbe zu identifizieren. Diese Bereiche beschreiben im Idealfall die von Konturen begrenzten Projektionen von Objektoberflächen und können so ebenfalls wichtige Informationen für eine Bildinterpretation bereitstellen. Das Homogenitätskriterium kann aber auch auf bereits extrahierte Informationen, wie Texturmerkmale oder Bewegungsinformation abgestimmt sein. In diesen Fällen werden Bereiche komplexerer Charakteristik als homogen identifiziert. Generell ist jedoch die Formulierung eines Homogenitätskriteriums problematisch, da aufgrund von Signalrauschen oder Fehlern in Vorverarbeitungsschritten die Eingangsdaten verfälscht sein können. Daher sind unterschiedlichste Verfahren entwickelt worden, die zum Teil in Abhängigkeit von der Bilddomäne homogene Bereiche detektieren.

In den folgenden Abschnitten werden die Segmentierungsverfahren beschrieben und deren Ergebnisse diskutiert, die für die Integration von Regionen und Konturen und das konturbasierte perzeptive Gruppieren eingesetzt werden. Sie bilden damit die Basis für alle in den folgenden Kapiteln entwickelten Gruppierungsmethoden. Im Zuge dieser Diskussion wird auch auf verschiedene weitere Segmentierungstechniken eingegangen.

2.1. Kontursegmentierung

Das Ziel einer Kontursegmentierung ist die Identifikation signifikanter Änderungen von Farb- oder Intensitätswerten in lokalen Bereichen eines Bildes. Diese Änderungen enthalten wichtige Hinweise auf die im Bild enthaltenen Strukturen und bilden so die Basis für eine weitere Interpretation des Bildinhaltes. Eine solche Interpretation kann z.B. eine Objekterkennung und -rekonstruktion sein, die auf eine symbolische Beschreibung der durch Farb- oder Intensitätsänderung definierten Objektgrenzen aufsetzt.

In den hier vorgestellten Kontursegmentierungsverfahren werden nur Grauwertbilder eingesetzt, womit sich die Detektion von Diskontinuitäten nur auf die Intensitätsanteile einer projizierten Szene beschränkt. In der Literatur finden sich auch Verfahren zur Kontursegmentierung, die eine Detektion von Konturen in *Multiband*-Bildern (z.B. Farbbildern) vorschlagen [Gev00]. Diese Verfahren nutzen wesentlich mehr Information zur Konturdetektion aus, haben aber den Nachteil, daß sie deutlich umfangreichere Be-

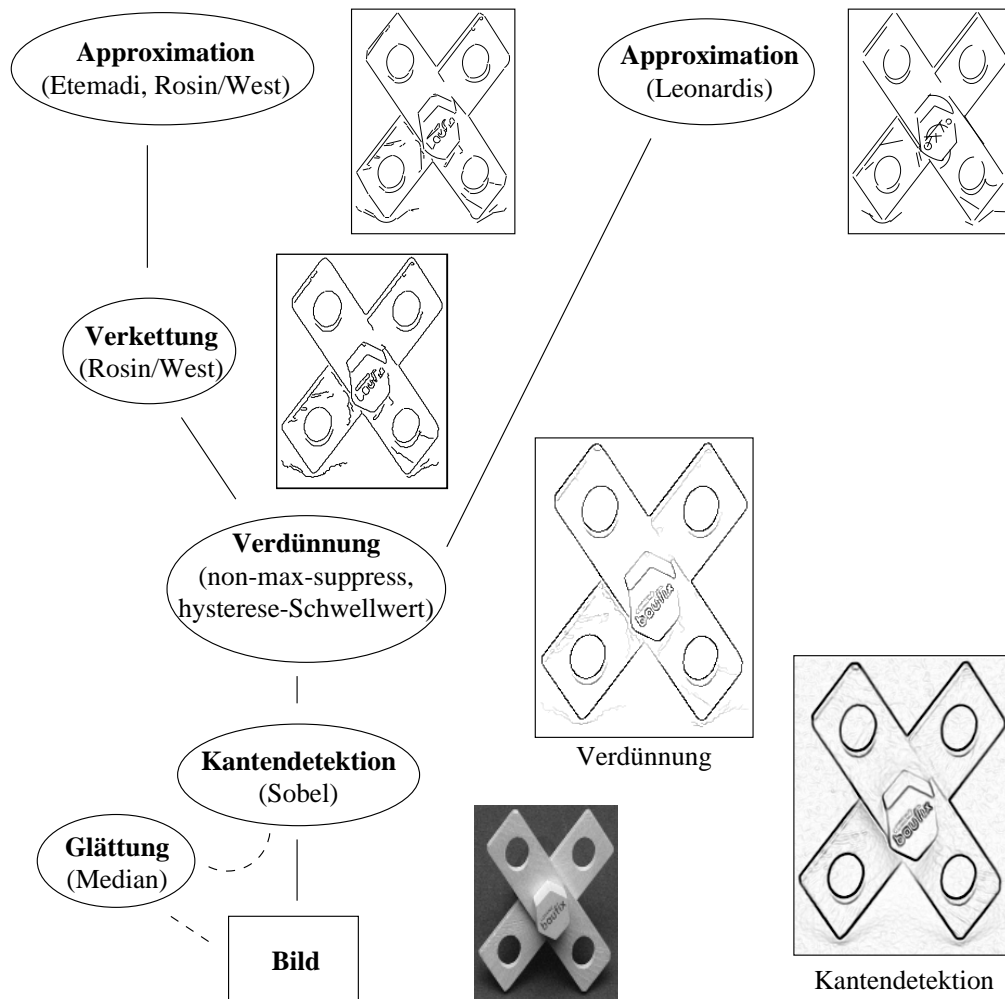


Abbildung 2.1.: Übersicht zum Prozeß der Kontursegmentierung (Angabe der hier diskutierten Verfahren). Die Glättung eines Bildes ist als optionaler Bestandteil anzusehen. Bei der Kantendetektion und der Verdünnung entsprechen dunkle Stellen im Beispielbild hohen Gradientenbetragswerten der Kantenelemente. Die Verkettung ist durch Verbindungslinien zwischen den verketteten Kantenelementen markiert. In der Approximation sind die parametrisierten Geradenstücke und Kreis- bzw. Ellipsenbögen eingezeichnet.

rechnungen und damit längere Bearbeitungszeiten benötigen. Darüber hinaus führt die Verarbeitung von Intensitätsbildern in dem gewählten Szenario zu einer ausreichenden Beschreibung von Konturen.

Der gesamte Prozeß der Kontursegmentierung gliedert sich in mehrere Teilschritte (siehe Abbildung 2.1). Aufgrund von Signalrauschen oder anderen physikalischen Ein-

flüssen können in den Bilddaten stärkere Variationen auch in homogenen Bereichen des Bildes auftreten. Zur Reduzierung von Fehlsegmentierungen können diese Variationen in einem ersten Schritt durch eine *Glättung* des Intensitätsverlaufs verringert werden. Danach erfolgt die Beschreibung von Konturen in einer ersten Phase wiederum in einer ikonischen Form. Der als *Kantendetektion* bekannte Prozeß markiert Bildpunkte als *Kantenelemente*, die in einer lokalen Nachbarschaft zu Diskontinuitäten im Intensitätsverlauf des Bildes führen. Als Ergebnis erhält man eine *Kante*, die in ihrer Breite durchaus mehrere Bildpunkte umfassen kann. Das liegt zum Teil in einem Signalrauschen entlang von Konturen begründet, wird aber auch durch den gewählten Kantendetektor oder eine nicht ideale Stufenfunktion zwischen den verschiedenen Bildbereichen hervorgerufen. Aus diesem Grund schließt sich eine *Verdünnung* der Kanten auf eine Breite von einem Bildpunkt orthogonal zum Kantenverlauf an. Diese ersten Teilschritte werden häufig allgemein unter dem Begriff der *Detektion* zusammengefaßt.

Die Kantenelemente bilden dann die Ausgangsbasis für weitere Schritte zur Kontursegmentierung. Eine erste einfache Form einer symbolischen Beschreibung von Konturen ist das *Verketteten* von Kantenelementen zu *Kontursegmenten*. Darauf aufbauend bietet sich eine analytische Beschreibung der Kontursegmente durch parametrisierte geometrische Figuren an. Dieser Prozeß der *Konturapproximation* beschreibt die Kontursegmente beispielsweise durch Geradenstücke, Kreis- oder Ellipsenbögen. Das in dieser Arbeit eingesetzte Approximationsverfahren von Leonardis [Leo93] vernachlässigt den Schritt des Verkettens und arbeitet direkt auf den detektierten Kantenelementen.

Im folgenden wird zunächst der Schritt der Detektion diskutiert und anschließend werden Methoden zur Approximation vorgestellt.

2.1.1. Detektion

Zentrales Ziel einer Kantendetektion ist die Ermittlung von realen Kantenpunkten im Bild. Das bedeutet, daß Bildpunkte, die in der realen Welt zu Konturen von Objekten oder Materialien gehören, auch tatsächlich im Bild als Kantenelemente identifiziert werden sollen. Eine weitere Anforderung ist die korrekte Lokalisation der Kantenpunkte im Bild. Kantenelemente sollen möglichst genau die Position der realen Kontur im Bild lokalisieren. Die Evaluierung von Kantendetektoren ist aufgrund der großen Datenmengen sehr aufwendig und erfolgt deshalb in dieser Arbeit durch eine visuelle Inspektion. Eine systematische Evaluierung durch visuelle Inspektion mit mehreren Versuchspersonen wird in [Hea97] vorgeschlagen. Die vorgeschlagene Methode verdeutlicht dabei auch die Schwierigkeiten und den Aufwand eines solchen Ansatzes.

Glättung des Bildes

Die Qualität einer Kantendetektion hängt entscheidend von der Qualität des zu verarbeitenden Bildes ab. Ein Bild kann für einen Betrachter als eine fehlerfreie Projektion der Szene erscheinen, weil die visuelle Wahrnehmung viele komplementäre Prozesse

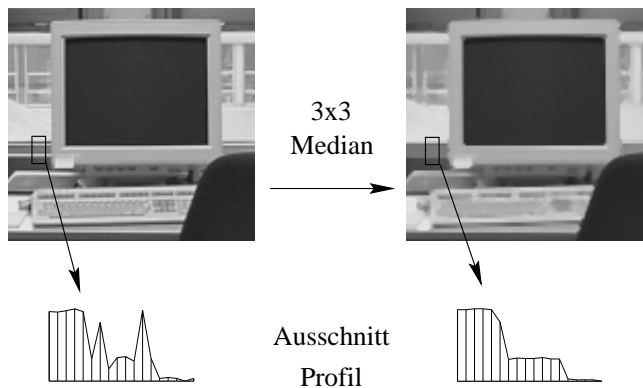


Abbildung 2.2.: Wirkung des Medianfilter bei Anwendung einer 3×3 -Maske auf eine Bildmatrix. Die Profildarstellung des Ausschnittes verdeutlicht die Eliminierung von abrupten Sprüngen, ohne den Sprung im Intensitätsverlauf zu verschmieren.

in die Erkennungsleistung integriert. Aus diesem Grund nimmt ein Mensch ein lokales Rauschen innerhalb einer homogen erscheinenden Region nicht wahr. Für eine Kantendetektion deuten diese Diskontinuitäten jedoch auf reale Konturen hin. Hervorgerufen wird ein Rauschen häufig durch physikalische Unzulänglichkeiten eines Aufnahmesystems. Beispielsweise kann die Qualität der elektronischen Bauteile in der Kamera oder dem Framegrabber (Analog-Digital-Wandler) minderwertig sein, oder das Signal wird auf dem Weg von der Kamera zum Rechner von externen Quellen gestört.

Das Rauschen kann durch den Einsatz von *digitalen Filtern* reduziert werden. Die Wahl des Filters ist dabei stark von den nachfolgenden Verarbeitungsschritten abhängig: Im Vorfeld einer Kantendetektion soll eine Glättung zu einer Eliminierung von Diskontinuitäten führen, die aufgrund von lokalem Signalrauschen entstanden sind. Gleichzeitig sollen jedoch signifikante Diskontinuitäten an projizierten realen Konturen erhalten bleiben. Ein hierfür geeigneter Filter ist der *Medianfilter*, der zu der Gruppe der *nichtlinearen Rangordnungsfilter* gehört. Der Medianfilter unterdrückt lokales Rauschen, ohne jedoch Diskontinuitäten an echten Kanten zu verschmieren (siehe Abbildung 2.2). Allgemein ist der Median einer Menge F mit N Werten definiert als: $med(F) = f_{\lfloor N/2 \rfloor}$, wobei $F = \{f_1 < f_2 < \dots < f_N\}$. Auf einer Bildmatrix wird der Medianfilter in der Regel durch eine quadratische Maske mit ungerader Anzahl Spalten und Zeilen realisiert, die über das Bild verschoben wird. Als Filterantwort wird im Ausgabebild das Zentralpixel des aktuell überdeckten Bildbereiches auf den Median der betrachteten Pixelwerte gesetzt. Mit diesem Vorgehen können einzelne Pixelwerte, die Spitzen oder Täler bezüglich einer lokalen Nachbarschaft im Bild beschreiben, unterdrückt werden.

Ein entscheidendes Problem in der Anwendung des Medianfilters ist das laufende Sortieren der von einer Maske überdeckten Intensitätswerte. In [Hua79] wurde deshalb eine effiziente Berechnung des Medians auf Bildmatrizen mit Histogrammen vorgeschla-

gen: Bei der Verschiebung der Operatormaske mit $M \times M$ Spalten/Zeilen über das Bild fallen immer nur M Werte aus einer zu sortierenden Liste, und es kommen nur M Werte hinzu. Wird diese Liste in einem kumulativen Histogramm abgebildet, so entspricht der Median dem Wert im Histogramm, bei dem die Häufigkeit von $\lfloor M^2/2 \rfloor$ überschritten wird. Nach einer Verschiebung der Maske werden im Histogramm die Akkumulatoren der herausfallenden Intensitätswerte dekrementiert und die Akkumulatoren der neu hinzukommenden Intensitätswerte inkrementiert. Der Median läßt sich dann wiederum über die Häufigkeit bestimmen. Das Histogramm muß jeweils nur am Anfang einer neuen Zeile neu initialisiert werden. Besonders bei größeren Masken ist diese Methode deutlich effizienter.

Da in dieser Arbeit überwiegend Bildaufnahmesysteme aus dem Konsumerbereich verwendet wurden, sind alle Bilder vor einer Kantendetektion mit einem 3×3 Median vorverarbeitet worden.

Detektion von Diskontinuitäten

Der erste Schritt von den Bilddaten zu einer Beschreibung der abgebildeten Konturen ist die Detektion von Diskontinuitäten. In der Literatur finden sich verschiedene Ansätze, die hierfür überwiegend eine Faltung des Bildes mit einer Operatormaske verwenden.

In der Praxis haben sich Operatoren bewährt, die auf einer Differenzbildung der Grauwerte innerhalb einer Maske beruhen. Diese *Gradientenoperatoren* realisieren mit Hilfe von Faltungen eine Approximation der ersten oder zweiten partiellen Ableitung des diskreten Bildsignals. Typische Vertreter sind der Roberts-, Sobel-, oder Laplaceoperator. Eine einfache erste Ableitung des Signals, die jedoch sehr störungsanfällig ist, wird vom 2×2 Robertsoperator durchgeführt. Zur Reduzierung einer Störanfälligkeit wird beim Sobeloperator eine 3×3 -Maske eingesetzt, die zusätzlich eine Glättung orthogonal zur Differenzierungsrichtung durchführt (Abbildung 2.3(a)). Aus den einzelnen

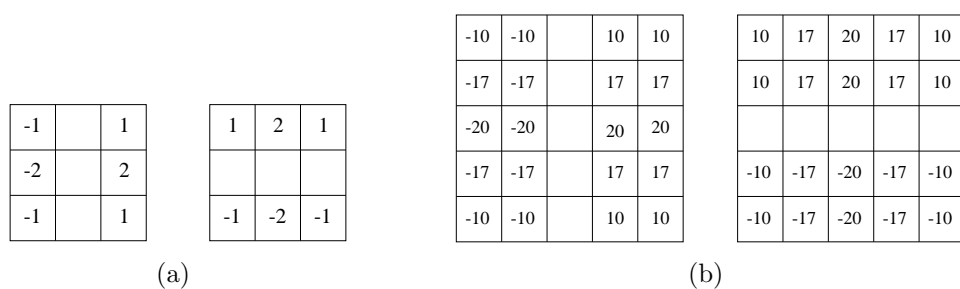


Abbildung 2.3.: (a) Operatormasken des Sobel, (b) erweiterte Masken nach [Bäs93].

Richtungsgradienten wird der Betrag, als Maß für die Stärke, sowie die Richtung der detektierten Diskontinuität bestimmt. Eine Kante im Bild wird von diesen Operatoren durch ein lokales Maxima des Gradientenbetrages angezeigt. Zur Realisierung der Ableitung zweiter Ordnung wird häufig der Laplaceoperator eingesetzt, der eine Kante

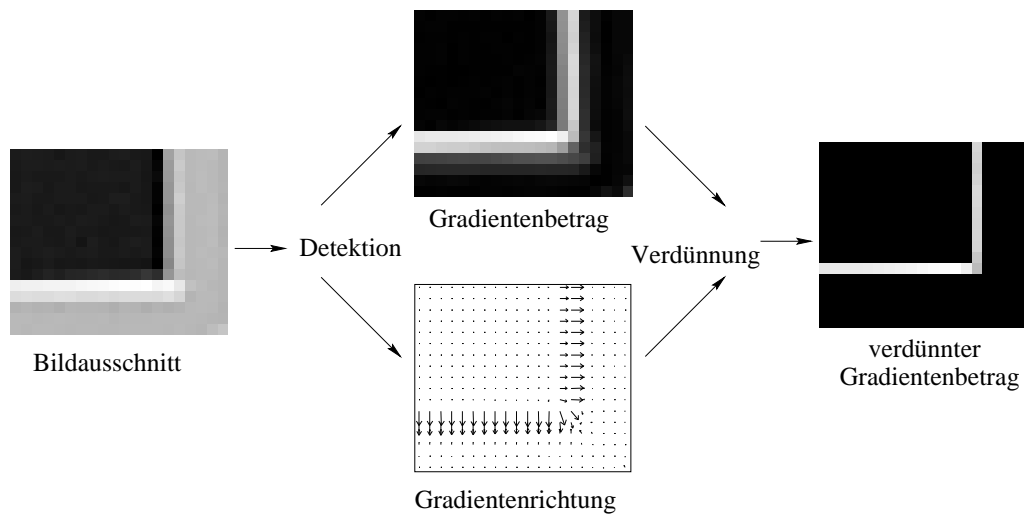


Abbildung 2.4.: Kantendetektion: Zunächst werden mit einem Sobeloperator Gradientenbetrag und -richtung ermittelt. Mit dieser Information wird eine Verdünnung der Kontur auf die stärksten Gradientenbeträge in einer lokalen Nachbarschaft durchgeführt.

über einen Nulldurchgang auf der zweiten Ableitung der Bildfunktion identifiziert. Jedoch ist auch seine Filterantwort ohne vorheriges Glätten sehr störanfällig. Neben diesen Differenzoperatoren werden noch Operatoren eingesetzt, die mit Modellkanten arbeiten (z.B. [Hue73]) oder einen Bandpaßfilter im Frequenzraum realisieren (z.B. [Mar80]). Ein guter Überblick über Randbedingungen und verschiedene Operatoren für die Kantendetektion findet sich in [Ste93].

In dieser Arbeit kommt der einfache Sobeloperator bzw. eine in Anlehnung an den Sobeloperator entworfene und in [Bäs93] vorgeschlagene 5×5 -Maske zum Einsatz (Abbildung 2.3(b)).

Verdünnung der Kante

Die Anwendung der Operatoren zur Kantendetektion liefern in ihren Filterantworten keine idealen Kanten, da die Glättung mit Median und Sobeloperator nur vereinzelte starke Störimpulse eliminiert. Damit führt ein unterschwelliges Grundrauschen im gesamten Bild zu geringen Diskontinuitäten, die nicht als Kantenelemente identifiziert werden sollen. Darüber hinaus liefern die Kantendetektoren, bedingt durch die Größe ihrer Operatormaske, immer einen breiteren Bereich signifikanter Diskontinuitäten, in dem die korrekten Kantenelemente identifiziert werden müssen.

Anschaulich läßt sich die Verteilung der Gradientenbeträge in der Nähe von Konturen wie die Höhen von langgezogenen Gebirgszügen beschreiben. Ziel der Verdünnung ist dann die Bestimmung der Kette von Gipfelpunkten. Hierzu wurde von [Can83] das

Verfahren des „*non-maximum-supression*“ vorgeschlagen. Dieses Verfahren führt die Verjüngung eines Bereiches von Diskontinuitäten auf Gipfelpunkte durch, indem es die Gradientenrichtung des Kantenelementes ausnutzt. Es werden alle Kantenelemente eliminiert, die in einer Nachbarschaft parallel zur Gradientenrichtung nicht den maximalen Gradientenbetrag aufweisen. Damit bleiben in einer lokalen Umgebung nur die stärksten Diskontinuitäten erhalten.

Die verbleibenden lokal maximalen Diskontinuitäten werden mit einem *Hysterese-Schwellwert*-Verfahren weiter bearbeitet [Can83]. Dieses Verfahren eliminiert in einem Durchlauf alle Kantenelemente aus dem Bild, die einen Schwellwert ν_{low} nicht überschreiten, und akzeptiert gleichzeitig alle Kantenelemente, deren Gradientenbetrag über einem Schwellwert $\nu_{high} > \nu_{low}$ liegen. In einem weiteren Verarbeitungsschritt werden ausgehend von den Kantenelementen mit einem Betrag größer ν_{high} nur Diskontinuitäten als Kantenelement akzeptiert, die transitiv über eine direkte Nachbarschaft erreichbar sind.

Abbildung 2.4 veranschaulicht den gesamten Ablauf der Detektion wie sie auch in dieser Arbeit eingesetzt wird.

2.1.2. Approximation

Mit der Detektion von Kantenelementen sind potentielle Konturen identifiziert, die noch in einer ikonischen und damit unhandlichen Beschreibungsform vorliegen. Daher wird für weitere Analyseprozesse ein stärkerer Abstraktionsgrad und eine Reduzierung der Datenmengen angestrebt. Dies wird in der Regel durch eine parametrische Approximation der Konturen erreicht. Im folgenden werden zwei prinzipielle Vorgehensweisen für die Approximation vorgestellt. Zum einen erfolgt eine Verkettung von Kantenelementen zu *Kontursegmenten*, die im weiteren stückweise durch geometrische *Konturprimitiva* approximiert werden. Das in dieser Arbeit eingesetzte Verfahren von Leonardis [Leo93] schlägt hingegen eine Vereinigung dieser Schritte für den Approximationsprozeß vor.

Verkettung

Die Grundidee aller Verfahren zur Verkettung von Kantenpunkten zu Kontursegmenten ist eine Zusammenhangsanalyse. Hierbei werden unter Ausnutzung der Betrags- und zum Teil auch der Richtungsinformation direkt benachbarte Kantenelemente zu einem Kontursegment verkettet. Das Hauptproblem ist dabei die Entscheidung, an welcher Stelle ein Kontursegment endet und ein oder mehrere neue Segmente beginnen. Dieses Problem tritt insbesondere an Kreuzungen und bei starken Richtungsänderung von Konturen auf.

In dem Verfahren von [Ros89] wird der Gradientenbetrag für die Verkettung ausgenutzt. Ausgehend von Kantenelementen mit nur einem Nachbarn wird sukzessive eine Kette um Kantenelemente erweitert, bis an einem Element keine weiteren Nachbarn vorhanden sind. In einem zweiten Schritt werden zyklische Ketten ausgehend von einem Kantenelement mit mindestens zwei Nachbarn gesucht. Die Problematik von Verzwei-

gungen wird dabei nicht behandelt. Die Auswahl des nächsten Nachbarelementes ist nur bestimmt durch die Reihenfolge der untersuchten Nachbarn. Dieser Ansatz wurde für diese Arbeit untersucht und um die Berücksichtigung von Richtungsinformation erweitert. Hierzu wird die Richtung des zuletzt hinzugefügten Kantenelementes an Kreuzungspunkten ausgewertet und mit einem Schwellwert über die Abweichung entschieden, mit welchem der benachbarten Kantenelemente das Segment erweitert werden soll. Dieser Ansatz führt zu weitaus besseren Ergebnissen. Generell ist die Repräsentation von Konturen durch Kontursegmente jedoch nur eingeschränkt für eine weitere Bildanalyse geeignet, und sie stellt in der Regel nur einen Zwischenschritt für die Approximation der Konturen dar.

Approximation

Die Approximation von Kontursegmenten verfolgt das Ziel, eine wesentliche Abstraktion und Reduktion der Bilddaten zu erreichen. Ansätze, die eine Approximation in Geradenstücke und Kreisbögen als Konturprimitiva verfolgen, finden sich z.B. in [Ros89, Ete92, Mok97]. Andere parametrisierte Beschreibungen werden beispielsweise in [Ros93] mit *Codons* und in [Pie97] mit *Splines* vorgeschlagen. Letztere beschreiben Konturen zwar in einer sehr eleganten Form, bieten jedoch nur wenige Ansatzpunkte für das im weiteren verfolgte konturbasierte perzeptive Gruppieren und werden deshalb nicht weiter behandelt.

Für diese Arbeit wurden die Verfahren nach Rosin und West [Ros89] und Etemadi [Ete92] untersucht. In diesen Ansätzen beruht die Approximation in Geradenstücke und Kreisbögen, wie in vielen anderen Ansätzen auch, auf einer Minimierung des quadratischen Abstandes der Kantenelemente zur parametrisierten Beschreibung der Kontur. Initial wird bei beiden Verfahren durch sukzessive Erweiterung entlang des Kontursegmentes eine polygonale Approximation des Segmentes erzeugt. Dabei kommen in dem jeweiligen Verfahren unterschiedliche Techniken und Randbedingungen zum Einsatz. Anschließend werden die Polygonkanten mit parametrischer Suche zu Geradenstücke oder Kreisbögen zusammengefaßt, die dann das Kontursegment stückweise approximieren. Mit dieser Vorgehensweise sind beide Verfahren explizit von den Ergebnissen der Konturpunktverkettung abhängig. Insbesondere können fehlende Segmente nicht approximiert und fälschlich generierte Segmente nicht verworfen werden. Darüber hinaus kann eine höhere Fragmentierung der Konturbeschreibung durch ungünstige Aufspaltungen der Segmente entstehen. In Abbildung 2.5 ist diese Abhängigkeit von einer Konturverkettung herausgestellt: Bei der Approximation nach Etemadi kann keine Korrektur fehlerhafter Verkettungen erfolgen (siehe Abbildung 2.5 oben rechts), wohingegen bei einer Approximation ohne explizite Verkettung kleinere Lücken oder Störungen überwunden werden können (Abbildung 2.5 unten rechts).

Eine Fragmentierung kann generell vom perzeptiven Gruppieren aufgelöst werden, jedoch ist zur Konzentration auf die wesentlichen Aspekte des Gruppierungsprozesses in dieser Arbeit eine initial gute Approximation der Konturen erwünscht. Aus diesem

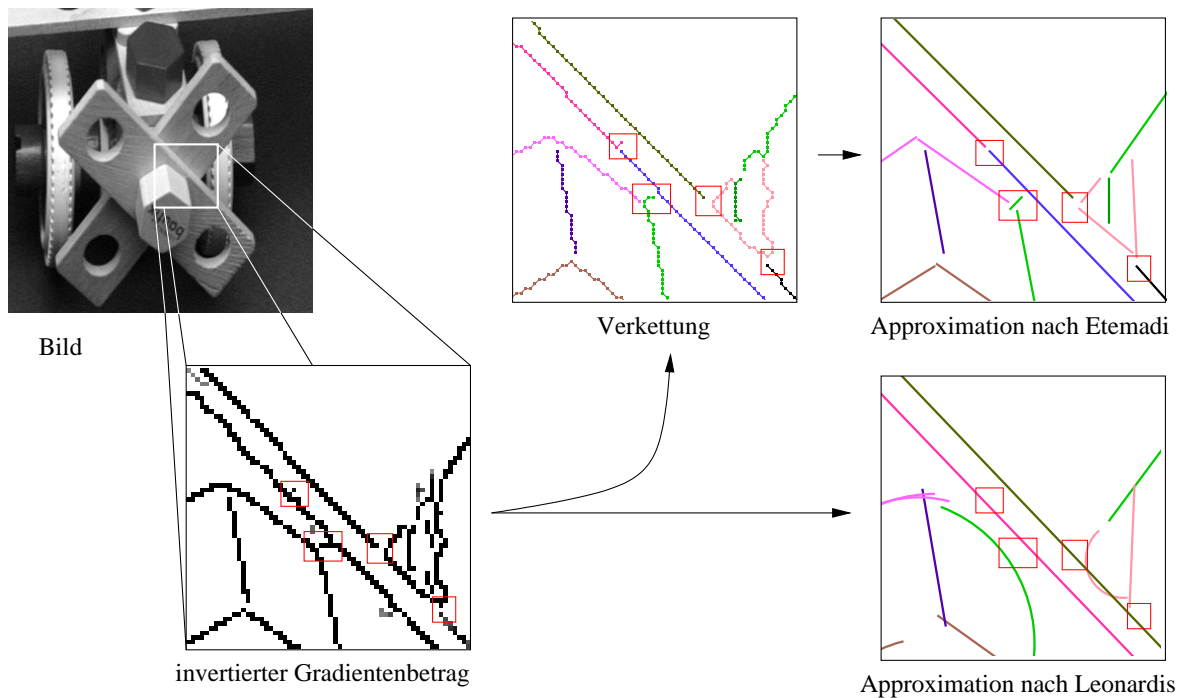


Abbildung 2.5.: Probleme bei der Konturapproximation: In dem Ausschnitt des Gradientenbetragsbildes sind problematische Bereiche rot umrandet. Die durch Verkettung ermittelten Kontursegmente sind jeweils farbig markiert (oben Mitte), wobei die fehlerhafte Interpretation deutlich zu erkennen ist. Die Konturprimitiva, die ein Kontursegment stückweise beschreiben, sind entsprechend eingefärbt (Rechts oben: Approximation auf Basis fehlerhaft verketteter Konturen. Rechts unten: Approximation ohne vorgelagerte Verkettung.)

Grund sind die Verfahren nach [Ros89] und [Ete92] nicht weiter eingesetzt worden und es wird der Approximationsalgorithmus nach Leonardis [Leo93] verwendet. Dieser wird im folgenden Abschnitt im Detail dargestellt.

2.1.3. Approximation nach Leonardis

Ein Verfahren zur parametrischen Beschreibung von Konturen, das eine angemessen gute initiale Approximation für die hier verwendete Bilddomäne liefert, wird von Leonardis [Leo93] vorgestellt.

In diesem Ansatz wird auf eine explizit vorangestellte Verkettung der Kantenelemente verzichtet und die Approximation direkt auf den Kantenelementen des verdünnten Gradientenbetragsbildes durchgeführt. In einer ersten Phase, der *Model Recovery*, werden

parametrisierte *Konturmodelle*¹ als Beschreibung der möglichen realen Konturen hypothetisiert. Dabei können mehrere Konturmodelle durchaus die gleichen Kantenelemente beschreiben. Dieser Schritt vereinigt somit die Verkettung und Approximation, läßt aber eine Entscheidung für die Begrenzung einzelner Segmente im Sinne einer Verkettung indirekt offen: Jedes hypothetisierte Konturmodell beschreibt *ein* Kontursegment, wobei erst in einer zweiten Phase, der *Model Selection*, entschieden wird, welche Konturmodelle die Konturen optimal beschreiben.

In [Leo93] werden Geradenstücke, Parabeln und Ellipsenbögen als Konturmodelle eingesetzt. Das Fehlermaß, oder auch der *Goodness-of-Fit*, wird über den mittleren quadratischen Abstand der Kantenelemente zur Approximation beschrieben. Für die Parameterschätzung setzt Leonardis eine Momentenmethode (Geradenstücke) und das Levenberg-Marquardt Verfahren (siehe z.B. [Pre88]) zur Lösung nichtlinearer Gleichungssysteme (Parabel, Ellipse) ein. In der für diese Arbeit verwendeten Implementation aus dem VISTA Bildverarbeitungspaket [Pop94] kommt zudem eine Approximationsmethode von Taubin [Tau91] mit *conics* zum Einsatz. Diese mit einer Eigenvektormethode ermittelten Beschreibungen lassen sich in Geradenstücke oder Ellipsenbögen klassifizieren. In der vorliegenden Arbeit werden nur Geradenstücke, Kreis- und Ellipsenbögen als Konturmodelle verwendet, da sie mit ihrer Parametrisierung eine gute Ausgangsbasis für den Einsatz im konturbasierten Gruppieren darstellen.

Model Recovery

Ausgehend von kleinen regelmäßigen Blöcken im Bild startet in der *Model Recovery* ein unabhängiger, iterativer Prozeß aus Approximation und Expansion für diese Blöcke (siehe Abbildung 2.6(a)).

Zunächst werden alle Kantenelemente innerhalb eines dieser sogenannten *Saatfenster* für die Approximation mit dem einfachsten parametrischen Modell (Geradenstücke) herangezogen. Die Approximation wird gespeichert, falls der mittlere Fehler akzeptabel ist²:

$$\overline{d_{\psi}^2}(P) := \frac{1}{|P|} \sum_{p \in P} d_{\psi}^2(p) < \Delta_{\text{accept}}(\cdot) \quad (2.1)$$

Hierbei bezeichnet P die Menge der approximierten Kantenelemente und $d_{\psi}^2(p)$ beschreibt den senkrechten quadratischen Abstand des Kantenelementes p von dem Konturmodell mit dem Parametersatz ψ . In dem Originalverfahren von Leonardis wird bei einem nicht akzeptablen initialen Approximationsfehler das gesamte Saatfenster verworfen. In der VISTA Implementation wurde hierzu eine Modifikation eingeführt: Ist der Approximationsfehler nicht akzeptabel wird das Saatfenster in vier gleichmäßige Blöcke unterteilt und mit einer neuen initialen Approximation einer Geradenstück für jedes dieser Fenster begonnen. Dieser Schritt der Unterteilung wiederholt sich im Falle einer

¹Der Begriff *Konturmodell* wird im folgenden synonym zum Begriff des *Konturprimitivum* oder der *Konturapproximation* benutzt und mit dem Symbol \cdot referenziert.

²Die hier verwendete Nomenklatur lehnt sich an Posch [Pos97] an.

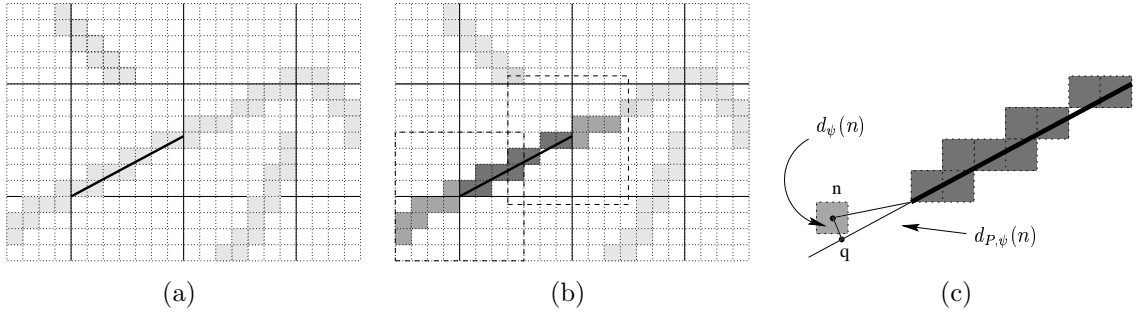


Abbildung 2.6.: Illustration zur *Model Recovery*: In (a) sind die Saatfenster mit durchgezogenen Linien dargestellt. Die Kantenelemente sind leicht schattiert markiert und die dicke Linie markiert eine initiale Approximation für ein Saatfenster. (b) Potentielle Expansionskandidaten n für eine Approximation sind in den Suchfenstern (gestrichelt gezeichnet) durch dunklere Schattierungen markiert. Bereits verwendete Kantenelemente P werden dunkelgrau dargestellt. (c) Visualisierung der Abstände, die die Auswahl von n einschränken.

fehlgeschlagenen Approximation bis eine Mindestgröße für die Saatfenster unterschritten wird. So können beispielsweise Saatfenster, die Kreuzungen von Konturen enthalten, trotzdem zur Initialisierung verwendet werden und darüber hinaus wird die Größe des Saatfensters vom Rest des Verfahrens unabhängig.

Ist der Fehler für eine initiale Approximation akzeptabel, wird das Konturmodell an den Endpunkten sukzessive um kompatible Kantenelemente expandiert. Hierzu wird an jedem Ende des Modells ein Fenster plaziert (Abbildung 2.6(b)) und die enthaltenen Kantenelemente müssen als Expansionskandidat n folgenden Einschränkungen genügen (Abbildung 2.6(c)):

$$d_{\psi}^2(n) \leq \Delta_{compatibility}(\cdot) \quad (2.2)$$

$$d_{P,\psi}(n) := \min_{p \in P} d_{\psi}(q, p) \leq \Delta_{continuity}(\cdot) \quad (2.3)$$

Gleichung (2.2) begrenzt den maximal erlaubten quadratischen Abstand von n zur aktuellen Approximation. Gleichung (2.3) schränkt die maximal erlaubte Lücke zwischen der Projektion q eines kompatiblen Punktes n und eines bereits für die Approximation verwendeten Punktes p ein. Dabei mißt $d_{\psi}(q, p)$ den Abstand der Punkte p, q entlang des Verlaufs der aktuellen Approximation. Alle Punkte, die obigen Bedingungen genügen, werden anschließend für eine temporäre neue Schätzung der Parameter für diesen Modelltyp verwendet. Ist für diese temporäre Approximation die Gleichung (2.1) erfüllt, werden die neuen Parameter ψ_{new} und die Kantenelemente P_{new} übernommen und der Prozeß der Expansion iteriert, bis eine akzeptable Erweiterung für diesen Modelltyp

Model Recovery

```

M = ∅ {Menge der Konturmodelle}
for all Saatfenster do
  M = M ∪ Initial Models(Saatfenster)
end for

for all m ∈ M do
  repeat
    Pnew = P ∪ c {Erweitere P um Kantenelemente c}
    if c = ∅ ∨
      Gleichung (2.2) ∨ Gleichung (2.3) für Pnew nicht erfüllt then
        next m {Expansion beenden}
      else
        berechne neue Parameter ψnew mit Pnew {Parameter schätzen}
        if Gleichung (2.1) für mψnew mit Pnew erfüllt then
          m = mψnew ; P = Pnew {Expansion fortführen}
        else
          approximiere Pnew mit nächst komplexeren {Komplexeres Modell wählen}
            Modelltyp mcomplex
          if kein komplexeres Modell mcomplex ∨
            Gleichung (2.1) für mcomplex mit Pnew nicht erfüllt then
              next m {Expansion beenden}
            else
              m = mcomplex ; P = Pnew {Expansion fortführen}
            end if
          end if
        end if
      until Expansion m beendet
    end for

```

Initial Models(*Saatfenster*)

```

m = Geradenapproximation mit allen Kantenelementen P aus Saatfenster
if Gleichung (2.1) für m mit P erfüllt then
  return {m}
else
  Mtmp = ∅
  subsaat[4] = 4 gleichmässige Blöcke aus Saatfenster
  for i = 1..4 do
    Mtmp = Mtmp ∪ Initial Models(Subsaat[i])
  end for
  return Mtmp
end if

```

Algorithmus 2.1: Model Recovery in Anlehnung an [Pos97].

nicht mehr möglich ist. Ist Gleichung (2.1) für eine Punktmenge nicht mehr erfüllt, wird das nächst komplexere Modell, ein Kreis oder eine Ellipse, zur Approximation und weiteren Expansion verwendet. Lassen sich auch für diese Modelltypen keine Parameter schätzen, endet die *Model Recovery* für dieses Saatfenster und das letzte akzeptable Konturmodell wird gespeichert. Anschließend wird für das jeweils nächste Saatfenster der iterative Prozeß aus Approximation und Expansion gestartet, bis eine Menge M von Konturmodellen für das gesamte Bild generiert wurde. Der gesamte Prozeß ist in Algorithmus 2.1 dargestellt.

Model Selection

Nachdem die *Model Recovery* für alle Saatfenster durchgeführt wurde, muß eine Selektion geeigneter Konturmodelle erfolgen. Da potentiell für jedes Saatfenster eine Konturbeschreibung generiert und expandiert worden ist, existieren in der Regel mehrere Konturmodelle, die die gleichen Kantenelemente beschreiben. Zur Auflösung dieser Redundanz sind verschiedene Aspekte zu betrachten. Die Konturmodelle zur Approximation einer realen Kontur sollen möglichst vollständig expandiert, möglichst genau lokalisiert und möglichst einfach, d.h. mit wenigen Parametern beschrieben, sein.

Zur Erfüllung dieser Kriterien wird von Leonardis eine *Objective Function* zur Selektion von Konturmodellen vorgeschlagen:

$$F(\mathbf{m}) := \kappa_1(\cdot)n(\mathbf{m}) - \kappa_2(\cdot)\xi(\mathbf{m}) - \kappa_3(\cdot)N(\mathbf{m}) \quad (2.4)$$

mit \mathbf{m} Teilmenge der Konturmodelle aus der Model Recovery
 $n(\mathbf{m})$ Anzahl der durch \mathbf{m} approximierten Kantenelemente
 $\xi(\mathbf{m})$ Approximationsfehler aller Konturmodelle in \mathbf{m}
 $N(\mathbf{m})$ Anzahl der Modellparameter der Konturmodelle in \mathbf{m}
 $\kappa_i(\cdot)$ Gewichte der Terme

Ziel ist es, diese Funktion mit einer Teilmenge \mathbf{m} der hypothetisierten Konturmodelle M zu maximieren. Zur Berücksichtigung der Interaktion (Überlappung) zwischen den einzelnen Approximationen wird die Funktion F erweitert:

$$F_{interact}(\mathbf{m}) := F(\mathbf{m}) + \sum_{\substack{m_i, m_j \in \mathbf{m}, \\ m_j \neq m_i}} \frac{\kappa_2(\cdot)\xi_{i,j} - \kappa_1(\cdot)\Gamma_{i,j}}{2} \quad (2.5)$$

mit $\Gamma_{i,j} := |P_i \cap P_j|$

Anzahl der durch m_i und m_j gemeinsam approximierten Kantenelemente und

$\xi_{i,j} := \max \left(\sum_{p \in P_i \cap P_j} d_{\psi_i}(p), \sum_{p \in P_i \cap P_j} d_{\psi_j}(p) \right)$ maximale Fehlersumme im Überlappungsbereich der Konturmodelle m_i, m_j .

Mit dieser Definition läßt sich durch ein Optimierungsverfahren die Teilmenge von Konturmodellen extrahieren, die die realen Konturen im Bild unter den obigen Bedingungen optimal beschreiben.

Ein häufiges Problem bei der Optimierung von Funktionen ist der sehr hohe Rechenaufwand für eine vollständige Lösung des Problems. Aus diesem Grund schlägt Leonardi einen *greedy* oder auch *Winner-takes-it-all* Algorithmus vor: Zunächst wird für jedes hypothetisierte Konturmodell mit Gleichung (2.4) eine individuelle Bewertung $F(m_i)$ berechnet und die Menge \mathbf{m} mit dem bestbewerteten Modell initialisiert. Anschließend wird jeweils das Modell m_j selektiert, das den Wert $F_{interact}(\mathbf{m} \cup \{m_j\})$ am stärksten erhöht. Die Auswahl ist beendet, wenn mit keinem der nicht selektierten Modelle eine Erhöhung von $F_{interact}(\mathbf{m})$ erreicht werden kann. Mit dieser Vorgehensweise wird in der Regel nur eine suboptimale Auswahl erreicht. In [Leo93] wird jedoch dargelegt, daß die resultierende Menge einen guten Kompromiß zwischen Laufzeit und Korrektheit im Sinne der obigen drei Bedingungen repräsentiert. Der Prozeß der Model Selection ist in Algorithmus 2.2 dargestellt.

Aufgrund der Strategie des *Winner-takes-it-all* Verfahrens können trotzdem weiterhin Konturmodelle selektiert werden, die zu einem großen Teil die gleichen Kantenelemente approximieren. Ziel ist es jedoch, eine möglichst disjunkte Beschreibung der Kantenelemente im Bild durch Konturmodelle zu erhalten. Aus diesem Grund wird in der vorliegenden Arbeit eine weitere Reduzierung der Menge der selektierten Modelle vorgenommen. Es werden diejenigen Konturmodelle verworfen, die mehr als einen prozentualen Anteil $\Delta_{edgel}(\cdot)$ der Kantenelemente eines anderen Konturmodells approximieren und gleichzeitig eine schlechtere individuelle Bewertung nach Formel (2.4) besitzen. Mit dieser Erweiterung werden redundante Konturmodelle eliminiert, ohne daß wesentliche Information verloren geht. Vielmehr ergibt sich eine entscheidende Reduzierung des Aufwands für den Gruppierungsprozeß in Kapitel 4, der häufig die fehlende Information überbrücken kann. Im folgenden Abschnitt werden die mit dem modifizierten Verfahren erzielten Ergebnisse und die hierzu verwendeten Parameter diskutiert.

Model Selection

```
 $\mathbf{m} = \emptyset, F_{max} = 0$   
 $m_{best} = \underset{m}{\operatorname{argmax}} F(m)$   
while  $F_{max} \leq F_{interact}(\mathbf{m} \cup \{m_{best}\})$  do  
   $\mathbf{m} = \mathbf{m} \cup \{m_{best}\}$   
   $F_{max} = F_{interact}(\mathbf{m})$   
   $m_{best} = \underset{m \notin \mathbf{m}}{\operatorname{argmax}} F_{interact}(\mathbf{m} \cup \{m\})$   
end while
```

Algorithmus 2.2: Model Selection in Anlehnung an [Pos97].

2.1.4. Ergebnisse der Kontursegmentierung

Die Ergebnisse der Kontursegmentierung bilden die Ausgangsbasis für die in dieser Arbeit diskutierten Gruppierungsprozesse (Kapitel 4). Aus diesem Grund erfolgt hier bereits eine Präsentation und Diskussion dieser Ergebnisse, mit denen eine Beschreibung des Bildinhaltes durch Geradenstücke, Kreis- und Ellipsenbögen erreicht wird.

Bilder

Die Evaluation der Kontursegmentierung wird unter anderem anhand von Bildern aus der **baufix**[®] Domäne vorgenommen. Diese Bilder stellen dabei eine exemplarische Auswahl von typischen Szenen dar, die sich im Rahmen des Projektes des „Situieren künstlichen Kommunikators“ ergeben. So werden acht Szenen mit einzelnen, separat positionierten Objekten und auch bereits montierten Baugruppen oder auch vollständig montiertem Roller oder Flugzeug betrachtet. Auch zwei Szenen mit anderen als den **baufix**[®] typischen Objekten werden hier diskutiert. Sie entstammen ebenfalls einer Spielzeug-Domäne, und bieten ähnliche Problemstellungen für eine konturbasierte Beschreibung und Strukturierung des Bildinhaltes. Die Bilder sind vor einem homogenen Hintergrund aufgenommen worden und sind in Anhang A auf Seite 197ff abgebildet. Die Kodierung der Bilder erfolgte im YUV-Farbraum, aus dem die Intensitätsinformation extrahiert wird, um eine Kontursegmentierung durchzuführen.

In der vorliegenden Arbeit wurden zur Aufnahme der Bilder Standardgeräte (Kamera, Framegrabber) eingesetzt, die zum Teil verrauschte Bildsignale bereitstellen (vermutlich begründet durch die mindere Qualität einzelner elektronischer Bauteile). Das erklärte Ziel einer Kontursegmentierung ist es, eine initial gute Beschreibung der relevanten Konturen im Bild zu erlangen, und die Anzahl der auf Rauschen basierenden Konturmodelle gering zu halten. Aus diesem Grund werden alle hier diskutierten Grauwertbilder vor dem eigentlichen Prozeß der Kantendetektion mit einem 3×3 Median gefiltert. Problematisch dabei ist, daß die Filterung immer zu einer Eliminierung relevanter, schwach ausgeprägter Diskontinuitäten führen kann, und damit die Detektion entsprechender Konturen verhindern kann. So kann beispielsweise die Veränderung der Beleuchtung bereits eine Filterung erforderlich oder vernachlässigbar machen. Die Entscheidung für oder gegen eine Vorfilterung muß also individuell für die gegebene Aufnahmesituation erfolgen. Für die verwendete Aufnahmeapparatur und das **baufix**[®] Szenario hat sich überwiegend eine Vorfilterung bewährt.

Der Prozeß der Kontursegmentierung verläuft wie bereits in Abbildung 2.1 dargestellt, wobei für die Kantendetektion eine 5×5 Operatormaske (vergleiche Abbildung 2.3(b)) und zur Approximation das Verfahren von Leonardis verwendet wird.

Parameter

Für die Kontursegmentierung sind in zwei Arbeitsschritten Parameter anzugeben. Zum einen wird in dem Schritt der Verdünnung zur Auswahl relevanter Kantenelemente

	Parameter	Wert
Verdünnung		
	ν_{high}	0.1 * maximale Gradientenstärke
	ν_{low}	0.3 * ν_{high}
Approximation		
	Saatfenster	7×7
	Expansionsfenster	4×4
	$\Delta_{accept}(\cdot)$	1.1
	$\Delta_{compatibility}(\cdot)$	1.0
	$\Delta_{continuity}(\cdot)$	2.0
	$\kappa_1(\cdot)$	1.0
	$\kappa_2(\cdot)$	0.5
	$\kappa_3(\cdot)$	3.5
	$\Delta_{edgel}(\cdot)$	0.85 %

Tabelle 2.1.: Der gewählte Parametersatz zur Ermittlung der initialen Konturmodelle. Die Parameter für die Verdünnung beschreiben Schwellwerte bezüglich der zugelassenen Gradientenstärke. Die Werte für die Saat- und Expansionsfenster sowie für $\Delta_{continuity}(\cdot)$ sind in Pixeleinheiten angegeben. Die Schwellwerte $\Delta_{accept}(\cdot)$ und $\Delta_{compatibility}(\cdot)$ beziehen sich auf eine quadratische Distanz in Pixeleinheiten. Die Werte κ_{1-3} sind dimensionslos.

mit dem Hysterese-Schwellwert Verfahren ein maximaler und minimaler Schwellwert (ν_{high}, ν_{low}) berücksichtigt. In der vorliegenden Arbeit wird die Wahl aus [Pos97] übernommen, in der der Wert für ν_{high} in Abhängigkeit vom maximal auftretenden Gradientenbetrag im Bild gewählt wird. Der Wert für ν_{low} wird in Abhängigkeit von ν_{high} bestimmt (siehe Tabelle 2.1).

Eine umfangreichere Anzahl von Parametern ist für die Konturapproximation in Geradenstücke, Kreis- und Ellipsenbögen mit dem Verfahren von Leonardis erforderlich. Hierfür werden geringfügig modifizierte Werte gegenüber der Implementation nach [Pop94] gewählt. Diese Werte sind auf einer umfangreichen Stichprobe für die **baufix**[®] Domäne ermittelt worden (siehe auch [Pos97]). Die Tatsache, daß hier nur minimale Modifikationen erforderlich sind, obwohl eine vollständig andere Domäne verwendet wird, unterstreicht in hohem Maße die Generalisierungsfähigkeit und Robustheit des gewählten Verfahrens zur Approximation. Der vollständige Parametersatz für die Kontursegmentierung ist in Tabelle 2.1 angegeben.

Ergebnisse der Konturapproximation

Der gesamte Prozeß der Kontursegmentierung ist in Abbildung 2.1 dargestellt, wobei exemplarisch auch Ergebnisse einer Kantendetektion dargestellt sind. Dieser Schritt der

Kantendetektion wird im weiteren nicht näher diskutiert, vielmehr wird hier auf die Ergebnisse der Konturapproximation eingegangen.

Das Ziel einer Konturapproximation ist die Identifikation von relevanten Konturen im Bild und deren Beschreibung durch Geradenstücke, Kreis- und Ellipsenbögen. Idealerweise beschreiben diese die Projektionen von Objektbegrenzungen, deren Detektion in der vorliegenden Arbeit unter anderem das angestrebte Ziel sein soll. Aufgrund der gewählten Domäne kann es jedoch ebenfalls zu einer Beschreibung von Texturen, wie etwa einer Holzmaserung auf den Leisten oder dem Gewinde von Schrauben, oder aber genereller, zu der Beschreibung von Schattenkanten kommen. Als weiteres Ziel einer Konturapproximation wird eine geringe Fragmentierung der realen Konturen angestrebt, so daß eine initial gute Beschreibung derselben vorliegt. Im folgenden erfolgt die Evaluation der erzielten Ergebnisse in qualitativer Form an acht Bildern aus der **baufix**[®] Domäne und zwei weiteren Bildern aus dem Umfeld von Modelllandschaften. Die Bilder und die entsprechenden Approximationsergebnisse sind in Anhang A (Seite 197ff) abgebildet. Zur eindeutigen Referenzierung werden die Bilder sukzessive durchnummeriert.

In allen Bildern können die wesentlichen Konturen und Strukturen detektiert und durch entsprechende Konturmodelle beschrieben werden. So werden beispielsweise für die Projektionen von den Leistenkörpern in den meisten Fällen deren sichtbare Umrisse durch entsprechende Geradenstücke approximiert (Bild 1-8). Fragmentierungen treten an Kreuzungen mit anderen Objekten auf (Bild 3) oder an Stellen, an denen der Kontrast der Leiste zum Untergrund im Intensitätskanal des Bildes schwach ausgeprägt ist (Bild 3, 6, 8). Zudem werden häufig Geradenstücke für die Approximation von Schattenkanten der Leisten (Bild 2,4) bzw. der kompletten Baugruppen generiert (Bild 7, 8), die zum Teil auch eine hohe Fragmentierung aufweisen (Bild 7). Die Löcher in den Leisten werden zum großen Teil durch fast vollständig geschlossene Ellipsenbögen approximiert (Bild 1, 8). An einigen Stellen fehlen jedoch die vorderen Konturen und es wird nur jeweils ein Ellipsenbogen für die Unter- und Oberkante im Inneren des Loches generiert (Bild 5, 6). Das Fehlen jeglicher Oberflächenstruktur, die auf die Holzmaserung der Leisten schließen ließe, ist auf die Filterung aus der Vorverarbeitung und Kantendetektion zurückzuführen. Dennoch können zumeist alle wesentlichen Konturen approximiert werden. Dies gilt zum Beispiel auch für die Rauten in den Bildern 4, 6, 7. Hier sind jedoch häufig die Schmalseiten zur unteren Kante der Raute nicht, oder nur durch kleine Ellipsenbögen approximiert worden (Bild 7). Auch dies läßt sich mit der Filterung begründen.

Die Detektion der Konturen für die Schraubwürfel gestaltet sich aufgrund ihrer komplexeren Struktur hingegen schwieriger. So können zwar die wesentlichen Umrandungen zumeist detektiert und durch Ellipsenbögen und zum Teil durch Geradenstücke approximiert werden (Bild 3, 5, 6), jedoch liegt allgemein eine höhere Fragmentierung vor. Auch die kreisrunden Oberflächen und Löcher in den Würfeln werden nur vereinzelt durch längere Ellipsenbögen approximiert (Bild 3). Dies begründet sich zum Teil durch die Perspektive der Aufnahme, aber auch durch Verdeckungen und schwachen Kontrast zum Untergrund oder auf dem Objekt. Häufig entstehen auch mehrere kurze, zum Teil überlappende Ellipsenbögen. Ähnlich starke Fragmentierungen der Ellipsenbögen lassen

sich auch auf den Reifen und Felgen beobachten (Bild 2, 6, 7). Hierfür ist der mangelnde Kontrast zu den dahinterliegenden Objekten (Bild 7) und in hohem Maße die texturähnliche Profilierung der Reifen verantwortlich. Solche Texturierungen führen auch zu einer nur mäßigen Segmentierung der Gewinde der Schrauben (Bild 3, 5). Die Windungen werden zwar schön durch Bögen und Geradenstücke beschrieben, der Umriss des Gewindes wird hingegen nur selten approximiert. So ist auch die Struktur eines Innengewindes der Grund für die stark fragmentierten Löcher der Schraubwürfel und Rautenmuttern (Bild 5, 6, 7). Die äußere Umrandung der Schraubköpfe wird zumeist durch Ellipsenbögen adäquat approximiert, während die Schmalseiten nur vereinzelt durch Geradenstücke approximiert werden können. Auch hierfür zeichnet sich zumeist die Filterung bzw. der schwache Kontrast zum Untergrund oder anderen Objekten verantwortlich.

Insgesamt kann geschlossen werden, daß ein Großteil der relevanten Konturen detektiert und approximiert werden konnte, wobei gleichzeitig eine Übersegmentierung aufgrund von Rauschen oder Textur wesentlich unterdrückt wurde. Dies zeigt sich auch in den komplexeren Baugruppen in den Bildern 7 und 8. Hier ist die Fragmentierung wesentlich durch Verdeckungen oder Schattenkonturen geprägt. Der allgemeine positive Eindruck bestätigt sich zudem durch die Ergebnisse auf einer anderen Bilddomäne (Bild 9, 10). So konnten auch für die Bilder mit den Modellen von Lkw's und Häusern alle wesentlichen Konturen detektiert und approximiert werden. In diesen Bildern begründet sich die Fragmentierung einzelner Konturen ebenfalls durch Textur (Schrift auf dem Lkw) und Schattenkonturen, oder vereinzelt auch schwachem Kontrast und Verdeckung. Eine Diskussion der Laufzeiten und der Anzahl der generierten Konturmodelle erfolgt im Zusammenhang mit der Ergebnisdiskussion der Gruppierungsprozesse in Kapitel 7.

2.2. Regionensegmentierung

Im Gegensatz zur Kontursegmentierung zielt die Regionensegmentierung auf die Identifikation homogener Bereiche in den Bilddaten ab. Diese *Regionen* stellen im Idealfall das Komplement zu den Konturen dar und beschreiben Projektionen von Objektflächen.

Die Homogenität einer Region bezieht sich dabei zumeist auf Merkmale wie Grauwert, Farbe, Textur oder Bewegungsinformation in den Bilddaten. Die Kriterien für Homogenität werden dann üblicherweise über die Differenz oder Varianz in den Merkmalen potentieller Nachbarn zu Referenzmerkmalen einer bereits vorhandenen Region definiert. Für die Anwendung des Homogenitätskriteriums werden in der Literatur verschiedene prinzipielle Verfahren vorgeschlagen. Sie verfolgen alle die Ziele

- homogene Regionen im Hinblick auf eine spezielle Charakteristik zu beschreiben,
- signifikante Unterschiede in der Homogenitätscharakteristik zwischen benachbarten Regionen zu garantieren und
- eine möglichst glatte, der Realität entsprechende Begrenzung homogener Regionen zu generieren.

Die einfachsten Techniken zur Regionensegmentierung sind *Clustering*- und *Schwelwert*-Verfahren. Sie basieren auf der Auswertung von Histogrammen, in denen Schwellwerte für eine geeignete Diskriminierung der Regionen gesucht werden. Mit Hilfe dieser Schwellwerte wird dann ein pixelweises *Region Labeling* durchgeführt. In der Regel werden diese Verfahren nur bei klar separierbaren Bildinhalten eingesetzt, da geringe Schwankungen in den Merkmalen oftmals zu pixelweisen Fehlsegmentierungen führen. Darüber hinaus ist der Einsatz von morphologischen Operatoren essentiell, da keine Ortsinformation berücksichtigt wird und lokale Störungen im Labeling nur so beseitigt werden können. Ein Überblick über die grundlegende Technik der Clustering- und Schwellwert-Verfahren und verschiedene Erweiterungen findet sich beispielsweise in [Har92].

Als weitere Verfahren zur Regionensegmentierung werden verschiedene Ausprägungen des *Region Growing* oder auch *Merging* eingesetzt. Im Falle des *Single Linkage Region Growing* werden direkt benachbarte Pixel auf die Erfüllung des Homogenitätskriteriums untersucht: Sind zwei Pixel bezüglich des Kriteriums ähnlich, werden sie zu einer Region zusammengefaßt. Eine iterative Anwendung dieser Regel auf benachbarte Pixel der Region führt zu einer sukzessiven Expansion der Region. Ein entscheidender Nachteil dieser Methode ist das sogenannte *Chaining* Problem. So können Pixel mit stark divergenter Homogenitätscharakteristik in einer Region zusammengefaßt werden, weil nur die direkten Randpixel einer Region die Referenzmerkmale für das Homogenitätskriterium stellen. Mit diesem Verfahren würde beispielsweise ein Graukeil im Bild als eine Region segmentiert. Aus diesem Grund werden im *Centroid Linkage Region Growing* die Referenzmerkmale einer Region durch die Merkmalsmittelwerte aller bereits zugeordneten Pixel repräsentiert. Mit dieser Methode können dann auch benachbarte Regionen mit ähnlichen Mittelwerten zu einer Region zusammengefaßt werden. Aber auch hier stellt sich die Problematik der Bearbeitungsreihenfolge. Bei unterschiedlichen Startpunkten können komplett unterschiedliche Ergebnisse generiert werden. Ein Verfahren, das diese Situation entschärfen will, ist das *Island Growing*, das in Kapitel 2.2.2 näher vorgestellt wird. Weitere Ansätze, die der Problematik der Bearbeitungsreihenfolge entgegenwirken wollen, sind Verfahren mit *Split-and-Merge* Technik: Initial wird das gesamte Bild als eine Region aufgefaßt und sukzessive in regelmäßige Blöcke, häufig entsprechend einem Quadtree, unterteilt, bis in jedem Block für alle Pixel nach dem Centroid-Linkage das Homogenitätskriterium erfüllt ist (Splitting). Aus diesen initialen Regionen werden dann benachbarte Regionen nach dem Centroid-Linkage Verfahren wieder verschmolzen (Merging, siehe z.B. [Gon91]). Andere Verfahren kombinieren ein Merging mit einer Kantendetektion, um eine bessere Trennung zwischen homogenen Bereichen zu erzielen. Diese Ansätze werden im Kapitel 5 ausführlicher diskutiert. Neben diesen domänenunabhängigen Verfahren, sind auch Verfahren entwickelt worden, die Informationen über die im Bild dargestellten Objekte ausnutzen. Ein Verfahren dieser Klasse kommt auch in der vorliegenden Arbeit zum Einsatz und wird im folgenden Abschnitt vorgestellt. Im Anschluß daran wird das domänenunabhängige Island Growing beschrieben. Im letzten Abschnitt dieses Kapitels werden die Ergebnisse beider Verfahren diskutiert.

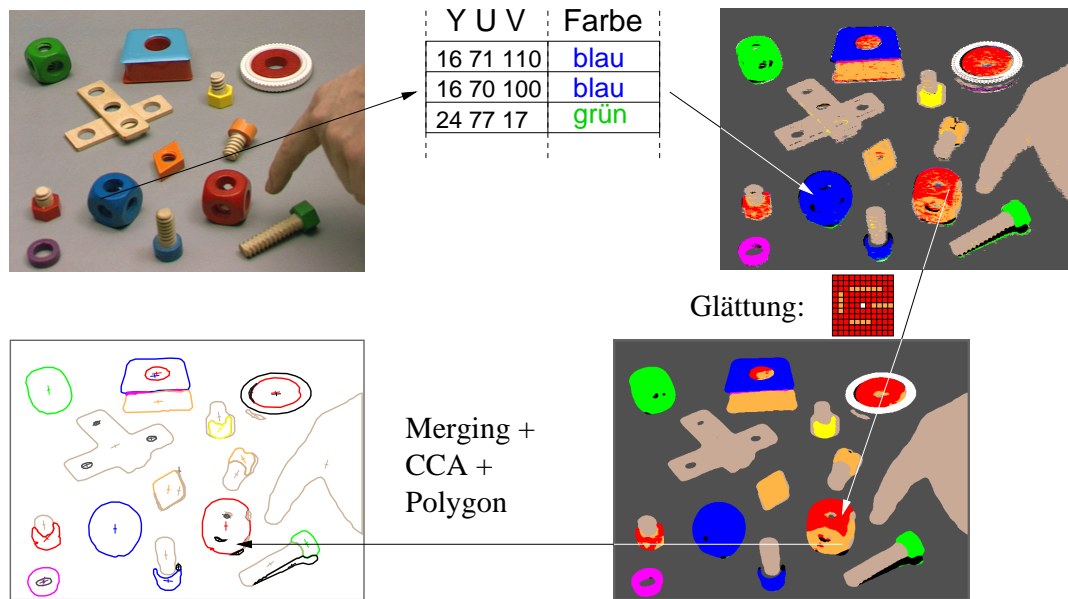


Abbildung 2.7.: Regionensegmentierung mit pixelweiser Klassifikation, Glättung, Merging, Connected Component Analyse (CCA) und Approximation durch Polygone. Hände werden als Holz klassifiziert, da die Farbcharakteristik von Holz ähnlich ist. Unbekannte Objekte (Klebefilmroller am oberen Bildrand) werden möglichst auf **baufix**[®] Farben korrekt abgebildet, was die Robustheit des Ansatzes unterstreicht. Objekte, die dem Farbenspektrum nicht entsprechen werden mit der Farbe *undefiniert* belegt.

2.2.1. Pixelweise Klassifikation

Die in der **baufix**[®] Domäne verwendete Segmentierung ist ein Verfahren, das Domänenwissen über den Bildinhalt zur Regionendetektion einsetzt. Es nutzt die Tatsache, daß in der **baufix**[®] Objektwelt nur 12 verschiedene Farben auftreten können. Diese Farben sind durch die Lackierung der **baufix**[®] Objekte (weiß, rot, gelb, orange, blau, grün, lila, holzfarbig, elfenbein, schwarz), den Untergrund auf dem die Objekte plaziert werden (grau) und eine nicht definierte Farbklasse (undefiniert) für unbekannte Objekte vorgegeben.

Der Ansatz startet mit der Klassifizierung jedes einzelnen Pixels des Eingangsbildes in eine der 12 **baufix**[®] Farben. Bei dieser Klassifizierung werden die YUV Farbkomponenten jedes Pixels als ein Merkmalsvektor aufgefaßt und mit einem Polynomklassifikators sechsten Grades klassifiziert. Zur Beschleunigung dieses zeitaufwendigen Prozesses sind alle YUV Kombinationen in einer *Look-Up* Tabelle gespeichert und vorklassifizierten Farben zugewiesen. Aus speichertechnischen Gründen wurde eine Reduzierung der Farbauflösung durchgeführt, die aber keine Verschlechterung der Ergebnisse bewirkt. Da zur Klassifikation nur lokale Information der einzelnen Pixel zur Verfügung steht, kann aufgrund von Signalrauschen eine Fehlklassifizierung in einzelnen Fällen auftreten. Eine

Reduzierung dieser Fehlsegmentierungen wird durch die Faltung des klassifizierten Bildes mit einer 11×11 Glättungsmaske erreicht. Dieser Operator setzt das aktuelle Pixel jeweils auf den Wert der maximal auftretenden Farbe.

Weitere Fehlsegmentierungen können durch Lichtreflektionen oder Objektschatten hervorgerufen werden. An dieser Stelle greift das Verfahren auf eine Merging Technik zurück: Benachbarte Regionen mit ähnlicher Farbe werden vereinigt, wenn das Verhältnis der Länge der gemeinsamen Regionengrenze zur Länge der Begrenzung der kleineren Region über einem Schwellwert liegt. Die Ähnlichkeit von Farben wurde hierzu auf der Basis von Fehlklassifizierungen in einer Stichprobe bestimmt. Die Begrenzung des Längenverhältnis wurde eingeführt, damit Regionen ähnlicher Farbe, die aber verschiedene Objekte im Bild beschreiben, nicht zusammengefaßt werden. Die aus der Vereinigung neu entstehende Region wird dann auf die Farbe der flächenmäßig größeren Region gesetzt. Nach diesem Merging wird für alle Regionen eine *Connected-Component* Analyse zum Region Labeling eingesetzt und schließlich die Regionengrenzen durch Polygonzüge approximiert. Gleichzeitig werden noch Regionenmerkmale wie Kompaktheit, Exzentrizität, etc. für eine spätere Objekterkennung berechnet. Abbildung 2.7 zeigt das Verfahren an einem Beispiel, für eine ausführlichere Diskussion siehe [Kum97]. In Abschnitt 2.2.3 werden Ergebnisse dieses Ansatzes vorgestellt und diskutiert.

2.2.2. Island Growing

Mit dem Island Growing nach [Pri93] wird nun ein Verfahren beschrieben, das auch in anderen Domänen mit einer entsprechenden Parametrisierung eingesetzt werden kann. Es gehört zu der Gruppe der Merging Verfahren, verfolgt dabei aber das Ziel eine weniger starke Abhängigkeit von der Verarbeitungsreihenfolge zu besitzen.

Die Idee des Island Growing beruht, ähnlich dem Ansatz von Leonardis zur Kontursegmentierung, auf der Einteilung des Bildes in ein regelmäßiges Raster, sogenannte *Inseln* (vergleiche Abschnitt 2.1.3, Abbildung 2.6(a)). Das Region Growing erfolgt dabei zunächst unabhängig, nur innerhalb der Grenzen einer Insel. Hierzu wird der Centroid Linkage Ansatz eingesetzt. Diese Begrenzung des Regionenwachstums auf die einzelne Insel minimiert den Einfluß einer Bearbeitungsreihenfolge, und eine parallele Bearbeitung ist prinzipiell denkbar. Ist der Region Growing Prozeß in jeder Insel abgeschlossen, werden benachbarte Inseln nach dem Prinzip eines Quadrees zu größeren Inseln zusammengefaßt. Innerhalb dieser größeren Inseln findet dann wieder ein Verschmelzen von benachbarten Regionen gemäß dem Centroid Linkage statt. Der gesamte Prozeß startet mit jedem Pixel als eine Insel und verfolgt eine stufenweise Vereinigung von Inseln, bis das ganze Bild einer Insel entspricht. Mit dieser Vorgehensweise wird das Chaining Problem lokal auf die Inseln begrenzt.

Für das Homogenitätskriterium können verschiedene Merkmale eingesetzt werden. In [Kor94] sind mehrere Varianten untersucht worden und der *City-Block Abstand* über die Farbmittelwerte der Regionen als besonders geeignet ermittelt worden. Die Hinzunahme statistisch motivierter Merkmale (Varianz, etc.) führen nach [Kor94] zu kei-

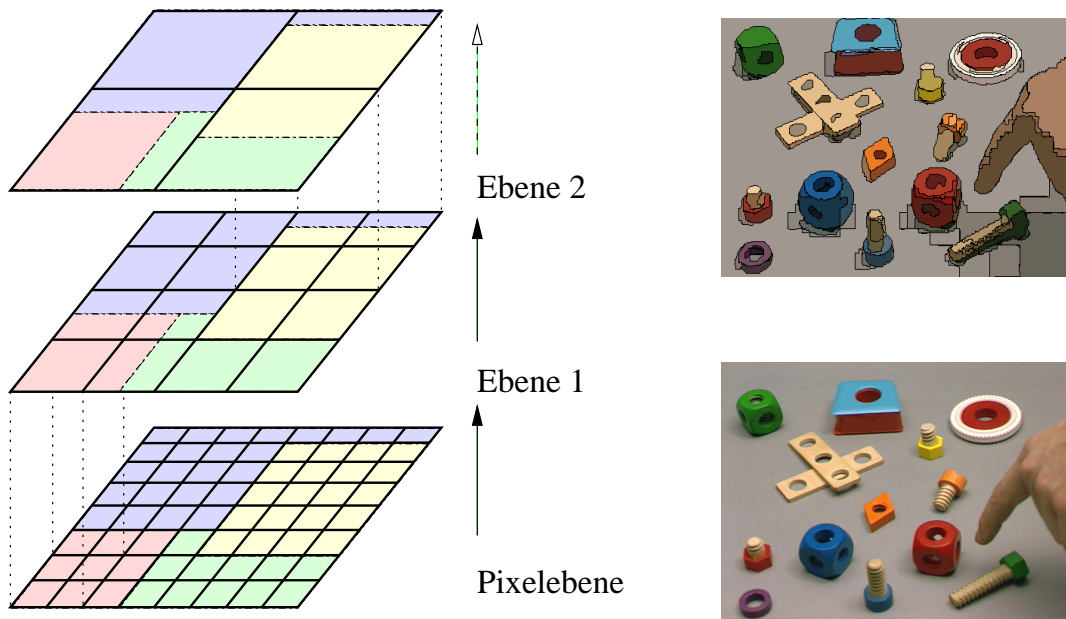


Abbildung 2.8.: Regionensegmentierung mit dem Island Growing. Im linken Teil ist die prinzipielle Verfahrensweise mit Region Growing und Vereinigung benachbarter Inseln dargestellt. Inselgrenzen sind mit durchgezogenen Linien und Regionengrenzen innerhalb der Inseln mit Strichpunkt-Linien markiert. Auf der rechten Seite ist das Ausgangsbild dargestellt und im oberen Teil die vollständige Segmentierung mit den homogenen Regionen ähnlicher Farbe gezeigt.

ner wesentlichen Verbesserung der Ergebnisse. Das Homogenitätskriterium wird über einen Schwellwert Δ_{island} bezüglich dem City-Block Abstand zwischen Referenzvektoren \mathbf{c} zweier Regionen definiert:

$$d(\mathbf{c}(R_1), \mathbf{c}(R_2)) \leq \Delta_{island}$$

In Abbildung 2.8 wird das prinzipielle Vorgehen veranschaulicht. Ein Vorteil gegenüber dem klassifikationsbasierten Verfahren aus Abschnitt 2.2.1 ist die Unabhängigkeit vom zu bearbeitenden Bildinhalt. Das Island Growing kann bei entsprechender Parametrisierung in jeder Domäne und jedem Farbraum eingesetzt werden, ohne ein neues Training für einen Klassifikator zu benötigen. In nächsten Abschnitt wird der Einsatz des Verfahrens auf einem Beispielbild und die entsprechende Wahl des Parameter Δ_{island} diskutiert.

2.2.3. Ergebnisse der Regionensegmentierung

Die Ergebnisse der Regionensegmentierung stellen eine weitere Ausgangsbasis für die in Kapitel 5 vorgestellten Gruppierungsprozesse zwischen Regionen und Konturen dar. Da

die Regionensegmentierung jedoch nicht zentraler Bestandteil der vorliegenden Arbeit ist, werden die Ergebnisse hier nur im Hinblick typischerweise auftretender Segmentierungsphänomene beschrieben. Dies erfolgt ebenfalls anhand der Bilder in Anhang A (Seite 197ff).

Das Ziel der klassifikationsbasierten Regionensegmentierung im Szenario des „Situieren künstlichen Kommunikators“ ist die Segmentierung vollständiger Objektoberflächen auf Basis der entsprechend trainierten Farben. Dies gelingt für den Großteil der abgebildeten Objekte sehr gut (Bild 1-8). Innerhalb eines Objektes sollen zudem möglichst die andersfarbigen Objektbestandteile oder -strukturen segmentiert werden. Dies gilt zum Beispiel für die Löcher in den Leisten (Bild 1) oder den Kopf und das Gewinde von Schrauben, die in der Regel durch zwei benachbarte Regionen beschrieben werden können (Bild 5). Die Löcher in Schraubwürfeln oder Rautenmuttern können hingegen in den meisten Fällen nicht geeignet segmentiert werden (Bild 3, 4, 6). Hier kann nur selten eine Trennung erfolgen, da die Lochinnenseiten im wesentlichen die gleiche Farbe wie die Objektoberfläche besitzen. Zumeist werden daher nur vereinzelte kleine und andersfarbige Regionen für das Loch gebildet, die dann zudem durch das Merging mit den der Gesamtregion des Objektes verschmolzen werden. Generell erfolgt in hohem Maße eine Verschmelzung von Regionen bei der Überdeckung von gleichfarbigen Objekten. So werden überlappende Leisten in den Bildern 2, 3, 4, 6 und 8 immer als eine Region segmentiert. Auch die Positionierung einer roten Schraube in einer roten Felge kann in Bild 4 nicht geeignet getrennt werden. In Bild 7 sind sehr starke Fehlsegmentierungen aufgetreten, wodurch beispielsweise die beiden Reifen überhaupt nicht segmentiert werden, bzw. wird ein Teil des vorderen Reifens als holzfarbig erkannt. Dies begründet sich aus der engen Nachbarschaft der Farbwerte von weiß und holzfarbig. Berühren sich die Projektionen von holzfarbigen und weißen Objekten, so wird häufig *eine* holzfarbige Region segmentiert. Auch das Verschmelzen der Schraube und der Felge zu einer blauen Region erklärt sich durch eine stark verrauschte Farbklassifikation der einzelnen Bildpunkte, die dann zu einer blauen Region verschmolzen werden. Zudem zerfallen die Regionen der Objekte aufgrund von Abschattungen oder Verdeckungen in diesem Bild sehr stark. Das gänzliche Fehlen des hinteren Reifens ist anschaulich nicht erklärbar. Insgesamt kann jedoch geschlossen werden, daß in den überwiegenden Fällen eine Segmentierung der Objekte gut erfolgen konnte, wobei mit der Komplexität der Baugruppen und ungünstigeren Beleuchtungsbedingungen eine deutliche Verschlechterung eintreten kann. Das Laufzeitverhalten für diesen Ansatz ist relativ unkritisch, da die Klassifikation von der Anzahl der Bildpunkte und das Glätten, Verschmelzen und die polygonale Approximation von der Anzahl und Größe der Regionen abhängt. Somit ergibt sich je nach Anzahl der farblich separierbaren Regionen eine Laufzeit von ca. 200-800ms/Bild auf einer Alpha Workstation 500/400MHz.

Das Verfahren des Island Growing wird in der vorliegenden Arbeit eingesetzt, um das Gruppieren von geschlossenen Figuren (Kapitel 4.4) auf einer anderen Bilddomäne und alternativen Segmentierung zu testen. Das hierfür verwendete Bild ist in Abbildung A.12 abgebildet, und zeigt einen Gullydeckel mit kreisförmig angeordneten Strukturen. Das

Island Growing wird im RGB-Farbraum durchgeführt, wobei für die gezeigte Segmentierung das Homogenitätskriterium über den City-Block Abstand definiert ist. Vor der eigentlichen Bearbeitung mit dem Island Growing wurde eine Filterung jedes Farbkanals mit einem 5×5 Median durchgeführt, um sehr feine Strukturen zu unterdrücken. Anschließend wurde der Wert $\Delta_{island} = 50$ gewählt, und eine minimale Regionengröße von 100 Bildpunkten gefordert. In Abbildung A.12 sind die Ergebnisse der Segmentierung mit diesen Parametern dargestellt, wobei nur Regionen gezeigt sind, die weniger als 2500 Bildpunkte enthalten, da die größeren Regionen keine signifikanten Strukturen für lineare Anordnungen im Bild beschreiben. Es ist sehr schön zu sehen, daß überwiegend eine gute Trennung der einzelnen Steine und Löcher erzielt werden konnte. Die ‚rechteckigen‘ Bereiche, in denen die Löcher auf dem Deckel enthalten sind, wurden zum Teil mit den größeren Regionen verschmolzen. Hingegen zerfallen einige Steine des Ringes in Randbereichen in kleinere Regionen. Insgesamt weist dieses Ergebnis darauf hin, daß dieses Verfahren mit einer entsprechenden Parametrisierung möglicherweise in unterschiedlichsten Bilddomänen eingesetzt werden kann.

Bevor in Kapitel 4 das konturbasierte Gruppieren und in Kapitel 5 die Kombination von Konturen und Regionen vorgestellt wird, erfolgt im nächsten Kapitel eine Einführung und Übersicht über die Prinzipien des perzeptiven Gruppierens.

3. Perzeptives Gruppieren

Im letzten Kapitel wurden verschiedene Methoden vorgestellt, die einen grundlegenden Bestandteil eines Bildanalyse-Systems darstellen. Ohne explizit darauf hinzuweisen, sind bei den dort diskutierten Bildsegmentierungstechniken bereits grundlegende Mechanismen des perzeptiven Gruppierens zum Einsatz gekommen. So entspricht beispielsweise die Approximation einer Pixelkette in ein Geradenstück auch einer Gruppierungsleistung. Im allgemeinen wird unter dem Begriff des perzeptiven Gruppierens die Fähigkeit eines biologischen oder auch künstlichen Wahrnehmungssystems verstanden, Sensordaten unter bestimmten Prinzipien in größere Einheiten zu gruppieren bzw. zu organisieren.

Perzeptives Gruppieren befähigt damit auch den Menschen, die ständig auf ihn einströmende Reizflut, seien es visuelle oder auditive Reize, spontan und unbewußt in strukturierten Gruppen wahrzunehmen. Dieses Gruppierungsverhalten unterstützt dann in starkem Maße höhere kognitive Prozesse, wie eine Objekterkennung und Spracherkennung. In Bezug auf die visuelle Wahrnehmung wurden diese Phänomene erstmalig zu Beginn des letzten Jahrhunderts von Wissenschaftlern aus dem Bereich der Psychologie untersucht und in Form von Gestaltgesetzen formuliert. Diese Gesetze beschreiben grundlegende Prinzipien, die ein perzeptives Gruppieren bzw. eine perzeptive Organisation von Teilen zu Gestalten bewirken. In Abschnitt 3.1 wird auf diese Gesetzmäßigkeiten und deren Relevanz für Bildanalyse-Systeme eingegangen.

Die in den Gestaltgesetzen formulierten Gruppierungsprinzipien sind bereits in Teilen in Bildanalyse-Systeme integriert worden. Eine entscheidende Problematik ergibt sich jedoch aus der ungenauen Spezifikation der Gesetzmäßigkeiten. Von den Psychologen sind die Gruppierungsphänomene nur qualitativ beschrieben worden, es fehlt eine quantitative formale Beschreibung für deren Anwendung oder Nachbildung. Aus diesem Grund sind viele Arbeiten zum perzeptiven Gruppieren entstanden, die unterschiedliche, zum Teil problemabhängige algorithmische Formulierungen dieser Gesetzmäßigkeiten vorschlagen. In Abschnitt 3.2 wird ein allgemeiner Überblick gegeben und einzelne Ansätze vorgestellt. In den Kapiteln 4, 5 und 6 werden die im Zusammenhang mit der vorliegenden Arbeit entwickelten Ansätze detailliert dargestellt.

3.1. Gesetzmäßigkeiten zur Gruppenbildung

Bereits am Anfang des letzten Jahrhunderts wurde von den Gestaltpsychologen beobachtet, daß die Wahrnehmungsleistung des menschlichen Sehsystems durch die Strukturierung und Organisation von einzelnen Empfindungen zu größeren Einheiten, den



Abbildung 3.1.: Beispiel zur Wahrnehmung von Gestalten nach Sarkar und Boyer [Sar94b].

Gestalten, bestimmt wird [Wer23]. Dabei ist die Wahrnehmung einer Gestalt nicht nur durch eine Summierung von Einzelwahrnehmungen zu erklären: „*Das Ganze ist verschieden von der Summe seiner Teile*“. Dieser Hauptlehrsatz der *Gestalt-Psychologie* untermauert die Komplexität der Wahrnehmung und stellt explizit heraus, daß erst das Zusammenwirken einzelner Stimuli in einem Kontext die eigentliche Struktur des Ganzen erzeugt. Dies läßt sich durch das Beispiel in Abbildung 3.1 motivieren: Die symmetrisch gespiegelten Buchstaben lassen Worte zunächst verschwinden, ihre Gestalt ist verändert, und die *spontane* Wahrnehmung von Schrift wird behindert. Wird der obere Teil abgedeckt, treten spontan die emergenten Eigenschaften der Gestalt von Worten in den Vordergrund. Neben diesen Effekten eines visuellen Organisationsprozesses, werden die Konzepte der Gestalt-Psychologie auch in anderen Bereichen beobachtet. So lassen sich beispielsweise ähnliche Phänomene in der akustischen Wahrnehmung ausmachen.

3.1.1. Gestaltgesetze

Aus den Beobachtungen zur visuellen Wahrnehmung beim Menschen entstanden die Bemühungen, die Gesetzmäßigkeiten der visuellen perzeptiven Organisation aufzudecken und zu beschreiben. Die Ergebnisse dieser Untersuchungen formulierten die Gestalt-Psychologen in den sogenannten *Gestaltgesetzen*. Diese Gesetze beschreiben einzelne Gruppierungsphänomene, die Bestandteile der Organisationsprozesse sind, und auch heute noch grundlegende Motivation für Forschungen im Bereich des perzeptiven Gruppierens sind. In Abbildung 3.2 sind einige dieser klassischen Gestaltgesetze anschaulich motiviert (vergleiche auch [Wer23, Roc91, Pal94]):

Nähe: Die Distanz zwischen den Geradenstücken in Abbildung 3.2 führt zu einer Gruppierung mit dem Eindruck gestreifter Rechtecke.

Ähnlichkeit: Die Gruppierung aufgrund von Ähnlichkeit kann durch Eigenschaften wie Farbe, Größe, Orientierung und Form von Einzelteilen hervorgerufen werden.

gemeinsamer Bereich: Objekte, die sich in *einem* abgeschlossenen Bereich befinden, werden vom Menschen zu einer Gruppe zusammengefaßt.

Geschlossenheit: Strukturen, die scheinbar eine geschlossene Form beschreiben, werden als geschlossene Einheit wahrgenommen.

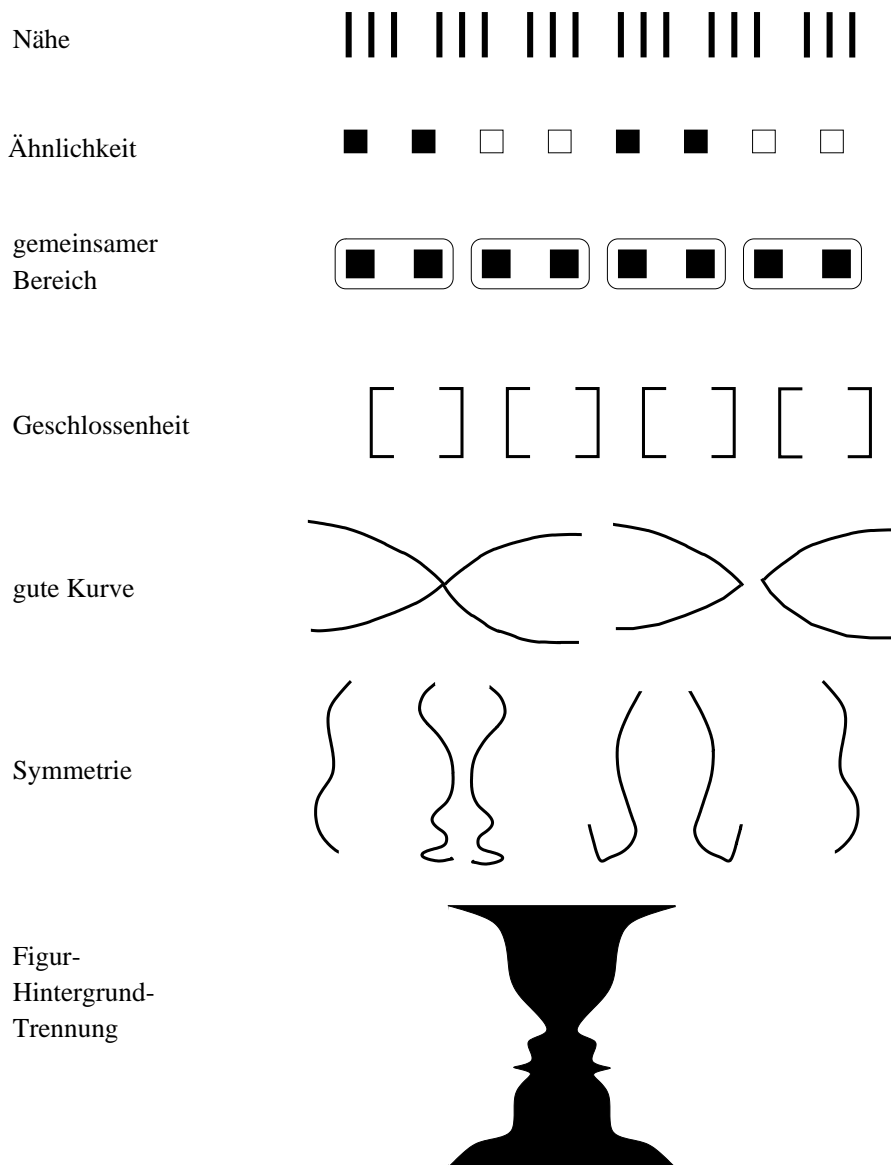


Abbildung 3.2.: Motivation einiger Gestaltgesetze.

gute Kurve: Linienzüge werden aufgrund der Glattheit ihres Verlaufs gruppiert. Die Störung der Glattheit bewirkt ein abweichendes Gruppierungsverhalten.

Symmetrie: Symmetrien werden ebenfalls bevorzugt gruppiert. Sie sind ein sehr wichtiges Gruppierungsphänomen, das in vielen Wahrnehmungen von entscheidender Bedeutung ist.

Figur-Hintergrund-Trennung: Dieses Gesetz ist entscheidend für die Wahrnehmung von

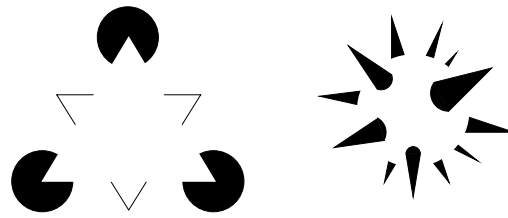


Abbildung 3.3.: Wahrnehmung von Scheinkonturen (siehe z.B. [Kan79]).

Objekten. Eine Region kann als Objekt oder als Hintergrund wahrgenommen werden. Im Beispiel ist Rubins Gesicht/Vase Illusion dargestellt.

Darüber hinaus befaßt sich die Gestalt-Psychologie noch mit weiteren visuellen Wahrnehmungsphänomenen. So wurde beispielsweise auch das Gesetz des *gemeinsamen Schicksals* formuliert. Bei diesem Gesetz findet eine Gruppenbildung von Bildinhalten statt, die sich mit identischer Geschwindigkeit in *eine* Richtung bewegen. Ein Beispiel für dieses Gesetz ist der keilförmige Formationsflug von Zugvögeln. Jeder Mensch erkennt sofort die keilförmige Struktur, die sich aus den Vögeln zusammensetzt. Ein weiteres Phänomen spiegelt sich in der Wahrnehmung von Scheinkonturen wider. Der Mensch tendiert dazu, verdeckte oder fehlende Konturen durch möglichst einfache und glatte Vervollständigungen zu ergänzen (siehe Abbildung 3.3). Dieses von Kanizsa intensiv diskutierte Phänomen [Kan79], wird zum Beispiel im Bereich der Bildanalyse in den Arbeiten [Tho96, Wil98] untersucht. Ein weiterer wichtiger Faktor in der Wahrneh-

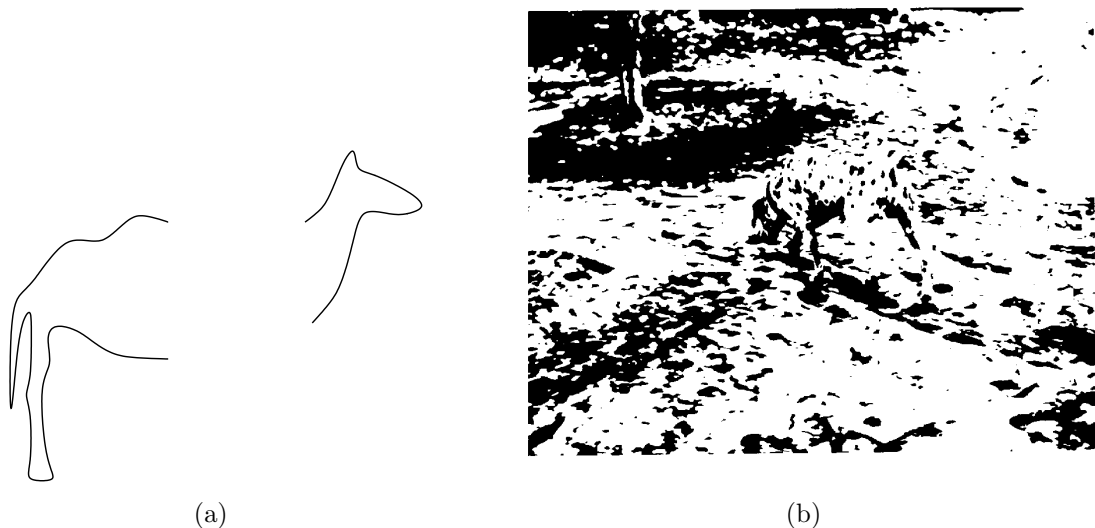


Abbildung 3.4.: Objektwissen und Erfahrung prägen die Erkennungsleistung.

mungsleistung ist der Einfluß von Erfahrung und Wissen aus früheren Wahrnehmungen. Werden in den visuellen Perzepten Strukturen oder Teile von Strukturen erkannt, die bereits in ähnlicher Form bekannt sind, so wird das Wahrgenommene leichter auf diese bereits bekannten Strukturen oder Objekte abgebildet. In Abbildung 3.4 sind Beispiele für diese Effekte gezeigt.

Eng verbunden mit diesem Aspekt ist der von der Gestalt-Psychologie eingeführte Begriff der *Prägnanz* einer Struktur. Dieser Begriff wird für Strukturen benutzt, die durch ihre Einfachheit und Stabilität (zum Teil durch eine starke Symmetrie begründet) die ausgeprägteste Struktur in der Szene darstellen. Je einfacher die Struktur, desto weniger Information ist zu ihrer Beschreibung nötig. Darum werden einfache, aber teilweise verdeckte Strukturen, vom Betrachter spontan ergänzt [Roc91]. Die Prägnanz befähigt den Menschen beispielsweise auch, einfache symmetrische Strukturen besser zu behalten, zu beschreiben und zu vergleichen. Die Bedeutung des Begriffes Prägnanz ist damit inhärent mit der Wechselwirkung verschiedener Gruppierungsphänomene und einer gemeinsamen kausalen Ursache verbunden. Auf diese Aspekte wird im nächsten Abschnitt eingegangen.

3.1.2. Wechselwirkungen und Kausalität

Die einzelnen Gesetze können nicht isoliert für die Beschreibung von Organisationsprozessen herangezogen werden. Dies wird bereits in der Abbildung 3.2 bei dem Gesetz der Ähnlichkeit deutlich. Werden die unterschiedlich gefärbten Quadrate näher zusammen gebracht, so gewinnt die Nähe-Gruppierungswirkung an Dominanz gegenüber der Ähnlichkeit. Verschiedene Arbeiten befassen sich mit dieser Thematik, wobei eine generalisierte Anschauung und Formulierung noch aussteht. Allen Untersuchungen auf diesem Gebiet ist gemeinsam, daß sie sich auf psychophysische Tests beziehen, um die Wechselwirkung der unterschiedlichen Gruppierungsprinzipien zu erklären. Im folgenden werden einige Arbeiten betrachtet, die sich mit der Wechselwirkung von Gruppierungen aufgrund von Nähe und Ähnlichkeit befassen.

In [Hoc76] wird die Wechselwirkung von Ähnlichkeit in der Helligkeit der Stimuli gegenüber einer Nähebeziehung zwischen den Stimuli untersucht. Hochberg und Silverstein verwenden hierzu Quadrate, die grau schattiert oder weiß sind und auf einem 6×6 Gitter angeordnet sind. Durch Veränderungen des Abstandes zwischen den Spalten und Veränderungen der Helligkeit, sowie den daraus resultierenden Wahrnehmungen der Versuchspersonen, postulieren sie eine Abhängigkeit des Ähnlichkeitsempfindens von der Nähebeziehung der betrachteten Objekte (siehe Abbildung 3.5(a)). Sie selbst merken jedoch an, daß diese Ergebnisse einigen Einschränkungen unterliegen. Die Stärkste davon ist, daß die Einfachheit der verwendeten Symbole und deren künstliche Anordnung generalisierende Rückschlüsse vermeiden. In ähnlicher Weise untersuchen Ben Av und Sagi in [Ben95] die Wechselwirkung von Nähe gegenüber Form-Ähnlichkeit und Helligkeits-Ähnlichkeit. Sie verwenden die unterschiedlichen Symbole „X“ und „L“ als Bildprimitiva (siehe Abbildung 3.5(b)). Neben der Variation von Abständen und Hel-

3. Perzeptives Gruppieren

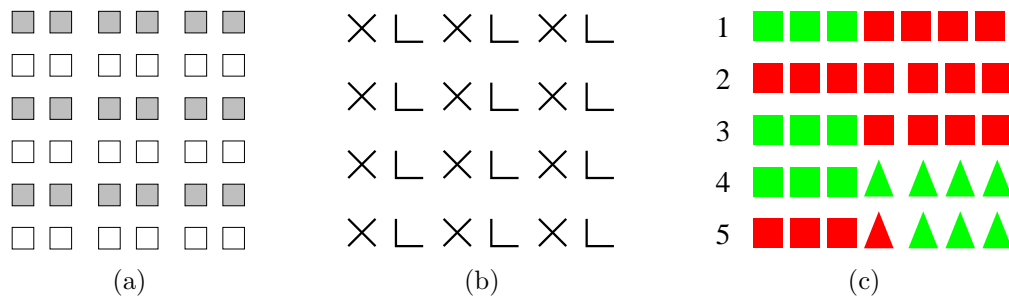


Abbildung 3.5.: Stimuli zur Untersuchung von Wechselwirkungen zwischen Nähe und Ähnlichkeits-Gruppierung. (a) Hochberg und Silverstein [Hoc76], (b) Ben Av und Sagi [Ben95], (c) Quinlan und Wilton [Qui98].

lichkeit, wurden den Probanden die entsprechend unterschiedlichen Konfigurationen für bestimmte Zeitintervalle präsentiert. In der Auswertung wird eine Dominanz von Nähegruppierungen bei kurzen Präsentationen festgestellt, wohingegen bei längerer Betrachtungsdauer die Ähnlichkeitsbeziehungen an Dominanz gewinnen. Auch in [Qui98] wird die Wechselwirkung von Ähnlichkeiten bezüglich Farbe und Form, sowie der Einfluß von Nähe untersucht. Entsprechend variierte Konfigurationen wurden den Probanden für einen Zeitraum von 200ms präsentiert, die dann eine Aussage über ihr spontanes Gruppierungsverhalten mit entsprechender Gruppierungsstärke anstellen sollten (siehe Abbildung 3.5(c), eine Zeile entspricht *einer* variierten Konfiguration). Als Ergebnis berichten Quinlan und Wilton, daß die Gruppierungswirkung aufgrund von Nähe in allen Darstellungen starken Einfluß hat. Stehen jedoch zum Beispiel die Farbähnlichkeit und die Nähe in Konkurrenz zueinander (Zeile 3 in Abbildung 3.5(c)), so entschieden sich die Hälfte der Probanden für eine stärkere Gruppierungswirkung aufgrund von Nähe, und die andere Hälfte für einen stärkeren Wirkung aufgrund von Farbe. Weiterhin weisen sie daraufhin, daß ein starker Einfluß von Farbe im Vergleich zu einer schwachen Gruppierungswirkung aufgrund der Form durch die Wahl der Farben und Formen erklärt werden kann. Während sich Grün und Rot stark unterscheiden, sind die Formunterschiede zwischen Dreieck und Quadrat weniger markant. Auch aktuelle Arbeiten beschäftigen sich mit dieser Problematik, und verfolgen das Ziel eine formale Beschreibung der Gruppierungsphänomene zu erlangen. So werden in [Kub00] beispielsweise aus psychophysischen Experimenten quantitative Formulierungen für Nähegruppierungen auf stabilen Punktgittern entwickelt. Dabei erfährt auch die Ähnlichkeit in der Helligkeit der Punkte Berücksichtigung. Für das in Abschnitt 4.4 entwickelte Verfahren zur Gruppierung linear angeordneter Regionen wird die Wechselwirkung von Form-Ähnlichkeit und Nähe ebenfalls berücksichtigt, und eine algorithmische Lösung vorgeschlagen (siehe auch [Lan00]).

Neben der Untersuchung von Wechselwirkungen der verschiedenen Gestaltgesetze, stellt sich darüber hinaus die Frage nach einer Begründung einer strukturierten Wahrnehmung. Hierzu haben verschiedene Wissenschaftler unabhängig voneinander eine These

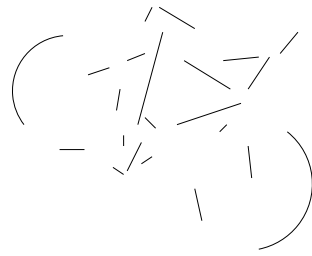


Abbildung 3.6.: Motivation zum Prinzip des *non accidentalness* (siehe auch [Low85]).

über die Kausalität einer wahrgenommenen Struktur aufgestellt. Das Prinzip des „*non-accidentalness*“ oder auch „*common cause*“ wurde von Witkin und Tennenbaum [Wit83], und Lowe [Low85] in ähnlicher Weise formuliert:

Ist die Wahrscheinlichkeit des Auftretens einer bestimmten Organisation sehr gering, so deutet die Erkennung einer solchen Organisation auf eine kausale Ursache hin.

Bei der Projektion der physikalischen Welt in ein Bild bleiben die inhärenten Strukturierungen und Regelmäßigkeiten der realen Welt zu einem hohen Anteil erhalten. Somit treten regelmäßige Strukturen wie Parallelität oder Symmetrie selten rein zufällig in Bildern auf. Sie deuten in der Regel auf einen kausalen Zusammenhang hin, und implizieren idealerweise die Projektion eines Objektes. Abbildung 3.6 zeigt ein Beispiel für dieses Prinzip. Die beiden Halbbögen erzeugen die charakteristische Struktur eines Fahrrades, fehlen sie kann der Mensch die gezeigte Struktur deutlich schwerer oder gar nicht identifizieren. Gerade künstliche Objekte zeichnen sich durch einfache, regelmäßige Strukturen, wie beispielsweise Parallelität aus. Jedoch besitzen auch natürliche Objekte eindeutige Strukturen, die eine Identifizierung der Objekte hervorrufen. Diese Strukturen sind ebenfalls oft symmetrisch und zeichnen sich durch ihre Einfachheit aus.

Das Prinzip des kausalen Zusammenhangs ist eine starke Motivation die Gruppierungsphänomene in der automatischen Bildanalyse algorithmisch nachzubilden. Deren Ziel ist es ja gerade, die Projektion der realen Welt zu analysieren und Objekte oder andere Ereignisse zu erkennen. Hierbei unterstützen die Gruppierungsmechanismen zum einen die Detektion solcher Strukturen, und zum anderen können diese Strukturen explizit für weitergehende Verarbeitungsprozesse repräsentiert werden.

3.2. Gruppierung in der Bildanalyse

Im letzten Viertel des letzten Jahrhunderts wurde zunehmend die Relevanz der Gestaltgesetze für eine automatische Bildanalyse erkannt [Wit83, Low85, Sar94b]. Dabei bestand und besteht heute noch die Problematik, daß die Gestaltgesetze nicht in operationaler Form vorliegen, sondern vielmehr grundlegende Prinzipien beschreiben, die zu

einer Gruppenbildung führen. Aus diesem Grund wird im folgenden etwas abschwächend auch von den Gestaltprinzipien oder Gruppierungsprinzipien gesprochen.

Zentrale Frage für die automatische Bildanalyse ist, *wie* perzeptive Organisation algorithmisch realisiert werden kann. Perzeptive Organisation wird durch das Gruppieren von primitiven Bildelementen erzeugt. Dabei werden räumliche Beziehungen oder Übereinstimmungen in charakteristischen Eigenschaften zwischen den Bildprimitiva ausgenutzt. Diese Organisationen beschreiben häufig aufgrund der unterschiedlichen Wechselwirkungen zwischen den Elementen Gestalten, die auf eine kausale Ursache, möglicherweise auf Objekte, hindeuten. Aus dieser allgemeinen Anschauung ergeben sich sofort konkrete Fragestellungen, die für eine Realisierung perzeptiver Organisation in einem Bildanalyse-System bearbeitet werden müssen.

- Welche Bildprimitiva sind auf welchen Abstraktionsstufen eines Systems zu benutzen?
- Welche Gestalten entstehen durch das Zusammenwirken von Gruppierungsprinzipien?
- Welche Gestaltprinzipien können eingesetzt werden und müssen entsprechend quantitativ formuliert werden?

In Bezug auf die erste Frage schlagen Sarkar und Boyer in [Sar94b] ein Klassifikationsschema vor, das perzeptive Organisation nach Abstraktionsstufen und der Dimension unterteilt, in der sie stattfindet (siehe Tabelle 3.1). Mit diesem Schema schlagen sie gleichzeitig auch eine einheitliche Terminologie vor, die bereits entstandenen und noch entstehenden Arbeiten zum perzeptiven Gruppieren in der Bildanalyse einzuordnen. Hierzu beschreibt die vertikale Achse der Tabelle die Abstraktionsebenen eines Bildanalyse-Systems, während die horizontale Achse die Dimension beschreibt, in der eine Organisation stattfindet. In jeder Abstraktionsebene benennt die obere Zelle die Strukturen, die durch die Organisation in der entsprechenden Ebene entstehen. Die darunterliegende Zelle führt die hierfür verwendeten Daten bzw. Bildprimitiva auf. Sarkar und Boyer gehen in dieser Klassifikation nicht auf die verwendeten Gruppierungsprinzipien ein, vielmehr liegt der Fokus auf den eingesetzten Bildprimitiva und den daraus entstehenden abstrakteren Bildprimitiva. Damit geben sie bezüglich des Resultats teilweise auch eine Antwort auf die zweite Frage, welche Organisationen entstehen können. Die Frage nach den einzusetzenden Gestaltprinzipien und deren Wechselwirkung zur Erzeugung von Gestalten wird häufig problemabhängig beantwortet und in Abschnitt 3.2.1 mit der Vorstellung einiger Gruppierungsansätze beleuchtet.

Wie bereits zu Beginn dieses Kapitels erwähnt, können auch die in Kapitel 2 dargestellten Verfahren zur Bildsegmentierung mit diesem Schema klassifiziert werden. Demnach gehören diese Verfahren der *Signalebene* und zum Teil der *primitiven Ebene* für 2D-Bilder an. Allerdings wird für Bildsegmentierungen seltener der Begriff des perzeptiven Gruppierens verwendet, gleichwohl essentielle Gruppierungsmechanismen ein-

Aggregat- ebene	Organisation von Strukturen; Anordnung von Polygonen	Gestaltorganisation von Oberflächen- clustern, polyedrische Volumen, etc.	Gestaltorganisation von Flußelement- clustern starrer Körper	Organisation von Fluß; Gruppen
	Strukturen der letzten Ebene	Strukturen der letzten Ebene	Strukturen der letzten Ebene	Strukturen der letzten Ebene
strukturelle Ebene	Ribbons, Ecken, Polygone, Regionen, 3D-Inferenzen, Stränge	nützliche Regionen- kombinationen, Polyeder, zylindrische Volumen, parallele Oberflächen	Organisation von Flußelementen, Polygone, Regionen, etc.	Flußelemente, Bewegungen starrer und flexibler Körper
	Kontur- & Regionenprimitiva	ko-parametrische Ober- flächen, Konturen	optische Linien, Konturelemente	optische Linien, Oberflächen
primitive Ebene	Regionengrenzen, Kontur- segmentierung	ko-parametrische Oberflächen, Detektion von Okklusion	Flußmuster	optische Linien, Flußmuster
	Regionen, Ketten von Kantenpixeln	Oberflächen-Patches, Cluster	optische Flußvektoren	3D-Bewegungs- vektoren
Signal- ebene	Punktewolken, Ketten von Kantenpixeln, Regionen, Textur	Oberflächen-Patches/ Diskontinuitäten, Punktewolken	optischer Fluß, Kantenstreifen	Bewegungs- vektoren, Oberflächen- segmentierung
	Punkte, markante Punkte, Grauwertbild	Tiefenbilder, Punkte im 3D-Raum	bewegte Punkte, Bildsequenzen	dynamische 3D-Szenen
	2D	3D (2$\frac{1}{2}$D)	2D + Zeit	3D + Zeit

Tabelle 3.1.: Klassifikation perceptiver Organisation in der Bildanalyse nach Sarkar und Boyer [Sar94b]. In jeder Ebene benennt die untere Zelle die Eingangsdaten und die obere Zelle die resultierenden Strukturen.

gesetzt werden (Ähnlichkeit, Nähe, gute Fortsetzung). Der überwiegende Teil der Arbeiten, die sich selbst dem Gebiet des perceptiven Gruppierens zuordnen, konzentriert sich auf die Untersuchung statischer Bilder (**2D**) in der *primitiven* und *strukturellen Ebene*. Darüber hinaus sind für die Verarbeitung von Bildsequenzen (**2D + Zeit**) Systeme vorgeschlagen worden, die großteils auf Untersuchungen zum optischen Fluß in Kombination mit Verfahren aus der Verarbeitung statischer Bilder aufbauen. Ansätze, die ein perceptives Gruppieren in den Dimensionen **3D** und **3D + Zeit** beschreiben, sind bisher im Wesentlichen nur durch die vielen Arbeiten zur Tiefenbestimmung oder Stereo-Bildanalyse vorgeschlagen worden. Für einen umfassenden Literaturüberblick sei auf [Sar94b, Boy99a, Boy00] verwiesen.

In der vorliegenden Arbeit werden Verfahren entwickelt, die sich mit der Organisation von Konturen in statischen Bildern (**2D**) beschäftigen. Diese Verfahren lassen sich auf den Abstraktionsstufen der primitiven und strukturellen Ebene (Abschnitte 4.2 und 4.3), sowie auf der Aggregatebene (Abschnitt 4.4) einordnen. In Abschnitt 6.3 wird

zudem ein Verfahren zur Kontursegmentierung vorgeschlagen, das Informationen aus der Verarbeitungshistorie von Bildsequenzen ausnutzt. Dieser Ansatz ist aufgrund seiner aktuellen Konzeption nur eingeschränkt der *primitiven Ebene* der Dimension **2D** + **Zeit** zuzuordnen, da bisher nur nicht bewegte Konturmodelle berücksichtigt werden. Als Ziel ist zukünftig die Realisierung eines räumlich-zeitlichen Gruppierens angedacht, welches zur Identifikation von temporal korrespondierenden Gruppen eingesetzt werden kann. Darüber hinaus wird in Kapitel 5 ein Ansatz entwickelt, der eine Kombination von Konturen und Regionen für das perzeptive Gruppieren realisiert. Damit wird auf die Problematik eingegangen, daß perzeptive Organisation nicht ein Prozeß ist, der sich nur auf die eine oder andere Modalität beschränkt, sondern vielmehr ein ganzheitliches Konzept darstellen sollte.

Im folgenden werden verschiedene Verfahren für die Dimension **2D** vorgestellt, wobei im Hinblick auf ein räumlich-zeitliches Gruppieren auch der Bereich **2D** + **Zeit** gestreift wird. Bedingt durch die Ausrichtung der vorliegenden Arbeit auf ein konturbasiertes Gruppieren, werden entsprechend nur konturbasierte Arbeiten beleuchtet. Arbeiten, die sich mit der Integration von Regionen und Konturen beschäftigen, werden in Abschnitt 5.1 vorgestellt.

3.2.1. Verfahren in der Dimension 2D

Die Signalebene des Klassifikationsschema nach Sarkar und Boyer beschäftigt sich allgemein mit der Organisation der ikonischen Bilddaten. Da in der vorliegenden Arbeit diese Organisationsform bereits durch die in Kapitel 2 vorgestellten Bildsegmentierungen (Konturdetektion, Segmentierung homogener Bereiche) realisiert wird, erfährt diese Ebene hier keine weitere Betrachtung. Weiterhin ist die Zuordnung der verschiedenen Verfahren in die Ebenen nicht immer eindeutig. Einige Verfahren lassen sich in mehrere Ebenen parallel einordnen, da sie in *einem* System unterschiedlich abstrakte Strukturierungen realisieren. Die Einteilung ist daher bei diesen Ansätzen nach dem Gewicht der untersuchten Gruppierungsleistung erfolgt.

Primitive Ebene

In der primitiven Ebene werden Verfahren eingeordnet, die eine Organisation von Bildelementen aufgrund eines linearen Charakters bzw. einer eindimensionalen Ausdehnung realisieren.

Einige dieser Arbeiten basieren auf der Idee einer *Saliency Map*, die erstmals von Sha'ashua und Ullmann [Sha88] vorgeschlagen wurde. In diesen Karten wird für jedes Bildprimitivum ein Maß für seine Auffälligkeit (*Saliency*) berechnet, mit der es zu globaleren Strukturen im Bild beiträgt. Zur Berechnung dieses Wertes werden überwiegend Mechanismen perzeptiven Gruppierens (Nähe, gute Fortsetzung, Ähnlichkeit) operationalisiert. Die damit entstehenden Gruppierungsverfahren werden dann auf ein Netzwerk von Elementen mit lokalen Nachbarschaftbeziehungen angewendet. In [Sha88] selbst

wird so beispielsweise Kantenelementen von langen glatten Konturen eine hohe Saliency zugewiesen. Hierzu wird ein Optimierungsverfahren eingesetzt, das auf Methoden der dynamischen Programmierung beruht und durch iterative Anwendung lokaler Operatoren entsprechend auffällige Kantenelemente für Konturen zunehmender Länge ermitteln kann. Die Detektion von Kontursegmenten kann anschließend aufgrund der Saliency der Kantenelemente und einer zusätzlich berechneten bevorzugten Konturrichtung erfolgen. Eine Erweiterung und Generalisierung dieses Ansatzes wird z.B. in [Alq98] im Rahmen eines hierarchischen Gruppierungssystems vorgeschlagen. Auf Basis eines generischen *Saliency Network* werden Kantenelemente zu Kontursegmenten bzw. Kontursegmente zu Ketten von Segmenten gruppiert, wobei für die Definition des Optimalitätskriteriums Konzepte aus dem Anwendungsfeld aktiver Konturen verwendet werden. Ein guter Überblick und Vergleich verschiedener Methoden zur Bestimmung der *Saliency* von Bildprimitiva findet sich in [Wil98]. Eine neuere Arbeit beschreibt in [Lin00a, Lin00b] eine probabilistische Interpretation von *Saliency*.

In [Cre99, Cre00] wird ein probabilistischer Ansatz zur Gruppierung von Geradenstücken und Kreisbögen in Ketten kollinearere Konturmodelle vorgestellt. Dieser Ansatz ist ganz wesentlich durch das Prinzip des non-accidentalness motiviert. Zunächst werden alle Paare von Konturmodellen gruppiert, deren Mittelpunktabstand kleiner als ihre mittlere Länge ist. Anschließend werden die Gruppierungen verworfen, die heuristisch gewählte Schwellwerte bezüglich Orientierungsdifferenz, Kontrast oder Distanzmerkmale überschreiten. Das Verhältnis aus den verbleibenden Gruppierungen zu allen möglichen Paaren beschreibt dann die a priori Wahrscheinlichkeit für das Auftreten einer Gruppierung aufgrund einer kausalen Ursache (non-accidentalness). Für eine weitere probabilistische Untersuchung der verbleibenden Paare wird ein Maß eingesetzt, das die Abweichung von einer optimalen Verbindung zwischen zwei Konturmodellen im Sinne des Gestaltprinzips der guten Fortsetzung mißt. Aus Simulationen von zufällig generierten Konturmodellen wird dann die Wahrscheinlichkeitsverteilung für kollineare Gruppierungen in Abhängigkeit von diesem Maß geschätzt (accidentalness). Mit dieser Wahrscheinlichkeit und der a priori Wahrscheinlichkeit für eine kausale Ursache, wird dann die a posteriori Wahrscheinlichkeit für das kausale Auftreten der einzelnen kollinearen Gruppierung durch Anwendung der Bayes-Formel berechnet. Ein empirisch ermittelter Schwellwert bezüglich dieser Wahrscheinlichkeit schränkt die Akzeptanz von Gruppierungen weiter ein. Zur Ermittlung von Ketten kollinearere Gruppierungen werden nun Paare von Gruppen gebildet, deren Verbundwahrscheinlichkeit den Akzeptanzschwellwert überschreiten. Eine iterative Anwendung des Verfahrens, bei der die zuletzt expandierten Gruppierungen als neue Konturmodelle berücksichtigt werden, ermöglicht sukzessive Überbrückungen größerer Fragmentierungen und führt schließlich zu maximal expandierten kollinearen Gruppierungen.

In [Ami98] stellen Amir und Lindenbaum einen generischen Gruppierungsansatz vor, der prinzipiell auf allen Ebenen des Klassifikationsschemas eingesetzt werden kann. In einem Graphen wird zunächst jedes Bildprimitiva durch einen Knoten repräsentiert (in [Ami98] z. B. Kantenelemente). Die Nachbarschaft der Knoten wird durch eine

für das Gruppierungsproblem adäquate Formulierung festgelegt (in [Ami98] K nächste Kantenelemente eines Kantenelementes). Hinweise auf gruppierende Eigenschaften zwischen zwei oder mehrelementigen Teilmengen von Knoten werden durch Funktionen auf binäre Relationen im Graphen abgebildet. Diese Funktionen beurteilen dabei die Konsistenz einer Gruppierungsannahme (z.B. Glattheit bezüglich einer zugrundeliegenden Kontur). Bei mehrelementigen Teilmengen wird eine *cue enhancement* Methode eingesetzt, die auf statistischen Verfahren beruht und letztendlich die höhere Ordnung wieder auf binäre Relationen abbildet. Die Detektion von Gruppen wird abschließend mit einem heuristischen Verfahren zur Ermittlung von Cliques mit maximaler Likelihood realisiert. Auf Basis dieses Ansatzes präsentieren Amir und Lindenbaum in [Ami99] eine Operationalisierung des Gestaltprinzips von Figur-Hintergrund-Trennung zur Isolierung von Kantenelementen in stark verrauschten Gradientenbetrags-Bildern, die die Kontur einer Struktur beschreiben. Dieser Ansatz ist ebenfalls generisch angelegt, und ermöglicht auch die Isolierung komplexerer Strukturen. Die entscheidende Erweiterung wird durch einen *Bootstrap* Algorithmus erreicht, der die Nachbarschaft zwischen Strukturen analysiert und gegebenenfalls Kanten eliminiert oder auch hinzufügt.

Strukturelle Ebene

In der strukturellen Ebene werden Bildprimitiva organisiert, die Strukturen mit einer zweidimensionalen Ausdehnung beschreiben, und damit über einen rein linearen Gruppierungscharakter hinausgehen. Die konstituierenden Bildprimitiva stehen beispielsweise in einer symmetrischen Beziehung zueinander oder formen geschlossene Figuren.

In [YJ92] wird ein Ansatz vorgestellt, der Geradenstücke in spiegelsymmetrische Strukturen (*Ribbons*) gruppiert. Zunächst werden für alle Paare von Geradenstücken geometrische Merkmale (Distanzen, Winkeldifferenz, etc.) und symbolische Attribute berechnet. Symbolische Attribute beschreiben beispielsweise, ob sich die Projektionsbereiche von zwei Geradenstücken auf ihrer Symmetrieachse überlappen. Anhand dieser Merkmale erfolgt eine Auswahl relevanter Gruppierungen. Dabei ist die Auswahl durch die entsprechend zu lösende Gruppierungsaufgabe vorgegeben, und wird mit Schwellwerten und logischen Relationen realisiert. Die verbleibenden Symmetrien werden anschließend zu linear angeordneten Paaren gruppiert. Auch hier erfolgt die Auswahl über geometrische Merkmale und symbolische Attribute, nun jedoch bezüglich der entsprechenden Symmetrieachsen. Abschließend werden diese Paare von Symmetrien rekursiv zu größeren Strukturen zusammengefaßt. Dies geschieht durch die Auswertung von Teilmengenbeziehungen zwischen den Paaren von Symmetrien.

Ein asynchroner Gruppierungsansatz zur Detektion von geschlossenen Figuren wird in [JD97] vorgeschlagen. Konturbeschreibungen (Geradenstücke, Kreisbögen) werden zunächst sequentiell für einen Gruppierungsprozeß herangezogen. Die Reihenfolge ergibt sich aus der Saliency, die für jedes Konturmodell aus dessen Länge und der Krümmung bestimmt wird. Jedes Konturmodell wird dann auf Kreuzungspunkte und gruppierende Eigenschaften mit bereits betrachteten Konturmodellen überprüft. Hierzu werden quan-

titative Formulierungen der Gestaltprinzipien Nähe, gute Fortsetzung und Parallelität eingesetzt. Die Konturmodelle, die Gruppierungen mit anderen Konturapproximationen bilden und damit zu relevanten Strukturen im Bild beitragen können, werden in einen Strukturgraphen eingefügt. In diesem Graphen beschreiben die Knoten die End- und Kreuzungspunkte der Konturmodelle, während die Kanten die Modelle selber und deren strukturellen Relationen untereinander repräsentieren. Zur Detektion von geschlossenen Figuren werden nun die kürzesten zyklischen Pfade in dem Graphen gesucht und auf Basis eines Alphabets vorgegebener Formen auf Gültigkeit überprüft.

In [Kri96] wird eine konturbasierte Detektion von Gebäudeumrissen in Luftbildern vorgestellt. Zur Lösung dieses Problems definieren die Geradenstücke einer Kontursegmentierung die Stellen eines Markov Random Fields¹. Das Label jeder Stelle dient als kontinuierliches Selektionskriterium für Geradenstücke, die Projektionen von Gebäudeumrissen beschreiben. Die Nachbarschaft zwischen den Stellen wird zum einen über die Distanz und die Winkeldifferenz zwischen den Geradenstücken, aber auch durch die aktuellen Labelwerte im Analyseprozeß definiert. Damit können sich die Nachbarschaften zwischen den Stellen während der Analyse verändern. Zur Kodierung charakteristischer geometrischer Eigenschaften von Gebäudeumrissen werden Potentialfunktionen definiert, die lange, nahe und parallel oder orthogonal angeordnete Geradenstücke unterstützen. Als zusätzlicher Term wird in der Energiefunktion die Entropie der Labelwerte berücksichtigt. Dies unterstützt eine fossierte Entscheidung für die extremen Labelwerte 0 oder 1 an den einzelnen Stellen. Zur Minimierung der Energie wird ein Gradientenabstiegsverfahren eingesetzt. Abschließend können Geradenstücke durch ihr Label (= 1) selektiert werden. Um die Selektion von Schattenkonturen zu vermeiden, werden nur Konturen gewählt, die helle Bereiche eingrenzen. Darüber hinaus werden unvollständig beschriebene Gebäudeumrisse durch den Einsatz von aktiven Konturen vervollständigt.

Eine weitere Anwendung von MRFs zum Gruppieren von Geradenstücken und Kreisbögen findet sich in [Mur93]. Auch hier bilden die Konturmodelle die Stellen des MRFs, dessen Zielkonfiguration die Gruppierung von Geradenstücken realisieren soll. Hierzu werden Potentialfunktionen definiert, die Gruppierungen von Geradenstücken aufgrund von Parallelität oder Nähe bewerten. Anschließend werden im Energieminimierungsprozeß entsprechende Gruppen über die Zuordnung identischer Indizes gebildet. Diese Gruppen werden in einem weiteren MRF als Knoten aufgefaßt, das durch Definition einer adäquaten Nachbarschaft und Potentialfunktionen eine primitive Modellierung von menschlichen Körpern realisieren soll. Die Potentialfunktionen werden dabei zur Modellierung von guter Fortsetzung eingesetzt. Im Minimierungsprozeß werden Teilstrukturen durch Zuordnung identischer Indizes gruppiert. Als Anwendung wird die Detektion von Passagieren in Bahnsteigszenen präsentiert.

In [Zhu99] wird ein theoretischer Formalismus vorgeschlagen, um natürliche geschlossene Formen durch perzeptive Organisation von Konturfragmenten in Markov Random Fields zu repräsentieren. Hierzu werden generische probabilistische Formmodelle aus

¹ Die Theorie zu Markov Random Fields wird in Abschnitt 6.1 ausführlich dargestellt.

einer Stichprobe typischer domänenabhängiger Objektformen mittels eines Maximum Entropie Ansatzes gelernt. Die gelernten Formen werden durch Gibbs Verteilungen modelliert, die die Wahrscheinlichkeitsverteilungen für entsprechende MRFs repräsentieren. Die Stellen dieser MRFs sind durch die Konturfragmente der geschlossenen Formen definiert, wobei deren Positionen als Label eingesetzt werden. Das zugeordnete Nachbarschaftssystem ergibt sich aus räumlichen Nähe- und Symmetriebeziehungen innerhalb der gesamten geschlossenen Kontur. Für diese MRFs werden dann die Potentialfunktionen für paarweise Relationen aus Merkmalshistogrammen gelernt, so daß die gelernten Formen bei einer stochastischen Simulation für ein MRF wiedergegeben werden. Die verwendeten Merkmale definieren quantitative Formulierungen der Gestaltprinzipien Nähe, gute Fortsetzung und Symmetrie. Für das Lernen, als auch für die Simulation, werden stochastische Methoden wie MCMC (*Markov Chain Monte Carlo*) eingesetzt. Damit steht abschließend ein Repräsentationsformalismus zur Verfügung, der aus gelernten natürlichen geschlossenen Formen durch stochastische Simulation ähnliche natürliche Formen erzeugen bzw. repräsentieren kann.

Die Problematik der ungenauen Definition der Gestaltgesetze wird in [Ral99] durch den Einsatz von Fuzzy-Methoden angegangen. Auf Basis initialer Geradenstücke werden Fuzzymengen definiert, die die Gruppierungswirkung aufgrund von Nähe, guter Fortsetzung und Symmetrie abbilden. Hierzu werden ebenfalls Distanzmerkmale und Winkeldifferenzen eingesetzt. Der Ansatz gruppiert zunächst kollineare Geradenstücke in sogenannte *Fuzzy Segmente*, die im Wesentlichen eine ungenauere Beschreibung der zugrundeliegenden realen Kontur abbilden sollen. Nach einem Verifikationsschritt, in dem gruppierte Geradenstücke mit großem Überlappungsbereich eliminiert werden, erfolgt die Detektion von Ecken und Symmetriebeziehungen zwischen den Fuzzy Segmenten. Dies geschieht durch Auswertung entsprechend winkel- und distanzabhängiger Fuzzymengen. Abschließend erfolgt unter Ausnutzung der Fuzzy-Segmente und der Ecken die Detektion geschlossener konvexer Strukturen. Hierzu wird nach alternierenden Folgen von Fuzzy-Segmenten und Ecken zwischen diesen Segmenten gesucht.

Ein hierarchischer Ansatz zur Gruppierung von Geradenstücken wird in [Moh89] vorgeschlagen. Zunächst werden initiale Geradenstücke in kollineare Gruppen organisiert, die dann weiter zur Generierung von Parallelitätshypothesen verwendet werden. Auf Basis dieser Gruppierungshypothesen werden U-Konturen gebildet, wenn eine Parallelitätshypothese durch die Nachbarschaft in den Endpunkten mit einem dritten Geradenstück das Erscheinungsbild eines U hervorruft. Paare von geeigneten U-Konturen können abschließend zur Hypothesisierung eines Rechtecks führen, wenn sie eine identische Parallelität enthalten. Für diesen Generierungsprozeß werden die Gruppierungsprinzipien für Nähe, gute Fortsetzung und Parallelität eingesetzt und durch Schwellwerte für entsprechende Merkmale operationalisiert. Zur Vermeidung redundanter und falscher Gruppierungshypothesen wird nach der Generierung eine Auswahl der korrekten Hypothesen vorgenommen. Diese Auswahl ist durch ein *Constraint Satisfaction Network* realisiert, das auf Basis eines neuronalen Hopfield Netzwerkes unterstützende und konkurrierende Beziehungen zwischen den Hypothesen auswertet und damit die Auswahl eines Satzes

konsistenter Hypothesen ermöglicht. In einer Verallgemeinerung dieses Ansatzes werden von Mohan und Nevatia auch allgemeinere Konturformen zugelassen [Moh92]. Hierzu wird neben den erwähnten Gruppierungsprinzipien noch das Prinzip der Symmetrie für Polygonzüge modelliert. Entsprechend signifikante Strukturen werden wiederum durch ein Constraint Satisfaction Network ermittelt. In [Lin98] präsentieren Lin und Nevatia als Anwendung die Detektion von Gebäudeumrissen in Luftbildaufnahmen.

Ein weiterer hierarchischer Ansatz wird von Sarkar und Boyer in [Sar94b] vorgestellt. Als Ausgangsbasis werden hier ebenfalls initial detektierte Geradenstücke und Kreisbögen verwendet. Diese werden in einer nächsthöheren Ebene zu Ribbons, geschlossenen Figuren, Parallelitäten und Kreuzungen organisiert. Darauf baut noch eine Ebene zur Organisation von beispielweise parallelen Geschlossenheiten auf. Die generierten Strukturen werden in jeder Ebene in Abhängigkeit vom Abstraktionsgrad mit verschiedenen Merkmalen beschrieben und durch geometrische Primitiva (Konturmodelle, Symmetrieachsen oder Schwerpunkte) lokalisiert. Die Generierung der Strukturen erfolgt dabei für jede Ebene getrennt und gliedert sich in zwei Phasen. In der preattentiven Phase wird für jedes der modellierten Gestaltgesetze (Nähe, gute Fortsetzung und Ähnlichkeit) ein sogenannter Gestalt-Graph aufgebaut. Darin definieren alle Strukturen der jeweils darunterliegenden Ebene die Menge der Knoten. Mit Hilfe von *Voting Methoden* werden zwischen den Knoten Verbindungen erzeugt, wenn die entsprechenden Strukturen das zugehörige Gestaltprinzip erfüllen. Durch verschiedene Graphalgorithmen und einfache Verknüpfungen zwischen den Graphen werden Gruppierungshypothesen aufgestellt und in der zweiten, attentiven Phase weiter analysiert. Für diese Phase wird ein *Perceptual Inference Network* (PIN) entwickelt, in dem auf Basis von Bayes-Netzwerken Wissen über strukturelle Abhängigkeiten von Gruppierungsbestandteilen kodiert ist. Die Generierung der einzelnen Gruppierungen führt im PIN zu Instantiierungen vordefinierter *Composite Nodes*, die primitive geometrische Strukturen repräsentieren (z.B. Rechtecke) und denen Information über ihren strukturellen Aufbau zugeordnet ist. In jedem Composite Node wird mittels einer bedingten Wahrscheinlichkeit die Sicherheit seiner Instantiierung als Strukturhypothese ermittelt. Hierzu werden Informationen über die konstituierenden Elemente der Struktur durch ein *Message Passing*-System integriert. Fehlende konstituierende Komponenten einer Struktur können so auch inferiert werden. Die Sicherheit der Composite Nodes wird mit der Integration zusätzlicher Gruppierungen sukzessive verstärkt. Konkurrierende Strukturhypothesen an nahezu identischen Positionen im Bild (Rechteck vs. Kreis) werden durch die Integration geeigneter Bildverarbeitungsmodule aufgelöst.

Aggregatebene

In der Aggregatebene werden Bildprimitiva organisiert, die beispielsweise die Projektionen von Objektoberflächen durch geschlossene Figuren beschreiben. Die resultierenden Gruppierungen beschreiben prinzipiell ähnliche Strukturen, wie in den darunterliegenden Ebenen, nur auf einem abstrakteren Niveau und mit einem globaleren Einflußbereich.

Eine der wenigen Arbeiten zur Realisierung von perzeptiver Organisation in der Aggregatebene wird von Rosin und West in [Ros91] vorgestellt. Bei der Projektion von zylindrischen Objekten oder Strukturen aus der realen Welt entstehen in der Bildebene normalerweise ellipsenförmige Abbildungen dieser realen Ereignisse. Mit dieser Motivation schlagen Rosin und West die Gruppierung von linear und parallel angeordneten Ellipsenbögen vor, die dann Rückschlüsse auf die 3D-Struktur und Lage von zylindrischen Objekten erlauben. Zunächst werden in einem ersten Schritt parallele Ellipsenbögen aufgrund der Orientierung ihrer Halbachsen und dem Verhältnis der Längen der Halbachsen mit Hilfe einer Hough-Transformation in Gruppen zusammengefaßt. Das Längenverhältnis der Halbachsen wird verwendet, da es Hinweise auf identische Neigungswinkel von kreisförmigen Rotationsflächen in der realen Welt beinhaltet. Anschließend werden die Gruppen wiederum mit einer Hough-Transformation weiter unterteilt. Die Untergruppen enthalten danach Ellipsenbögen, deren Mittelpunkte auf einer Linie liegen, die senkrecht zu den großen Halbachsen der Ellipsenbögen verläuft. Zur Bewertung dieser Untergruppen wird abschließend nach Kantenelementen und Geradenstücken gesucht, die über die Verbindungslinie der Mittelpunkte symmetrisch zueinander lokalisiert sind. Mit den symmetrischen Geradenstücken wird schließlich die Projektion der äußeren Kontur der Objekte beschrieben.

3.2.2. Verfahren für Bildsequenzen

In der perzeptiven Organisation einer Szene hat auch Bewegung eine stark organisierende Wirkung. Dieses wurde von den Gestaltpsychologen durch das Gesetz des *gemeinsamen Schicksals* zum Ausdruck gebracht und wird in dem Klassifikationsschema von Sarkar und Boyer in den Dimensionen **2D + Zeit** und **3D + Zeit** berücksichtigt. Bei der Wahrnehmung von gemeinsam bewegten Strukturen werden diese häufig als Einheit empfunden und heben sich damit von ihrer Umgebung ab. Für die automatische Bildanalyse ergibt sich hierbei die Schwierigkeit, einzelne Strukturen in zeitlich aufeinanderfolgenden Bildern als korrespondierend zu identifizieren. Für das sogenannte *räumlich-zeitliche* Gruppieren lassen sich prinzipiell drei Paradigmen aufstellen (vergleiche [Boy00]):

sequentielles Gruppieren: Die Bilder einer Sequenz werden zunächst einzeln segmentiert und räumlich organisiert. Anschließend werden die Strukturen von Bild zu Bild als korrespondierend identifiziert und zeitlich gruppiert. Das im Rahmen dieser Arbeit vorgeschlagene Verfahren aus Abschnitt 6.3 basiert auf diesem Konzept, allerdings nur in Bezug auf nicht bewegte Strukturen.

interaktives Gruppieren: Dieses Konzept ist durch eine Art *Hypothesisierung* und *Verifikation* motiviert. Das zeitliche Gruppieren von Bildprimitiva über eine Folge von Bildern kann eingesetzt werden, um räumliche Gruppierungen zu korrigieren bzw. anzupassen, die zu einem früheren Zeitpunkt generiert wurden. Können beispielsweise Konturmodelle über die Zeit als korrespondierend gruppiert werden, so würde

durch diesen Zugewinn an Strukturierungsinformation die Korrektur von anfangs fehlerhaften oder fehlenden räumlichen Gruppierungen angestrebt.

multidimensionales Gruppieren: In einem multidimensionalen Konzept wird Organisation in einem drei- oder vierdimensionalen Raum realisiert, wobei *eine* Dimension der Zeitskala entspricht. Aufgrund der parallel zur Verfügung stehenden Informationen birgt dieses Konzept das größte Potential ein raum-zeitliches Gruppieren adäquat zu realisieren. Jedoch führt die Erweiterung der Dimensionalität auch zu einer deutlichen Steigerung der Komplexität der Problemstellung. In [Bom97, Pos97] wird ein Verfahren zur Identifikation von korrespondierenden Regionen und gemeinsam bewegten Regionen vorgestellt. Das Verfahren kann diesem Paradigma zugeordnet werden, auch wenn der Gedanke eines voll integrierten raum-zeitlichen Gruppierens in der Realisierung nicht gegeben ist.

Im folgenden werden einige konturbasierte Verfahren beschrieben, die eine Korrespondenzbestimmung von Konturen über die Zeit realisieren.

Eine Erweiterung von aktiven Konturen zur Verfolgung von Konturbewegungen wird von Denzler in [Den97] vorgestellt. Hierbei wird in einem ersten Schritt die Energiefunktion der aktiven Kontur um eine temporale Komponente erweitert. Zusätzlich wird eine Bewegungsvorhersage der Konturpunkte über einen Kalman-Filter im Energieterm der aktiven Kontur berücksichtigt und mit dem parallel ermittelten Schätzfehler gewichtet. Anwendung findet dieses Verfahren beispielsweise in [Den96] zur konturbasierten Objektverfolgung. In [Par97, Par98] wird, ebenfalls auf Basis einer erweiterten Energiefunktion von aktiven Konturen, ein Verfahren zur Schätzung des optischen Flusses vorgestellt. Hierzu werden korrespondierende Konturpunkte in aufeinanderfolgenden Bildern über eine Minimierung bezüglich der Krümmungsänderung in einem Konturpunkt bestimmt. Insgesamt wird dann eine robuste Verfolgung aktiver Konturen über eine Bildfolge realisiert. Diese beiden Verfahren sind dem Paradigma des sequentiellen Gruppierens zuzuordnen.

In [Cha97] werden zunächst in jedem Bild unabhängig voneinander Geradenstücke detektiert und über eine normale Parametrisierung (Position, Orientierung, Länge) hinaus durch entsprechende statistische Merkmale charakterisiert. Auf Basis von statistischen Abstandsmaßen werden anschließend Geradenstücke sowohl in der Bildfolge als korrespondierend, als auch innerhalb eines Bildes aufgrund von Nähe, Parallelität räumlich gruppiert. Hierzu wird ein kooperatives *Relaxation Labeling* eingesetzt, das unter der Bedingung ähnlicher Bewegung räumliche Gruppierungen und zeitliche Zuordnungen von Geradenstücken bewertet und entsprechend verstärkt oder abschwächt bzw. eliminiert. Dieses Verfahren läßt sich ansatzweise in das Paradigma des interaktiven Gruppierens einordnen, wobei der eigentliche Gedanke einer Verifikation über einen längeren Zeitraum nicht realisiert wurde. Jedoch merken Chang und Aggarwal an, daß neben der Erweiterung der verwendeten Primitiva auch eine Langzeit-Gruppierung und Verifikation angestrebt wird.

In [Sar94a] stellt Sarkar einen Ansatz vor, der konturbasierte Strukturen in Bildern zur Auflösung der zeitlichen Korrespondenz ausnutzt. In jedem Bild einer Bildsequenz werden konturbasierte Bildprimitiva (Linien und Kreisbögen) in strukturelle Einheiten organisiert und mit einem statistischen Ansatz bewertet. Hierzu wird das bereits beschriebene Verfahren von attentiven und preattentiven Gruppierungsphasen eingesetzt [Sar94b] (vergleiche Seite 47). Anschließend werden alle Strukturen aus Bildern einer Sequenz in ein *composite image* projiziert und durch Schwerpunkte und gegebenenfalls durch Geradenstücke für Symmetrieachsen repräsentiert. Durch den Einsatz von *Voting*-Methoden werden paarweise Relationen zwischen den Strukturen aufgrund von Nähe, Ähnlichkeit und Parallelität bestimmt, und in entsprechenden Gestalt-Graphen repräsentiert. In diesen Graphen beschreiben die Kanten Beziehungen zwischen Strukturen aus unterschiedlichen Bildern. Durch logische Operationen werden diese Graphen in *einem* Graphen zusammengefaßt, dessen Pfade mögliche Trajektorien der Strukturen über die Bildfolge beschreiben. Unter Berücksichtigung eines Ähnlichkeitsmaßes führt eine Suche nach den kürzesten Pfaden zu Hypothesen über Trajektorien einzelner Strukturen, die im Hinblick auf ihre Kohärenz bewertet werden. Abschließend findet eine strukturelle Organisation des gesamten Bewegungsverhaltens im Bild statt. Hierzu werden die gefundenen Trajektorien approximiert und erneut im zeitlichen Raum organisiert.

Nach dieser Übersicht verschiedener Verfahren zum perzeptiven Gruppieren in der Bildanalyse, wird in dem folgenden Kapitel der im Rahmen der vorliegenden Arbeit entwickelte konturbasierte hierarchische Ansatz zum perzeptiven Gruppieren vorgestellt. In Abschnitt 6.3 wird ein Konzept für ein Verfahren vorgeschlagen, das eine initiale Kontursegmentierung in Bildfolgen realisiert und so eine zeitliche Korrespondenz von Konturmodellen ermöglicht.

4. Konturbasiertes hierarchisches Gruppieren

In diesem Kapitel wird ein Ansatz zur Modellierung von perzeptiven Gruppierungsleistungen auf Basis von Konturmodellen vorgestellt. Dieser Ansatz ist zusammen mit Posch und Maßmann in [Maß95b, Maß95a, Maß96, Maß97, Ack97] vorgeschlagen worden und verfolgt eine Quantifizierung der Gestaltprinzipien für Konturmodelle mit folgenden Zielsetzungen: Zum einen soll eine Detektion von inhärenten Strukturen im Bild auf Basis der Konturinformation erfolgen. Damit wird strukturelle Information für weitere Analyseschritte bereitgestellt, beispielsweise für eine Objekterkennung. Darüber hinaus sollen die bereits in Kapitel 2.1.4 diskutierten Fragmentierungen von Objektkonturen aufgelöst werden, die aufgrund lokaler Bildstörungen oder Verdeckungen entstehen.

Der vorgestellte Ansatz gliedert sich in der Klassifikation perzeptiver Organisation nach Sarkar und Boyer [Sar94b] in die strukturelle Ebene für statische 2D-Bilder ein. Dem System liegt eine Gruppierungshierarchie zugrunde, die ausgehend von den Ergebnissen einer Konturapproximation eine strukturelle Organisation der Konturmodelle zunehmenden Komplexitätsgrades modelliert. Dabei beschreibt diese Hierarchie im wesentlichen räumliche Relationen zwischen Konturgruppierungen, die in Kapitel 4.1 näher erläutert werden. Die Generierung der einzelnen Konturgruppierungen beruht hauptsächlich auf der Auswertung lokaler Nachbarschaftsbeziehungen zwischen den Konturmodellen oder bereits vorhandenen Gruppierungen. Dabei wird mit der Generierung zunehmend komplexerer Gruppen die abstrakte Begrifflichkeit von Lokalität deutlich. So werden auch Gruppierungen selbst wieder in komplexeren Strukturen organisiert. In den Abschnitten 4.2 bis 4.4 wird auf die verschiedenen Teilschritte des Generierungsprozesses ausführlich eingegangen.

Mit dem Prinzip des *non-accidentalness* und der Erkenntnis, daß *das Ganze verschieden ist von der Summe seiner Teile*, wird die Entscheidung, welche Gruppierungsannahmen eine konsistente Interpretation des Bildinhaltes darstellen, erst in einem Bewertungsverfahren im Anschluß an die Generierung entschieden (siehe Kapitel 6.2). Bis zu dieser Bewertung werden im folgenden die Gruppierungen auch als *Gruppierungshypothesen* bezeichnet.

4.1. Gruppierungshierarchie

In dem Ansatz nach [Maß95b, Maß95a, Maß96, Maß97, Ack97] wird eine Gruppierungshierarchie entwickelt, die verschiedene konturbasierte Gruppierungstypen definiert. Diese Gruppierungstypen sind durch die Gestaltprinzipien für Nähe, Ähnlichkeit, gute Fort-

4. Konturbasiertes hierarchisches Gruppieren

setzung, Symmetrie und Geschlossenheit motiviert. Mit einem Gruppierungsprozeß wird das Ziel verfolgt, Konturapproximationen entsprechend eines Gruppierungsprinzips zu Einheiten oder Gestalten zusammenzufassen. Diese Einheiten stellen dann auf einer abstrakteren Betrachtungsebene wiederum elementare Bildprimitiva für höhere Interpretationsaufgaben dar, beispielsweise für eine Objekterkennung.

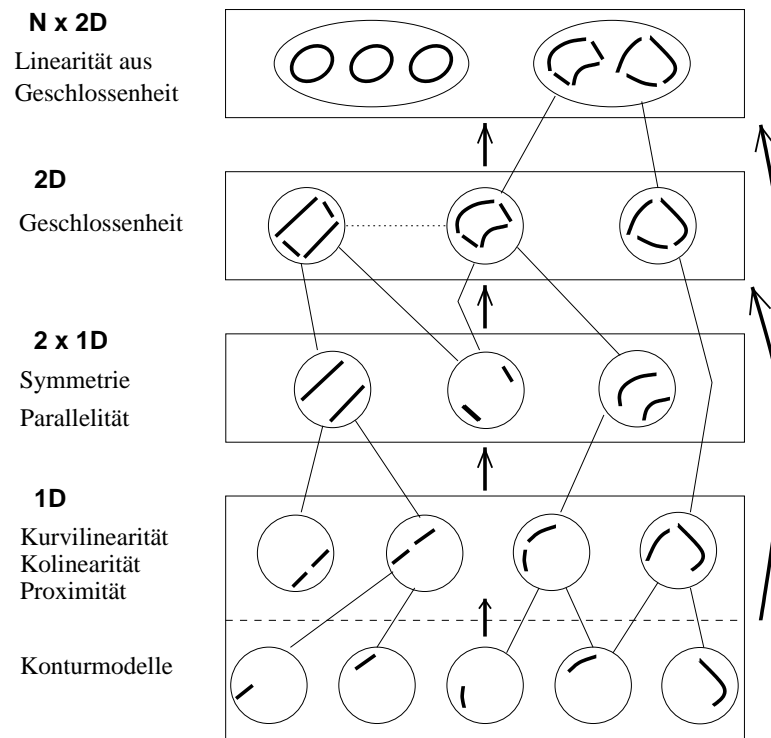


Abbildung 4.1.: Gruppierungshierarchie zur Modellierung konturbasierter Gruppierungsphänomene. Die Kreise stellen die einzelnen Gruppierungshypothesen dar. Dünne Linien zwischen den Ebenen markieren Bestandteilsbeziehungen von Gruppierungen, während dicke Pfeile den Generierungsprozeß andeuten. Gepunktete Linien zwischen Gruppierungen in einer Ebene beschreiben eine konkurrierende Beziehung im Sinne einer inkonsistenten Interpretation (siehe Kapitel 6.2).

Die Basis der Gruppierungshierarchie bilden die Konturmodelle aus der Bildsegmentierung (siehe Abschnitt 2.1.3). Diese Geradenstücke und Ellipsenbögen werden in der nächst höheren Ebene (**1D**) unter dem Gesichtspunkt der guten Fortsetzung und der Nähe zu *Kollinearitäten* und *Kurvilinearitäten* zusammengefaßt. Die Namensgebung dieser Ebene resultiert aus der räumlich eindimensionalen Ausdehnung und dem damit verbundenen linearen Charakter dieser Gruppierungsprinzipien. In dieser Ebene stellt beispielsweise die Gruppierung von zwei oder mehr Geradenstücken mit ähnlicher Orientierung eine neue linienhafte Interpretation der zugrundeliegenden Kantenelemente dar.

Eine erneute Approximation der zusammengefaßten Kantenelemente beschreibt dann eine abstraktere Struktur als die ursprüngliche Beschreibung durch mehrere einzelne Teilstücke. Diese neue abstraktere Struktur hebt sich jedoch nicht wesentlich von den initialen Teilstücken ab. Auch sind Objektkonturen häufig schon nach der Kontursegmentierung vollständig beschrieben und können als *primitive Gruppierungen* betrachtet werden. Aus diesen Gründen wird eine strikte Trennung dieser unteren Ebenen vermieden und die initialen Konturmodelle werden strukturell auf eine Ebene mit gruppierten Geradenstücken gestellt. Diese gleichwertige Betrachtung ist in Abbildung 4.1 durch die Verbindung von Konturmodell-Ebene und **1D**-Ebene angedeutet. Eine analoge Betrachtungsweise gilt für die Gruppierung von Ellipsenbögen zu Kurvilinearitäten. Im weiteren werden kollineare und kurvilineare Gruppierungen gemeinsam mit den konstituierenden Konturmodellen unter dem Begriff der *linearen Gruppen* zusammengefaßt.

Ein weiterer Gruppierungstyp innerhalb der **1D**-Ebene ist die Gruppierung von linearen Gruppen aufgrund des Prinzips der Nähe. Diese *Proximitäten* beschreiben Paare von linearen Gruppen, die *nur* aufgrund des geometrischen Abstandes zwischen ihren Endpunkten hypothetisiert werden. Um eine Redundanz von Proximitäten zu Kol- und Kurvilinearitäten zu vermeiden, werden Proximitäten nur zwischen linearen Gruppen generiert, die über eine Endpunktnähe hinaus dem Prinzip der guten Fortsetzung nicht genügen. Allerdings ist die Bedeutung der Proximitäten als strukturbeschreibende Elemente des Bildinhaltes eher untergeordnet anzusehen. Vielmehr sind sie wichtige Elemente zur Generierung stärker strukturierender Gruppierungen der **2D**-Ebene.

In der **2×1D**-Ebene werden lineare Gruppen aufgrund von Symmetrieeigenschaften gruppiert. Die Bezeichnung ist motiviert durch den Umstand, daß zwei lineare Gruppen zwar durch Symmetrie eine zweidimensionale Ausdehnung implizieren, aber keinen geschlossenen Bereich umschließen. Zudem ist bei einer Symmetrie die Ausdehnung orthogonal zur Symmetrieachse auch von linearem Charakter. Die Modellierung der Hierarchie ist auf eine Beschreibung von allgemeinen symmetrischen Strukturen angelegt. In der *baufix*[®] Domäne sind symmetrische Strukturen jedoch vornehmlich in Form von parallelen Geradenstücken zu beobachten. Aus diesem Grund wird hier nur der Sonderfall der *Parallelität* von Geradenstücken berücksichtigt. Für eine Interpretation des Bildinhaltes deuten Parallelitäten in starkem Maße auf Objektbegrenzungen oder Oberflächenkonturen hin und beinhalten damit eine starke Evidenz für Gruppierungen innerhalb der **2D**-Ebene.

Die **2D**-Ebene beschreibt Strukturen, die mit dem Gestaltprinzip von *Geschlossenheit* formuliert werden. Für ihre Hypothesierung werden initiale Konturmodelle und Gruppierungen aus der **1D**- und **2×1D**-Ebene als strukturbildende Elemente verwendet. Insbesondere sind Proximitäten von entscheidender Bedeutung, da sie die notwendigen Nachbarschaftsrelationen zwischen den Elementen einer geschlossenen Figur etablieren. Innerhalb der Ebene wird zwischen den rein auf Proximitäten beruhenden Geschlossenheiten und den Geschlossenheiten, die zusätzlich Parallelitäten beinhalten, unterschieden. Letztere tragen mit ihrer stärkeren Vorstrukturierung implizit eine höhere Evidenz für eine kausale Ursache der Strukturen.

Werden die bereits ermittelten Strukturen auf einer abstrakteren Ebene betrachtet, so lassen sich einzelne geschlossene Figuren wiederum als Basiselemente komplexerer Strukturen auffassen. Dieser Ansatz wird in der $\mathbf{N} \times \mathbf{2D}$ -Ebene verfolgt, in der geschlossene Figuren oder auch Regionen, motiviert durch die Prinzipien der Nähe, Ähnlichkeit und guten Fortsetzung, in lineare Gruppen organisiert werden. Diese Ebene stellt eine entscheidende Erweiterung der Hierarchie dar, die in der vorliegenden Arbeit entwickelt wurde. Mit einer Approximation der abstrakten linearen Gruppen durch Geradenstücke und Ellipsenbögen liegt eine Beschreibungsform vor, die eine erneute Gruppierung in den Ebenen $\mathbf{1D}$ bis $\mathbf{N} \times \mathbf{2D}$ ermöglicht. Somit läßt sich ein rekursives Modell der Hierarchie entwickeln, das Strukturen mit immer abstrakterem und komplexerem Erscheinungsbild beschreiben kann (siehe auch [Lan00]).

Die Generierung der einzelnen Gruppierungshypothesen gliedert sich, entsprechend dem strukturellen Aufbau der Gruppierungshierarchie, in mehrere sequentielle Teilschritte. Zunächst werden mit Hilfe von *Bereichen perzeptiver Aufmerksamkeit* nach [Maß95b, Maß95a] Gruppierungen der $\mathbf{1D}$ und $\mathbf{2} \times \mathbf{1D}$ -Ebenen generiert. Die Ermittlung dieser Bereiche und deren Anwendung werden im nächsten Abschnitt in Anlehnung an [Pos97] vorgestellt. Diese Gruppierungen werden anschließend zur Generierung von geschlossenen Figuren in der $\mathbf{2D}$ -Ebene eingesetzt. Abschließend werden mit einer Modellierung mehrerer Gruppierungsprinzipien für Regionen die geschlossenen Figuren wiederum in strukturelle Einheiten organisiert. Eine Bewertung und Auflösung inkonsistenter Gruppierungshypothesen erfolgt mittels eines Markov Random Fields, in dem Wissen über strukturelle Kontextabhängigkeiten zwischen den Gruppierungstypen modelliert ist und das in Kapitel 6.2 eine detaillierte Vorstellung erfährt. Als Bezeichnungen für die verschiedenen Gruppierungstypen und Elementbeziehungen wird im weiteren folgende Nomenklatur verwendet:

$GElem(g)$	Menge der konstituierenden Elemente der Gruppe g
$g_1 - g_2$	kollineare Gruppierung mit $GElem(g_1 - g_2) := \{g_1, g_2\}$
$g_1 \frown g_2$	kurvilineare Gruppierung mit $GElem(g_1 \frown g_2) := \{g_1, g_2\}$
$g_1 \bullet g_2$	Nähe-Gruppierung (Proximität) mit $GElem(g_1 \bullet g_2) := \{g_1, g_2\}$
$g_1 \parallel g_2$	Parallelitäts-Gruppierung mit $GElem(g_1 \parallel g_2) := \{g_1, g_2\}$
$g_1 \circ \dots \circ g_n$	Geschlossenheit basierend auf linearen Gruppen, $GElem(g_1 \circ \dots \circ g_n) := \{g_1 \dots g_n\}$
$g_1 = \dots = g_n$	Geschlossenheit basierend auf Parallelitäten und linearen Gruppen, $GElem(g_1 = \dots = g_n) := \{g_1 \dots g_n\}$
$g_1 \oplus \dots \oplus g_n$	Gruppierung mit Geschlossenheiten als Basiselemente, $GElem(g_1 \oplus \dots \oplus g_n) := \{g_1 \dots g_n\}$

Die hier vorgestellte Gruppierungshierarchie besitzt Parallelen zu den hierarchischen Ansätzen von Mohan und Nevatia [Moh89, Moh92] sowie Sarkar und Boyer [Sar94b]. Jedoch wird mit der hier definierten Hierarchie ein flexibleres Gruppierungskonzept vorgeschlagen, das im Gegensatz zu Mohan und Nevatia die Beschreibung von allgemeinen Proximitätsrelationen und Geschlossenheiten erlaubt. Zudem werden in [Moh89, Moh92] die Bestandteilsrelationen zwischen Gruppierungen über alle Ebenen explizit durch Kanten modelliert, während in dem vorliegenden Ansatz dies über transitive Bestandteilsrelationen der jeweiligen Gruppierungstypen realisiert wird. Eine Bewertung und Auflösung von inkonsistenten Gruppierungen erfolgt bei Mohan und Nevatia über ein neuronales Hopfield Netzwerk, wohingegen in dieser Arbeit ein Markov Random Field eingesetzt wird (vergleiche Abschnitt 6.2). Die Hierarchie von Sarkar und Boyer faßt die hier definierten Ebenen **1D**, **2×1D** und **2D** in einer ersten Ebene zusammen und ist damit wesentlich grober unterteilt. Ihre zweite Ebene beschreibt Organisationen von symmetrischen Strukturen, während in der Gruppierungshierarchie in der **N×2D**-Ebene direkt eine Organisation von allgemeinen geschlossenen Figuren vorgeschlagen wird. Auch das Bewertungsverfahren von Sarkar und Boyer unterscheidet sich von dem in dieser Arbeit verwendeten. So setzen Sarkar und Boyer ein Bayes-Netzwerk zur Bewertung der Gruppierungen ein, daß zum einen nur gerichtete Unterstützung zwischen Hypothesen modelliert, und zum anderen nach jeder Generierungsphase aktiviert wird. Konkurrierende Gruppierungen müssen nachträglich durch entsprechende Verarbeitungsschritte aufgelöst werden. Diese Bewertung erfolgt in der vorliegenden Arbeit, wie auch bei Mohan und Nevatia, nach der vollständigen Generierung aller Hypothesen in einem Schritt (vergleiche Abschnitt 6.2). Zudem wird in dieser Arbeit mit der **N×2D**-Ebene das Potential für ein rekursives Gruppieren geschaffen, was in keiner der beiden erwähnten Arbeiten verfolgt wird.

4.2. Gruppierungen mit linearem Charakter

Das Ziel der **1D** und **2×1D**-Ebene ist die Organisation von Konturmodellen in Strukturen mit linearem Charakter. Für die Generierung dieser Gruppierungen sind zwei Aspekte zu betrachten. Welche Elemente sollen unter den verschiedenen Prinzipien gruppiert werden, und wie werden potentielle Partner ausgewählt?

Wie bereits in der Hierarchie angedeutet werden für die Kol- und Kurvilinearitäten jeweils nur Konturmodelle *gleichen Typs*, also Geradenstücke oder Ellipsenbögen, verwendet. Dies gilt auch für den Spezialfall der Parallelität in der **2×1D**-Ebene, in dem nur Geradenstücke für die Hypothesisierung verwendet werden. In Abbildung 4.2(a) sind Situationen dargestellt, die eine Gruppierung nach dem Prinzip der guten Fortsetzung erwarten lassen. In diesen Situationen erfolgt hier jedoch eine Gruppierung aufgrund des Prinzips der Nähe zu einer Proximität, da diese Gruppierungen nicht durch ein hier verwendetes parametrisches Modell (Geradenstück/Ellipse) beschrieben werden könnten. Auch werden lineare Gruppen gleichen Typs in einer Proximität zusammengefaßt, wenn

sie nicht parallel oder kol-/kurvilinear zueinander sind (siehe Abbildung 4.2(b)).

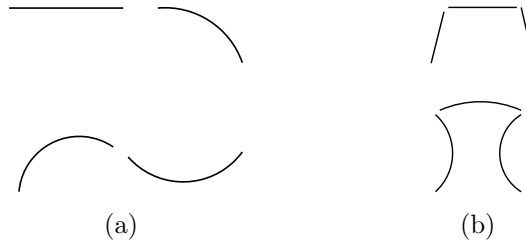


Abbildung 4.2.: Verschiedene Ausprägungen von Proximitäten: In (a) entstehen Gruppierungen aufgrund guter Fortsetzung. Da diese aber nicht durch ein lineares Konturmodell (Geradenstück/Ellipse) beschrieben werden können, werden hier Proximitätsgruppierungen generiert. (b) Strukturen mit nicht linearem oder nicht parallelem Charakter werden durch Proximitäten gruppiert.

Neben dem Typ der konstituierenden Gruppierungselemente, muß die Frage nach der Auswahl von gruppierenden Konturen diskutiert werden.

4.2.1. Bereiche perzeptiver Aufmerksamkeit

Das Ziel der Bereiche perzeptiver Aufmerksamkeit, oder auch Aufmerksamkeitsbereiche, ist die Modellierung eines Kriteriums zur Auswahl potentiell gruppierender Konturmodelle für die $1D$ und $2 \times 1D$ -Ebene. Da Objektkonturen fragmentiert sein können oder durch Verdeckung nicht sichtbar sind, stellt sich zunächst die Frage nach der Lokalität von gruppierenden Konturmodellen: In welchen Konstellationen und Bereichen erzeugen Konturen Aufmerksamkeit zu *einer* realen Kontur zu gehören? Welche Form haben diese Bereiche und wie stark sind sie ausgedehnt? In vielen Ansätzen werden, vermutlich aus pragmatischen Gründen der Handhabbarkeit und Allgemeinheit, Kreise oder Rechtecke als Suchmasken für gruppierende Konturen eingesetzt (siehe z.B. [Sar94b]). Sie werden an den Enden der Konturbeschreibungen als Suchbereiche für potentielle Gruppierungspartner benutzt. Diese Suchmasken sind jedoch keine Modelle, die die gruppierende Wahrnehmung des Menschen adäquat beschreiben. In [Maß95a] wird deshalb ein Ansatz entwickelt, der eine domänenspezifische Modellierung solcher Aufmerksamkeitsbereiche vorschlägt. Auf einer Stichprobe von Bildern werden Konturmodelle von Hand unter Berücksichtigung der verschiedenen Gestaltprinzipien als gruppierend markiert. Eine Auswertung dieser Stichprobe führt zu Suchmasken, in denen potentielle Partner für Gruppierungen lokalisiert sein können. Die Ermittlung und Anwendung dieser Masken nach [Maß95a] werden im folgenden näher erläutert.

Stichprobe

Für die Durchführung der Stichprobe wurden 40 Bilder mit prototypischen Objektkonstellationen des **baufix**[®] Szenarios aufgenommen (siehe Abbildung 4.3 für Beispielbilder). In diesen Konturbildern mußten Versuchspersonen unter Berücksichtigung der Gestaltprinzipien der Nähe, guten Fortsetzung und Parallelität paarweise gruppierende Konturmodelle markieren. Die Interpretation der Gestaltprinzipien wurde dabei den Versuchspersonen überlassen, wodurch ausschließlich Gruppierungen von Objektkonturen vorgenommen wurden. Dieses Vorgehen resultiert in einer Abhängigkeit von der **baufix**[®] Domäne, die sich jedoch durch Verallgemeinerungen reduzieren läßt.

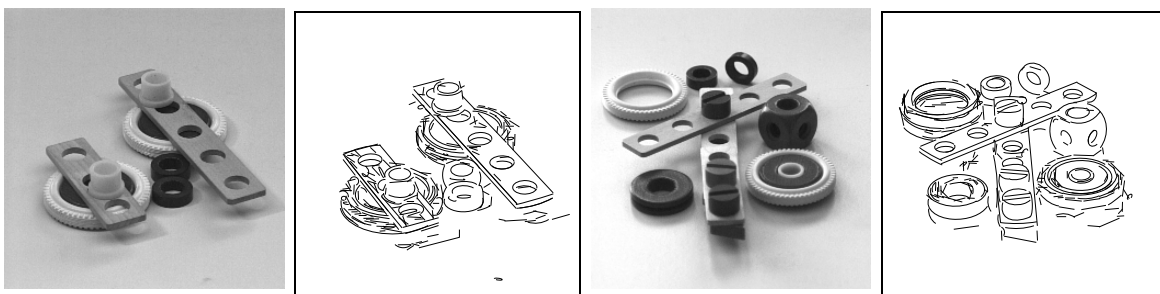


Abbildung 4.3.: Beispielbilder prototypischer Objektkonstellationen mit den Ergebnissen der initialen Konturapproximation, wie sie in der Stichprobe zur Ermittlung der Aufmerksamkeitsbereiche eingesetzt wurden.

Ermittlung der Masken

Für die Auswertung der Stichprobe müssen zunächst Bezugspunkte zwischen den manuell gruppierten Konturmodellen festgelegt werden. Für Gruppierungen der **1D**-Ebene wird das Paar der jeweils nächstgelegenen Endpunkte von gruppierten Konturmodellen als gegenseitiger Bezugspunkt definiert. Bei Parallelitäten hingegen werden die Bezugspunkte durch die Mittelpunkte der gruppierten Geradenstücke vorgegeben. Abbildung 4.4 zeigt die Lage der Bezugspunkte für die unterschiedlichen Gruppierungsprinzipien. Aus der relativen Lage der Bezugspunkte aller Gruppierungen läßt sich dann eine positionsabhängige Häufigkeitsverteilung für das jeweilige Gruppierungsprinzip ableiten. Jedoch erfolgt vor der Bestimmung dieser Verteilungen eine Normierung der relativen Lage L von Bezugspunkten, da Gruppierungsphänomene mit unterschiedlich starker Wirkung auftreten können. In Abbildung 4.4(d) wird diese Problematik veranschaulicht: Beide abgebildeten Parallelitäten besitzen eine identische relative Lage L . Ohne eine Normierung bezüglich der Länge der gruppierten Elemente, würde in der Gruppierungswirkung keine Differenz zwischen beiden Gruppen ermittelt. Folgende Normierungsvorschriften werden eingesetzt:

4. Konturbasiertes hierarchisches Gruppieren

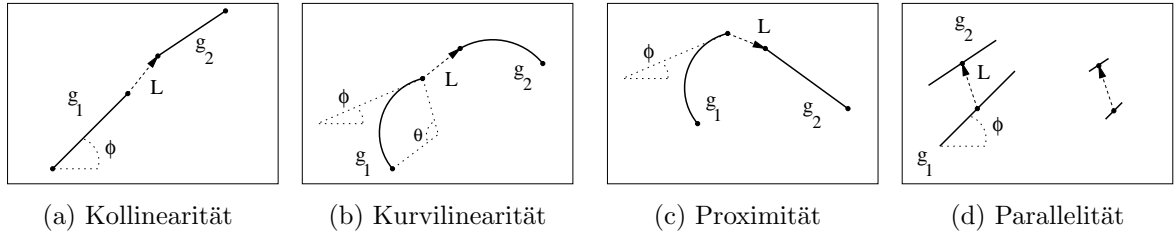


Abbildung 4.4.: Lage der Bezugspunkte von gruppierten Konturmodellen für die verschiedenen Gruppierungstypen. Die relative Lage zwischen den Konturen ist in allen Beispielen nahezu identisch gewählt. Für Parallelitäten ist ein Beispiel gezeigt, das die unterschiedliche Wirkung von Gruppierungsphänomenen und die Notwendigkeit einer Normierung der relativen Lage verdeutlicht (siehe Text).

Kollinearität: Für Kollinearitäten ergibt sich die Normierung der relativen Lage L zum einen durch Skalierung mit der Länge $l(g)$ des aktuell betrachteten Geradenstücks:

$$\tau_g := \frac{1}{l(g)} \quad (4.1)$$

Neben der Längenskalierung wird die relative Lage zusätzlich bezüglich der Orientierung ϕ des Geradenstücks durch eine Rotation $R_{(\frac{\pi}{2}-\phi)}$ korrigiert und ergibt sich zu:

$$L_{norm} := \tau_g R_{(\frac{\pi}{2}-\phi)} L$$

Die Häufigkeitsverteilung für kollineare Gruppierungen ist damit derart normiert, daß jedes Geradenstück parallel zur y -Achse eines Koordinatensystems ausgerichtet und von identischer Länge ist. Dieselbe Normierung wird auch für die anderen Gruppierungsannahmen angestrebt.

Kurvilinearität: Für Kurvilinearitäten wird eine zu Kollinearitäten analoge Normierung durchgeführt. Bezüglich der Längenskalierung ist jedoch die zirkuläre Eigenschaft von elliptischen Konturmodellen zu berücksichtigen. Bei zunehmenden Schließungsgrad eines Ellipsenbogens e nimmt die lineare Gruppierungswirkung ab, da eine stärkere Tendenz zur Wahrnehmung einer geschlossenen Gestalt erzeugt wird. Mit dieser Einschränkung ergibt sich die Längenskalierung zu:

$$\tau_e := \begin{cases} \frac{1}{l(e)} & : 0 \leq \theta < \pi \\ \frac{1}{u-l(e)} & : \pi \leq \theta < 2\pi \end{cases} \quad (4.2)$$

Hierbei bezeichnet u den vollen Umfang der entsprechend parametrisierten Ellipse und θ den aufgespannten Öffnungswinkel des Ellipsenbogens (siehe Abbildung 4.4(b)). Die Normierung bezüglich der Richtung wird bei Kurvilinearitäten

anhand der Tangentorientierung ϕ im Endpunkt des Konturmodells durchgeführt. Darüber hinaus muß eine Berücksichtigung der Krümmung erfolgen, da nur Ellipsenbögen mit gleicher Krümmungsrichtung κ als kurvilinear gruppiert werden. Um die Krümmungsrichtung zu beachten, wird deshalb gegebenenfalls eine Spiegelung an der x-Achse durchgeführt:

$$L_{norm} := \tau_e \begin{pmatrix} \text{sgn}(\kappa) \\ 0 \end{pmatrix} R_{(\frac{\pi}{2}-\phi)} L$$

Proximität: Bei Proximitäten werden die Normierungsvorschriften bezüglich Länge und Orientierung analog zu Kol- und Kurvilinearität in Abhängigkeit des betrachteten Konturmodells gewählt:

$$L_{norm} := \tau R_{(\frac{\pi}{2}-\phi)} L, \quad \text{mit} \quad \tau = \begin{cases} \tau_g & \text{aktuelles Modell ist Geradenstück} \\ \tau_e & \text{aktuelles Modell ist Ellipsenbogen} \end{cases}$$

Parallelität: Auch für Parallelitäten werden die Normierungsdefinitionen von Kollinearitäten übernommen. Für die Normierung der Orientierung der relativen Lage wird das gruppierenden Geradenstück parallel zur y-Achse angenommen:

$$L_{norm} := \tau_g R_{\phi} L$$

Bei der Durchführung der Stichprobe wurden die Gruppierungen immer paarweise zwischen Konturmodellen gebildet. Aus diesem Grund erfolgt die Bestimmung der Häufigkeitsverteilungen jeweils für beide Konturmodelle als gruppierendes Element. Diese Häufigkeitsverteilungen der normierten relativen Lagen werden anschließend mit Gaußfiltern unterschiedlicher Breite verschmiert, um den Einfluß stärkerer Häufigkeiten auf benachbarte Positionen zu erhöhen. Hierzu nimmt die Breite der Filter in Abhängigkeit von der Häufigkeit an der jeweiligen Position zu. Nach einer Schwellwertoperation wird dann auf die verbleibenden Bereiche ein iteratives morphologisches Closing angewendet. Die so ermittelten normierten *Bereiche perceptiver Aufmerksamkeit* werden schließlich durch die Bestimmung eines einhüllenden konvexen Polygons beschrieben. Mit dieser Approximation wird die Abhängigkeit von der Stichprobe und auch allgemein von der **baufix**[®] Domäne bereits reduziert. In Abbildung 4.6 sind die resultierenden Polygone dargestellt, während in Abbildung 4.5 beispielhaft die normierte gefilterte Häufigkeitsverteilung für Kollinearitätsgruppierungen gezeigt ist.

Die Ergebnisse der Stichprobe werden zudem für die Bestimmung weiterer Parameter eingesetzt. So werden mit den normierten relativen Lagen Einschränkungen bezüglich der erlaubten Orientierungsdifferenzen für potentielle Gruppierungspartner ermittelt. Diese werden im nächsten Abschnitt für den jeweiligen Gruppierungstyp angegeben. Darüber

4. Konturbasiertes hierarchisches Gruppieren

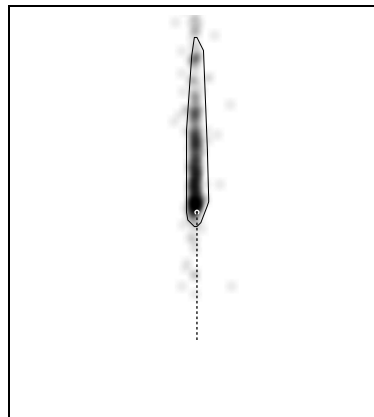


Abbildung 4.5.: Normierte gefilterte Häufigkeitsverteilung für Kollinearitätsgruppierungen mit markiertem Aufmerksamkeitsbereich und normiertem Konturmodell (gestrichelt dargestellt).

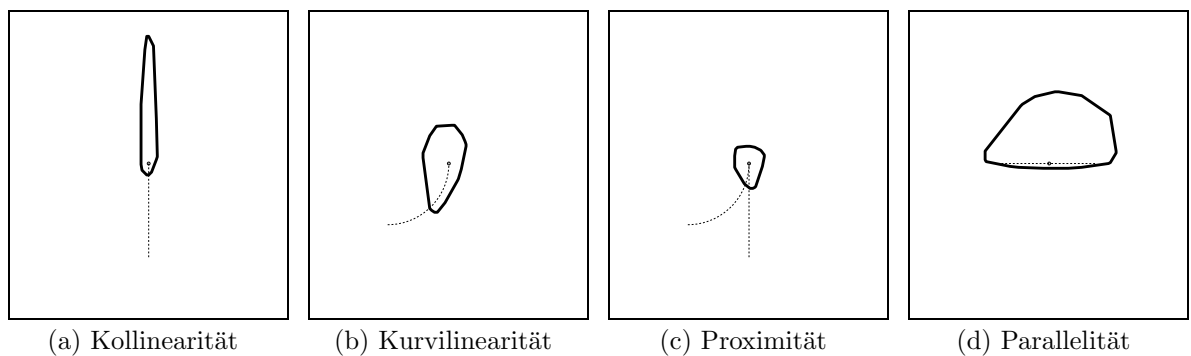


Abbildung 4.6.: Suchmasken für die Bereiche perzeptiver Aufmerksamkeit bezogen auf normierte Konturmodelle (gestrichelt dargestellt). Die Bezugspunkte der Konturmodelle befinden sich jeweils in der Mitte des Bildes und sind mit einem Punkt markiert. Die Krümmung von Ellipsenbögen bei kurvilinearen Gruppen und Proximitäten ist einheitlich positiv gewählt (siehe Text).

hinaus wird für die Hypothesisierung von Parallelitäten die minimal erforderliche prozentuale Überlappung von zwei Geradenstücken orthogonal zu ihrer Symmetrieachse bestimmt. Mit diesen zusätzlichen Schwellwerten und den normierten Polygonen für die Aufmerksamkeitsbereiche kann nun eine automatische Generierung von Gruppierungen der **1D** und **2×1D**-Ebene erfolgen.

4.2.2. Hypothesisierung

Die Generierung der Gruppierungshypothesen mit linearem Charakter erfolgt, wie bereits in der Abbildung 4.1 angedeutet, sequentiell mit zunehmendem Komplexitätsgrad. Nach der initialen Approximation werden zunächst die kollinearen Gruppierungsannahmen vollständig generiert. Anschließend erfolgt die Hypothesisierung von Kurvilinearitäten. Den letzten Schritt für Gruppierungen der **1D**-Ebene bildet immer die Generierung von Proximitätsannahmen, da sie zwischen allen linearen Gruppen, also auch zwischen Kol- und Kurvilinearitäten, gebildet werden können. Im Anschluß erfolgt die Generierung der Parallelitätshypothesen, die aber prinzipiell auch bereits nach der Ermittlung der Kollinearitäten erfolgen kann, da nur Geradenstücke verwendet werden. Im folgenden wird das für alle Gruppierungsprinzipien identische algorithmische Vorgehen mit den jeweiligen typabhängigen Besonderheiten beschrieben.

Bearbeitungsreihenfolge und Generierung

Bei der Betrachtung einer Szene stellt sich zunächst die Frage, welche Konturmodelle oder Strukturen die stärkste gruppierende Wirkung besitzen. In dem beschriebenen Ansatz wird diese Auswahl mit der analytischen Länge $l(g)$ der linearen Gruppen betrieben: Alle konstituierenden linearen Gruppen, die für die Hypothesenbildung eines Gruppierungstyps eingesetzt werden können, werden zunächst in einen nach der Länge absteigend sortierten Stapel O eingefügt. Diese Sortierung begründet sich durch die Annahme, daß längere lineare Gruppen eine stärkere Aufmerksamkeit für die hier untersuchten Gruppierungsphänomene erregen und somit eine höhere Evidenz für eine kausale Ursache im Bild vorliegt. Nach der Sortierung wird von diesem Stapel jeweils immer das oberste Element g_{cur} für eine Untersuchung auf Gruppenbildung mit anderen linearen Gruppen $g_i \in O \setminus \{g_{cur}\}$ gewählt. Es werden, unter Berücksichtigung der später beschriebenen jeweiligen Auswahlkriterien und Einschränkungen, alle zulässigen Gruppierungshypothesen erstellt, für die g_{cur} als gruppierendes Element auftritt. Die weitere Bearbeitung hängt dann vom aktuell untersuchten Gruppierungstyp ab.

Kol- und Kurvilinearität: Kann eine kol- oder kurvilineare Gruppierungshypothese mit g_{cur} aufgestellt werden, so werden die Kontursegmente der konstituierenden Elemente zusammengefaßt und durch eine neu approximierete Gruppe beschrieben:

$$g_{new} = g_{cur} - g_i \quad \text{oder} \quad g_{new} = g_{cur} \cap g_i$$

Diese neue lineare Gruppe wird in den Stapel entsprechend ihrer Länge einsortiert:

$$O_{new} = \text{sort}(O \cup \{g_{new}\})$$

Sind alle zulässigen Gruppierungen mit g_{cur} als gruppierendes Element hypothesiert, wird dieses Element vom Stapel entfernt:

$$O_{new} = O \setminus \{g_{cur}\}$$

Mit der Berücksichtigung von $g_{new} \in O$ ist eine sukzessive Expansion kleinerer Strukturen zu größeren Einheiten gewährleistet. Lokale Evidenzen führen damit zu globalerem Einfluß in der Strukturierung des Bildinhaltes.

Parallelität und Proximität: Bei diesen Gruppierungsprinzipien wird nach der Generierung aller zulässigen Gruppierungshypothesen die aktuelle Gruppe g_{cur} vom Stapel O entfernt, und mit dem nächsten Element von O fortgefahren.

Ist der gesamte Stapel für einen Gruppierungstyp abgearbeitet ($O = \emptyset$), werden *offensichtlich redundante* Hypothesen dieses Typs eliminiert, um eine weitere Verarbeitung effizient zu gestalten. Welche Arten von Redundanz für die einzelnen Prinzipien auftreten, wird weiter unten erläutert. Die Generierung *identischer* Gruppierungen wird allerdings schon in der Auswahl potentieller Partner unterbunden.

Auswahl potentieller Partner

Zur Auswahl von potentiellen Gruppierungspartnern des jeweiligen Gruppierungstyps, werden die Suchmasken für Aufmerksamkeitsbereiche und die daraus ermittelten einschränkenden Parameter eingesetzt: Für jede aktuelle lineare Gruppe g_{cur} wird die dem Gruppierungstyp entsprechende Suchmaske an den Bezugspunkten von g_{cur} plazierte. Es liegt nahe, die Suchmasken anhand der ihnen zugeordneten Normierungsvorschrift L_{norm} und mit den Parametern der aktuellen linearen Gruppe g_{cur} invers zu skalieren und rotieren. In Bezug auf die Rotation ist dies korrekt und wird entsprechend durchgeführt. Bezüglich der Längenskalierung hat sich jedoch in umfangreichen Experimenten die inverse lineare Skalierung als nicht adäquat erwiesen. Aus diesem Grund wurde folgende inverse Längenskalierung $1/\tau'$ empirisch ermittelt:

$$\frac{1}{\tau'} := \begin{cases} 1 - (\frac{1}{\tau} - 1)^2 & : 0 \leq \frac{1}{\tau} < 1 \\ 1 + \frac{1}{2}\sqrt{\frac{1}{\tau}} & : 1 \leq \frac{1}{\tau} \end{cases}, \text{ mit } \tau = \begin{cases} \tau_g : g_{cur} = \text{Gerade, Def. (4.1)} \\ \tau_e : g_{cur} = \text{Ellipse, Def. (4.2)} \end{cases} \quad (4.3)$$

Mit dieser Skalierung erhalten dann Gruppierungen mit einer normierten Länge $[0, 1]$ einen größeren Suchbereich, als mit einer linearen Skalierung. Hingegen werden die Suchbereiche von Gruppierungen mit einer normierten Länge $[1, \infty[$ zunehmend weniger expandiert. Diese Modellierung hat sich als angemessen herausgestellt. Alle linearen Gruppen g_i , die dann einen Bezugspunkt innerhalb dieses skalierten und rotierten Suchbereichs besitzen, werden auf die einschränkenden Bedingungen untersucht. Davon ausgenommen sind Gruppierungen g_j , mit denen g_{cur} bereits eine Gruppierung gebildet hat. Sie werden nicht weiter betrachtet. Dieses Vorgehen ist für alle Gruppierungstypen identisch und in Abbildung 4.7 veranschaulicht. Bei Parallelitäten ist zu beachten, daß die Anwendung der Aufmerksamkeitsbereiche im gleichen Bezugspunkt auf beide Seiten der aktuell betrachteten Gruppe erfolgt, da auf der jeweiligen Seite der Gruppe potentielle Gruppierungspartner lokalisiert sein können.

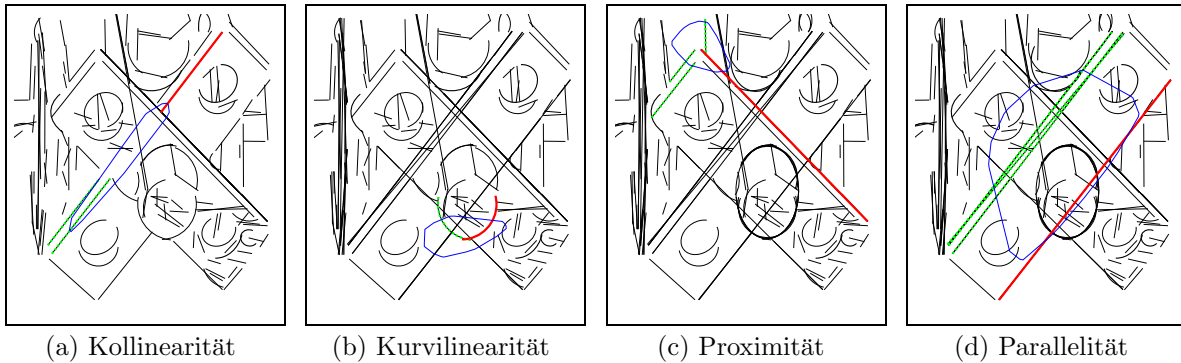


Abbildung 4.7.: Anwendung der Aufmerksamkeitsbereiche als Suchmasken für potentielle Gruppierungspartner. Das aktuell gruppierende Element (rot) ist dicker markiert. Die rotierten und skalierten Aufmerksamkeitsbereiche (blau) sind an den Bezugspunkten angesetzt und potentielle Gruppierungspartner (grün) sind gepunktet hervorgehoben.

Validierung der Einschränkungen

Nachdem eine Menge von potentiellen Partner in den Suchbereichen der aktuellen Gruppe identifiziert ist, werden für jedes Element dieser Menge die einschränkenden Gruppierungsbedingungen überprüft. Bei allen Gruppierungstypen betrifft dies zunächst die Orientierungsdifferenz zwischen dem gruppierenden Element g_{cur} und dem potentiellen Partner g_i :

$$|\phi_{g_{cur}} - \phi_{g_i}| \leq \Delta_\phi$$

Die Schwellwerte der zulässigen Abweichungen sind für das jeweilige Gruppierungsprinzip aus der Stichprobe ermittelt worden und in Tabelle 4.1 angegeben. Weitere Einschränkungen beziehen sich auf verschiedene Gruppierungsprinzipien:

Kurvilinearität: Für die Generierung von Kurvilinearitäten wird neben der Einschränkung bezüglich der Orientierungsdifferenz eine übereinstimmende Krümmung der gruppierten Ellipsenbögen gefordert. Andernfalls werden diese Ellipsenbögen in einer Proximität zusammengefaßt (siehe Beispiel in Abbildung 4.2).

$\Delta_\phi(-)$	2.9°
$\Delta_\phi(\frown)$	17.2°
$\Delta_\phi(\parallel)$	2.8°
Δ_{length}	30 %

Tabelle 4.1.: Schwellwerte zur Einschränkung von Orientierungs- und Längenabweichungen bei der Hypothesisierung der verschiedenen Gruppierungstypen.

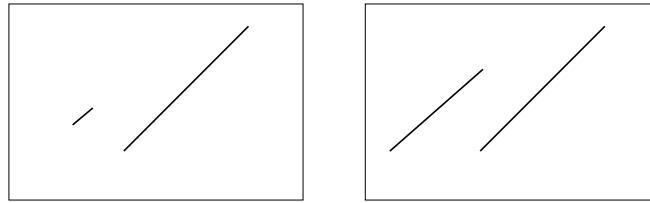


Abbildung 4.8.: Motivation zur Gruppierungswirkung bei Parallelitäten bei nur geringfügiger und stärkerer Überlappung gruppierender Elemente.

Proximität: Um eine redundante Hypothesierung von Proximitäten zu vermeiden, werden lineare Gruppen gleichen Typs nur dann in Proximitäten gruppiert, wenn sie die zulässigen Schwellwerte für eine Gruppierung aufgrund guter Fortsetzung überschreiten (für Beispiele siehe Abbildung 4.2). Des Weiteren wird bei Proximitäten aufgrund der fehlenden Richtungsselektivität gefordert, daß die Bezugspunkte wechselseitig in den Suchmasken der Gruppierungspartner zu liegen kommen. Ansonsten würde die Anzahl der Proximitäten unüberschaubar werden.

Parallelität: Eine weitere Einschränkung greift bei der Generierung von Parallelitäten: Zwei parallele lineare Gruppen (Geradenstücke oder Kollinearitäten) müssen mindestens eine bestimmte projektive Überlappung auf ihrer Symmetrieachse aufweisen, um eine gültige Parallelität zu bilden. Diese Einschränkung ist in Abbildung 4.8 an einem Beispiel nachzuvollziehen. Zur effizienten Berechnung dieser Einschränkung wird approximativ das Verhältnis der Segmentlängen verwendet:

$$\frac{\min(l(g_1), l(g_2))}{\max(l(g_1), l(g_2))} \geq \Delta_{length}$$

Die minimal geforderte Überlappung ist in Tabelle 4.1 angegeben.

Eliminierung redundanter Hypothesen

Eine automatische Anwendung der Aufmerksamkeitsbereiche und die Evaluierung von einschränkenden Bedingungen kann zu mehreren Hypothesen führen, die redundante Beschreibungen einer Struktur darstellen. Diese Redundanzen können in vielen Fällen nach der Generierung aller Hypothesen eines Typs eliminiert werden, wodurch die Komplexität weiterer Bearbeitungsschritte reduziert werden kann. Im folgenden werden die Redundanzphänomene des jeweiligen Gruppierungstyps motiviert und deren Auflösung beschrieben.

Kollinearität: Für Kollinearitäten werden Gruppierungshypothesen als redundant definiert, wenn zwei oder mehr Gruppierungen die identischen Endelemente besitzen,

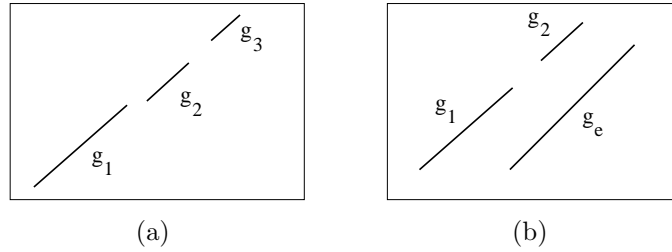


Abbildung 4.9.: Motivation zur Eliminierung redundanter (a) Kollinearitäten und (b) Parallelitäten (siehe Text).

und die Gruppierungen eine echte Teilmengenbeziehung zueinander haben. Abbildung 4.9(a) veranschaulicht dieses Phänomen. Durch das algorithmische Vorgehen entstehen hier folgende Gruppierungen:

$$g_a = g_1 - g_2 \quad , \quad g_b = g_1 - g_3 \quad , \quad g_c = g_2 - g_3 \quad , \quad g_d = g_1 - g_2 - g_3$$

Die Gruppierung g_b entsteht durch die räumliche Ausdehnung des Aufmerksamkeitsbereichs von g_1 , der zu einer Überbrückung von g_2 führt. Damit besteht eine Redundanz zwischen g_b und g_d , da sie identische Endelemente besitzen und $GElem(g_b) \subseteq GElem(g_d)$ gilt. Die Gruppierung g_d interpretiert jedoch die zugrundeliegenden Kantenelemente in einer vollständigeren Weise, als es die Gruppierung g_b leistet. Aus diesem Grund wird g_d als resultierende Beschreibung für diese Struktur gewählt und g_b verworfen.

Parallelität: Die Redundanz von Parallelitäten ist ebenfalls über eine Teilmengenbeziehung definiert und in Abbildung 4.9(b) motiviert. Alle Parallelitäten, bei denen eine Hälfte durch die identische lineare Gruppe g_e beschrieben wird, werden paarweise auf Redundanz untersucht. Redundante Hypothesen liegen vor, wenn die zu g_e parallelen linearen Gruppen von zwei Parallelitätshypothesen eine Teilmengenbeziehung zueinander haben. In Abbildung 4.9(b) sind die Gruppierungen

$$g_l = g_e \parallel g_1 \quad , \quad g_m = g_e \parallel g_2 \quad , \quad g_n = g_e \parallel (g_1 - g_2)$$

redundante Beschreibungen der gezeigten parallelen Struktur. Als korrekte Gruppierung wird g_n gewählt, da es die vollständige Gestalt der Parallelität beschreibt. Die Hypothesen g_l, g_m werden verworfen.

Mit der Beschreibung der Verfahrensweise zur Redundanzauflösung ist der gesamte Prozeß der Generierung von Gruppierungshypothesen der **1D**- und **2×1D**-Ebene vorgestellt. Die Ergebnisse einer Hypothesisierung auf Beispielbildern werden in Verbindung mit den Ergebnissen der nun folgenden Gruppierungsprinzipien in Kapitel 7.1 diskutiert.

4.3. Geschlossene Figuren

Die Wahrnehmung von geschlossenen Figuren ist als ein weiteres wichtiges Gruppierungsphänomen zur Identifikation komplexer Strukturen anzusehen. In der Gruppierungshierarchie werden Strukturen diesen Typs in der **2D**-Ebene eingeführt. Wie bereits in Kapitel 4.1 angesprochen, baut diese Ebene auf Hypothesen der darunterliegenden Ebenen auf. Im folgenden wird eine *Geschlossenheit* als eine zyklische Sequenz von linearen Gruppen definiert, die über Proximitätshypothesen eine Näherrelation in ihren jeweiligen Endpunkten besitzen. Die Detektion von Geschlossenheiten beruht damit auf der Suche nach solchen Sequenzen.

Proximitätsgraph

Für die Generierung von Geschlossenheiten wird zunächst ein ungerichteter *Proximitätsgraph* aufgebaut. In diesem Graphen definieren die Endpunkte von linearen Gruppen die Knoten, während die Kanten des Graphen in zwei Typen unterschieden werden:

Gruppierungskante: Die *Gruppierungskanten* werden von den linearen Gruppen gebildet und beschreiben deren geometrischen Verlauf zwischen den Endpunkten.

Proximitätskante: Die *Proximitätskanten* beschreiben die Verbindung von zwei linearen Gruppen zwischen den beiden Endpunkten, die in einer wechselseitigen Proximitätsbeziehung zueinander stehen. Hierbei ist zu beachten, daß zwei lineare Gruppen durchaus über zwei Proximitätskanten in ihren jeweiligen Endpunkten verbunden sein können. Ein Beispiel hierfür ist in Abbildung 4.10(a) gezeigt. Beide Ellipsenbögen stehen über die entsprechenden Endpunkte in einer Näherrelation zueinander und können so eine zweielementige Geschlossenheit bilden.

Parallelitätshypothesen werden in der Struktur des Graphen nicht berücksichtigt, da sie implizit durch ihre konstituierenden linearen Gruppen vertreten sind. Sie werden erst bei der endgültigen Generierung einer Geschlossenheitsgruppierung als zusätzliche Evidenz für die Struktur ausgenutzt.

Pfadsuche

Für die Bestimmung von zyklischen Pfaden werden bei dem Aufbau des Graphen nur lineare Gruppen eingesetzt, die in beiden Endpunkten eine Proximitätsbeziehung zu anderen linearen Gruppen besitzen. So kann im Hinblick auf die Effizienz einer Suche schon frühzeitig eine Reduzierung der Kantendichte im Graphen erfolgen. Anschließend wird mit einer beliebigen Gruppierungskante g_{cur} in dem Proximitätsgraphen eine Suche nach zyklischen Pfaden gestartet. Hierzu wird eine vollständige Tiefensuche durchgeführt, die als Ziel die Verbindung der beiden Endpunkte von g_{cur} über eine alternierende Folge von Proximitätskanten n_i und Gruppierungskanten g_j hat. Während dieser Suche wird als einzige Restriktion die Vermeidung einer Überquerung des bereits zurückgelegten Weges

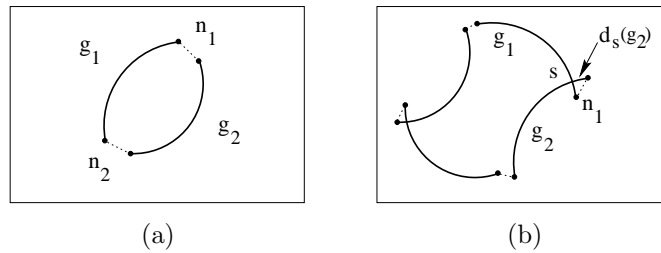


Abbildung 4.10.: Einfache Proximitätsgraphen in denen Gruppierungskanten durchgezogen und Proximitätskanten gepunktet markiert sind. Die Knoten sind durch dicke Punkte angedeutet. (a) Zwei lineare Gruppen können direkt über zwei Proximitäten n_1, n_2 verbunden sein und eine Geschlossenheit bilden. (b) Überschneidungen zwischen linearen Gruppen g_1, g_2 sind bei der Pfadsuche im Bereich ihrer gemeinsamen Näherrelation n_1 zulässig, wenn der Abstand $d_s(g_2)$ klein genug ist.

gefordert. Kleinere Überschneidungen zwischen zwei Gruppierungskanten g_1, g_2 im Bereich der gemeinsamen Näherrelation n_1 sind jedoch zulässig, wie es in [Abbildung 4.10\(b\)](#) veranschaulicht ist. Die Zulässigkeit einer Überschneidung wird durch den Schwellwert $\Delta_{intersect}$ (siehe [Tabelle 4.2](#)) eingeschränkt. Bezeichnet $d_s(g)$ den Abstand des Schnittpunktes s vom Endpunkt der linearen Gruppe, so muß folgende Bedingung für einen weiterhin gültigen Pfad erfüllt sein:

$$d_s(g) \leq \Delta_{intersect} l(g),$$

wobei $l(g)$ die Länge der linearen Gruppe bezeichnet. Für jeden gültigen zyklischen Pfad über g_{cur} wird schließlich eine Geschlossenheit hypothetisiert. Sind alle Pfade von g_{cur} aus verfolgt worden, wird diese Kante und die angrenzenden Proximitätskanten als vollständig expandiert markiert. Anschließend wird mit den verbleibenden Gruppierungskanten jeweils eine neue Suche gestartet, bei der alle bereits vollständig expandierten Gruppierungskanten nicht mehr berücksichtigt werden. Das Verfahren iteriert bis alle Gruppierungskanten vollständig expandiert sind und liefert die Menge aller zulässigen Geschlossenheitshypothesen.

$\Delta_{intersect}$	0.15
Δ_{closed}	0.85
$\Delta_{overlap}(\circ)$	0.95

Tabelle 4.2.: Empirisch ermittelte Schwellwerte zur Einschränkung der Pfadsuche, der Hypothesisierung und der Redundanzauflösung von Geschlossenheiten.

Hypothetisierung

Ein gefundener zyklischer Pfad wird sofort während des Suchprozesses als Geschlossenheit G hypothetisiert. Hierzu werden die Gruppierungskanten des aktuellen Pfades als Menge der konstituierenden Elemente $GElem(G)$ verwendet. Die Proximitäten werden nicht berücksichtigt, da sie, wie bereits in Kapitel 4.1 erwähnt, eher einen strukturbildenden als einen strukturbeschreibenden Charakter besitzen. Zur Bestimmung von Parallelitätsgeschlossenheiten werden alle Elemente von G , die paarweise eine Parallelitätshypothese p bilden, durch diese entsprechende Parallelität $g_i \parallel g_j$ ersetzt:

$$g_1 \circ \dots \circ g_i \circ \dots \circ g_j \circ \dots \circ g_n \quad \rightarrow \quad g_1 \equiv \dots \equiv p \equiv \dots \equiv g_n$$

Einzelne Hypothesen tragen somit eine höhere Evidenz für einen kausalen Zusammenhang und werden in Parallelitätsgeschlossenheiten konvertiert. Auch bei Ellipsenbögen, die nahezu vollständig geschlossen sind, besteht eine Evidenz für eine kausale Ursache. Aus diesem Grund werden Ellipsenbögen (initiale Konturmodelle oder Kurvilinearitäten) als einelementige Geschlossenheit hypothetisiert, falls sie folgende Bedingung erfüllen:

$$\frac{l(e(\theta))}{l(e(2\pi))} \geq \Delta_{closed}$$

Hierbei bezeichnet $l(e(\alpha))$ die analytische Länge eines parametrisierten Ellipsenbogens e mit dem Öffnungswinkel α und θ den aktuellen Öffnungswinkel von e . Der Wert für Δ_{closed} ist in Tabelle 4.2 angegeben. Für die weitere Verarbeitung wird jede Geschlossenheit schließlich durch einen geschlossenen Polygonzug geometrisch approximiert.

Eliminierung redundanter Hypothesen

Bedingt durch die Verfahrensweise zur Generierung von Proximitätshypothesen können redundante Geschlossenheiten entstehen. Zur Hypothetisierung von Proximitätsgruppierungen werden die Aufmerksamkeitsbereiche entsprechend der Länge des aktuell gruppierenden Elementes skaliert. Dadurch kann es zu Überbrückungen von kürzeren linearen Gruppen kommen, und es entstehen Näherrelationen, die zu redundanten Geschlossenheiten führen. In Abbildung 4.11(a) ist dieses Phänomen exemplarisch dargestellt. Zur Auflösung dieser Redundanz wird, wie bei der Eliminierung von kollinearen Gruppierungen, eine Teilmengenbeziehung ausgewertet. Darüber hinaus wird eine Betrachtung der von einer Geschlossenheit umschlossenen Fläche vorgenommen. Folgende Bedingungen identifizieren zwei Geschlossenheit G_1, G_2 als redundant:

$$GElem(G_1) \subseteq GElem(G_2) \quad \wedge \quad \frac{A(G_1) \cap A(G_2)}{\max(A(G_1), A(G_2))} \geq \Delta_{overlap}(\circ),$$

wobei $A(G)$ die Fläche des Polygonzuges von G bezeichnet und $\Delta_{overlap}(\circ)$ in Tabelle 4.2 angegeben ist. Sind diese Bedingungen für zwei Geschlossenheiten erfüllt, so wird die

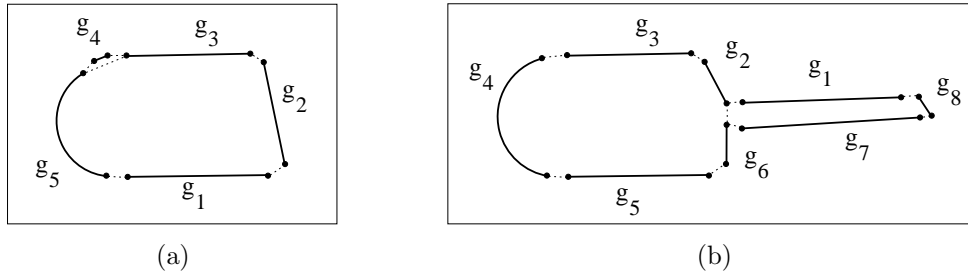


Abbildung 4.11.: Motivation zur Generierung von Geschlossenheiten mit Parallelitäten und zur Redundanzauflösung bei Geschlossenheiten.

(a) Die Gruppierung $g_1 \circ g_2 \circ g_3 \circ g_4 \circ g_5$ wird durch die Geschlossenheit $g_2 = \{g_1 \parallel g_3\} = g_4 = g_5$ ersetzt. Die redundante Hypothese $g_2 = \{g_1 \parallel g_3\} = g_5$ wird aufgrund schwächerer Evidenz gegenüber $g_2 = \{g_1 \parallel g_3\} = g_4 = g_5$ verworfen.

(b) Trotz einer Teilmengenbeziehung zwischen $g_2 = \{g_3 \parallel g_5\} = g_4 = g_6$ und $g_2 = \{g_3 \parallel g_5\} = g_4 = g_6 = \{g_1 \parallel g_7\} = g_8$ bleiben aufgrund einer deutlichen Abweichung in der überlappenden Fläche beide Hypothesen erhalten.

Hypothese eliminiert, die weniger konstituierende Elemente besitzt ($|G_{Elem}(G_1)| < |G_{Elem}(G_2)|$). Bei gleicher Anzahl konstituierender Elemente entscheidet die Anzahl stärker strukturierender Elemente (Parallelitäten), welche der beiden Geschlossenheiten entfernt wird.

Mit der Eliminierung redundanter Hypothesen liegen abschließend alle zulässigen Geschlossenheitshypothesen vor. Die Anwendung des Verfahrens auf Beispielbildern wird im Zusammenhang mit den anderen Gruppierungsergebnissen in Kapitel 7.1 diskutiert. Ein Verfahren zur Generierung von Geschlossenheiten, falls Nähegruppierungen aufgrund schwach ausgeprägter Kontraste nicht erzeugt werden konnten, wird in Kapitel 5 vorgestellt.

4.4. Gruppierung von geschlossenen Figuren

Die oberste Ebene der Hierarchie beschreibt Gruppierungsphänomene, die bereits erzeugte Strukturen als Entitäten für weitere komplexere Strukturierungen verwenden. Das Ziel dieser $\mathbf{N} \times \mathbf{2D}$ -Ebene ist die Beschreibung linearer Anordnungen von geschlossenen Figuren, die sowohl durch die Gruppierungen der $\mathbf{2D}$ -Ebene als auch durch eine Regionensegmentierung gegeben sein können. Für diese Aufgabe wird vorausgesetzt,

daß die zu organisierenden Strukturen durch Polygone¹ approximiert sind. Idealerweise beschreiben diese Polygone genau die Projektion einer Objektoberfläche, eines Objektbestandteiles oder eines vollständigen Objektes.

Für die Realisierung einer automatischen Generierung von Gruppierungen in dieser Ebene ergeben sich verschiedene Fragestellungen (siehe auch [Lan00]):

- Welche Gestaltprinzipien wirken, um das Gruppierungsphänomen guter Fortsetzung auf komplexen Figuren hervorzurufen?
- Wie kombinieren sich die einzelnen Gesetzmäßigkeiten, um eine adäquate Beschreibung solcher Gruppierungsphänomene zu erhalten?
- Wie können diese Gesetzmäßigkeiten überhaupt quantifiziert werden?

Bezugnehmend auf die Problematik, welche Gestaltprinzipien eine Gruppierung bewirken, ist zunächst die Regelmäßigkeit des Auftretens von Strukturen für eine Gruppenbildung auffallend. Hierbei ist scheinbar die regelmäßig nahe, zum Teil äquidistante, Anordnung von Objekten oder Objektstrukturen von Bedeutung. Es ist zu erwarten, daß das Gesetz der Nähe eine wichtige Komponente für die Wahrnehmung solcher Strukturen ist (siehe auch [Low85]). Jedoch ist die Nähe nicht allein verantwortlich für eine Gruppenbildung. Betrachtet man beispielsweise eine Szene in der Autos und Lastkraftwagen in willkürlicher Abfolge, aber regelmäßiger Anordnung auf einem großen Parkplatz abgestellt sind, so lassen sich Strukturen durch die regelmäßige Anordnung erkennen. Sind die Fahrzeuge dann aber in Gruppen gleichförmiger Bauart aufgestellt, so ist neben der Nähe auch eine Gruppenbildung über die Ähnlichkeit der Fahrzeuge festzustellen. Die Ähnlichkeit ist also ein weiteres Merkmal zur Gruppierung von komplexen Strukturen. Für die Bestimmung von Ähnlichkeit ist sicherlich Farbe ein wichtiges Merkmal. Diese steht jedoch bei dem konturbasierten Gruppierungsansatz nicht zur Verfügung. Aus diesem Grund wird im weiteren Ähnlichkeit nur über die Form von geschlossenen Figuren definiert. Neben der Nähe und Ähnlichkeit von Objektformen, ist auch deren Anordnung für die Strukturbildung entscheidend. Sind in einer Szene verschiedene Objekte chaotisch mit nahezu identischem Abstand zueinander plaziert, so bewirkt eine geraden- oder ellipsenförmige Anordnung gleichförmiger Objekte eine Strukturierung mit linearem Charakter. Aus dieser Vorstellung ergibt sich daher die Berücksichtigung des Prinzips der guten Fortsetzung. Im Übrigen läßt sich mit der Beschreibung solcher Strukturen durch Geradenstücke und Ellipsenbögen die Gruppierungshierarchie hervorragend auf einem abstrakteren Niveau wiederverwenden. Weitere Gesetzmäßigkeiten der Gestalttheorie finden in der $\mathbf{N} \times \mathbf{2D}$ -Ebene der Gruppierungshierarchie keine Berücksichtigung. Es werden daher folgende Gestaltprinzipien für die Gruppierung von geschlossenen Figuren eingesetzt:

¹Die Begriffe *Polygon* und *Region* werden in Bezug auf die geometrische Form geschlossener Figuren oder Geschlossenheiten synonym zueinander benutzt.

- Prinzip der Nähe,
- Prinzip der Ähnlichkeit und
- Prinzip der guten Fortsetzung.

Mit der Auswahl der einzusetzenden Gestaltprinzipien stellt sich die Frage nach der Kombination dieser verschiedenen Prinzipien. Hierzu bieten sich unterschiedliche Vorgehensweisen an: Angenommen, eine Gruppierung mit linearem Charakter ist von der Nähe und Ähnlichkeit der gruppierten Elemente zunächst unabhängig. Es können dann zuerst alle Strukturen nach dem Prinzip der guten Fortsetzung gebildet werden. Dieses Vorgehen entspricht einer *ModelRecovery*-Phase wie bei der Konturapproximation nach Leonardis (siehe Abschnitt 2.1.3, Seite 17). Daran anschließend erfolgt eine Selektion von linearen Strukturen, die aufgrund der Gruppierungswirkung von Nähe und Ähnlichkeit zwischen den Elementen eine korrekte Strukturierung des Bildes vornehmen (*ModelSelection*-Phase, Abschnitt 2.1.3, Seite 20). In dieser Arbeit wird jedoch ein anderes Vorgehen gewählt. Motiviert durch die Untersuchungen von Ben-Av und Sagi [Ben95], die beim Menschen eine zeitlich schnellere Wahrnehmung von Näherelationen gegenüber Ähnlichkeitsbeziehungen postulieren, wird zuerst eine Bestimmung von Näherelationen vorgenommen. Anschließend wird die Ähnlichkeit von geschlossenen Figuren ermittelt. Erst dann wird eine Detektion linearer Strukturen unter Berücksichtigung dieser zuvor bestimmten Eigenschaften durchgeführt. Mit dieser Vorgehensweise wird zunächst der Fokus auf die Beziehungen in einer Nachbarschaft von strukturbildenden Elementen gelegt, bevor dann die Detektion von Strukturen mit linearem Charakter erfolgt.

Die notwendige quantitative Formulierung der Prinzipien Nähe und Ähnlichkeit zwischen Polygonen wird in dieser Arbeit mit einem Klassifikator vorgenommen und in Abschnitt 4.4.3 beschrieben. Die hierfür eingesetzten Merkmale werden in den folgenden Abschnitten 4.4.1 und 4.4.2 vorgestellt. Die Verwendung der Klassifikationsergebnisse zur Bestimmung linearer Strukturen von Polygonen wird in Abschnitt 4.4.4 diskutiert.

4.4.1. Nähe

Das Gestaltprinzip der Nähe ist als eines der grundlegenden Prinzipien anzusehen, die eine Gruppenbildung hervorrufen (siehe auch [Low85]). Diese Annahme wird durch die Tatsache untermauert, daß der Mensch Näherelationen schneller wahrnehmen kann als Ähnlichkeitsbeziehungen und so eine Reduzierung der Komplexität einer Szenenanalyse vornimmt. Es stellt sich jedoch die Frage, was *Nähe* überhaupt bedeutet. Wie in Kapitel 4.2 vorgestellt, wird in dieser Arbeit die Nähe von Konturmodellen über eine Endpunktnähe definiert. Die Wirkung der Nähe wird dabei durch Aufmerksamkeitsbereiche festgelegt, die aus einer Stichprobe ermittelt wurden. Eine allgemeine Definition von Nähe ist jedoch nicht zu formulieren, da das Zusammenspiel der verschiedenen Gruppierungsphänomene bisher noch nicht vollständig verstanden ist.

Für die Formulierung von Nähebeziehungen zwischen polygonal approximierten geschlossenen Figuren fällt zunächst auf, daß innerhalb einer Gruppe eine paarweise Näherelation erzeugt wird. Diese Relation ist jedoch zwischen den Gruppenelementen unterschiedlich stark ausgeprägt. Erst das Zusammenwirken der verschiedenen Nähebeziehungen führt zu einer übergreifenden Gruppenbildung. Dies motiviert eine quantitative Beschreibung von Nähe zwischen Polygonen durch eine Funktion w_n , die die gruppierende Wirkung von zwei Polygonen G_1, G_2 auf eine reelle Zahl abbildet:

$$w_n : (G_1, G_2) \rightarrow \mathbb{R}$$

Das Ziel dieser Funktion ist dabei nicht eine genaue Abstandsmessung, sondern vielmehr eine angemessene Abbildung der menschlichen Wahrnehmung von Nähe zwischen Polygonen zu beschreiben. Für die im folgenden vorgestellte Quantifizierung sind verschiedene Faktoren zu berücksichtigen:

Distanz: Die geometrische Distanz von Polygonen ist ein grundlegender Baustein in der Definition einer Näherelation. Häufig wird die Distanz sogar als alleiniges Kriterium für die Beschreibung einer Nähebeziehung verwendet [Qui98].

Ausdehnung: In Analogie zu der Skalierung der Aufmerksamkeitsbereiche, die abhängig von der Länge der Konturmodelle ist, muß auch hier die Größe bzw. die Ausdehnung einer geschlossenen Figur in der gruppierenden Wirkung von Nähe berücksichtigt werden.

Nachbarschaft: Die Konstellation der Nachbarschaft ist ebenfalls von Bedeutung. Hierzu stellt sich die Frage, in welchem Maße die Gruppierungswirkung durch andere Polygone zwischen dem betrachteten Polygonpaar gestört sein?

Dichte: Ein weiterer beeinflussender Faktor in der Gruppierungswirkung ist die Dichte von Polygonen in einer lokalen Umgebung.

Auf die Berücksichtigung dieser Faktoren wird nun im Detail eingegangen.

Distanz und Ausdehnung

Ein einfache Formulierung eines Abstandsmaßes zur Beschreibung von Nähe ist der euklidische Abstand der Schwerpunkte von zwei Polygonen. Diese Definition von Nähe kommt sehr häufig zum Einsatz, ist aber für einen allgemeinen Ansatz mit beliebiger Größe und Form von Polygonen nicht aussagekräftig. Die gruppierende Wirkung von zwei kleineren Regionen ist schwächer ausgeprägt als von zwei größeren Regionen, wenn der Abstand der Schwerpunkte jeweils identisch ist (siehe Abbildung 4.12(a)). Aus dieser Anschauung heraus muß eine Abstandsdefinition erfolgen, die die Größe und auch die relative Lage von Objekten berücksichtigt (Abbildung 4.12(b)). Um die gruppierende Wirkung zwischen der Projektion zweier realer Objekte unter Berücksichtigung ihrer Größe

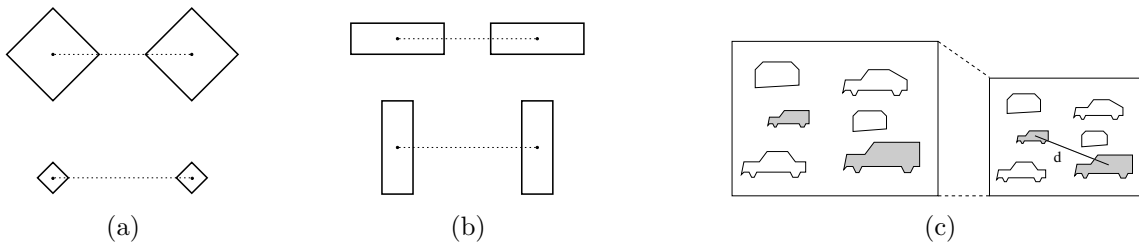


Abbildung 4.12.: Nahe Gruppierungswirkung bei einer einfachen Abstandsdefinition mit identischer Distanz: (a) ohne Berücksichtigung der Größe, (b) ohne Berücksichtigung der relativen Lage zueinander bei gleicher Größe. Die Distanz d_{com} ist gepunktet dargestellt. (c) Illustration zur Normierung einer Szene bezüglich zweier Polygone (grau schattiert dargestellt).

adäquat zu beschreiben, wird zunächst eine Normierung des betrachteten Polygonpaares bezüglich ihrer Flächen vorgeschlagen. Die Fläche wird dabei als Beschreibungsmerkmal für die Größe eines Polygons gewählt.

Bei unterschiedlichen Abbildungsmaßstäben ist die Quadratwurzel $W_i = \sqrt{A(G_i)}$ der Polygonflächen $A(G_i)$ proportional zu einem noch zu definierenden Abstand d zwischen den Polygonen. Damit ergeben sich zwei Vorgehensweisen für einen Normierungsansatz:

1. Berechne einen Abstand d zwischen zwei Polygonen und normiere d bezüglich der Polygonflächen mit W_i .
2. Normiere die Polygone jeweils mit W_i auf eine Normgröße und berechne anschließend einen Abstand d (siehe Abbildung 4.12(c)).

Unabhängig von der Vorgehensweise ist darüber hinaus zu klären, wie aus den beteiligten Polygonflächen ein gemeinsamer Normierungswert zu ermitteln ist. Hier bieten sich unter anderem folgende Definitionen an:

$$N_1 := \max\{W_1, W_2\} \quad (4.4) \qquad N_3 := \frac{W_1 + W_2}{2} \quad (4.6)$$

$$N_2 := \sqrt{\frac{W_1^2 + W_2^2}{2}} \quad (4.5) \qquad N_4 := \sqrt{W_1 W_2} \quad (4.7)$$

Bei identischen Polygonflächen ergeben alle Definitionen den gleichen Normierungswert. Unterscheiden sich die Werte für die Polygonflächen, so läßt sich eine Rangordnung angeben:

$$N_1 > N_2 > N_3 > N_4 \quad (4.8)$$

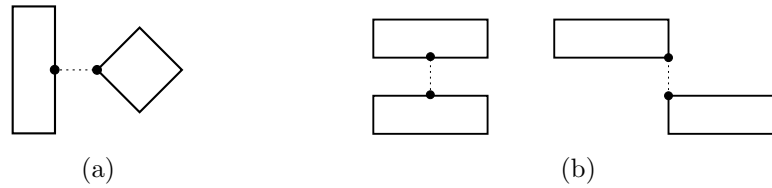


Abbildung 4.13.: Gruppierungswirkung bei der Verwendung der kürzesten Distanz: (a) einfacher Fall, (b) mit Auswirkung der relativen Lage. Die Distanz d_{short} ist gepunktet dargestellt.

Aus dieser Rangordnung ergibt sich wiederum ein Zusammenhang für eine Abstandsdefinition d_{N_i} in Abhängigkeit von N_i :

$$d_{N_1} < d_{N_2} < d_{N_3} < d_{N_4} \quad (4.9)$$

In dieser Arbeit wird die Normierungsdefinition N_4 gewählt. Eine nähere Begründung für diese Wahl wird in der Diskussion der Ergebnisse in Abschnitt 7.1 gegeben. Die folgende Diskussion zu Abstandsdefinitionen bezieht sich jeweils auf eine normierte Szene (siehe Abbildung 4.12(c)). Alle folgenden Abstandsdefinitionen setzen also voraus, daß vor der Anwendung der Definition die Szene mit einem Normierungsfaktor skaliert worden ist. Dieser Normierungsfaktor ergibt sich dabei jeweils durch die beiden betrachteten Polygone.

Wie bereits in Abbildung 4.12 veranschaulicht, sind unterschiedlichste Faktoren bei einer Abstandsdefinition zu beachten. Generell lassen sich zwei verschiedene Kategorien unterscheiden:

1. Die Distanz zwischen zwei Polygonen ergibt sich aus einer euklidischen Abstandsmessung zwischen geeigneten Referenzpunkten der Polygone.
2. Die Distanz wird durch ein Flächenmaß gegeben, das den Zwischenraum der Polygone beschreibt. Hier ist insbesondere zu beachten, daß die relative Lage der Polygone starken Einfluß auf den Distanzwert nimmt.

Für die 1. Kategorie sind folgende Definitionen untersucht worden:

Distanz der Schwerpunkte: Der Abstand zwischen zwei Polygonen ergibt sich aus der euklidischen Distanz der beiden Schwerpunkte \overline{com}_i :

$$d_{com}(G_i, G_j) := \|\overline{com}_i - \overline{com}_j\| \quad (4.10)$$

Sowohl in Abbildung 4.12(a) als auch in 4.12(b) ist zu erkennen, daß diese Abstandsdefinition nicht optimal ist. Das untere Paar erscheint jeweils wesentlich entfernter als das Obere.

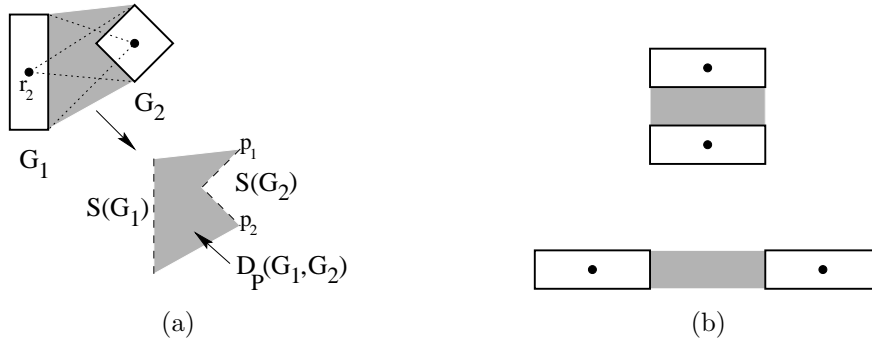


Abbildung 4.14.: Gruppierungswirkung bei der Verwendung der Punkt-Sichtbarkeit: (a) Anschauung der Definitionen, (b) problematische Konstellation, die bei gleichem Abstand $d_{vis-point}$ nicht der Gruppierungswirkung von Nähe entspricht. Die Distanz $d_{vis-point}$ ist grau schattiert und die Sichtbarkeitsabschnitte $S(G)$ sind gestrichelt markiert.

Kürzeste Distanz: Bei der Definition des kürzesten Abstands zwischen den Polygonen, wird in stärkerem Maße die relative Lage der Polygone zueinander berücksichtigt. Er ergibt sich aus den beiden nächstgelegenen Punkten \vec{p}_i auf den Polygonen:

$$d_{short}(G_i, G_j) := \min_{\substack{\vec{p}_i \in G_i, \\ \vec{p}_j \in G_j}} \|\vec{p}_i - \vec{p}_j\| \quad (4.11)$$

In Abbildung 4.13 ist dieses Abstandsmaß veranschaulicht. In Abbildung 4.13(b) ist ein Problem mit dieser euklidischen Abstandsmessung gezeigt. Beide Konstellationen haben den identischen Abstand, obwohl eine unterschiedliche Gruppierungswirkung vorliegt.

Anhand der Beispiele wird deutlich, daß bei diesen Definitionen eine Diskrepanz zwischen der Zweidimensionalität der Polygone und der Eindimensionalität der Abstandsmessung besteht. Daraus motiviert sich der Ansatz eine flächenmäßige Abstandsberechnung zu wählen, die den Bereich zwischen den Polygonen in stärkerem Maße beachtet. Für eine flächenmäßige Abstandsberechnung wurden folgende Definitionen untersucht:

Punkt-Sichtbarkeit: Die Idee des Punkt-Sichtbarkeitsabstandes gründet auf der Vorstellung, daß ein Polygon G_i von einem externen Bezugspunkt \vec{r}_i betrachtet wird (siehe Abbildung 4.14(a)). Von diesem Bezugspunkt aus ist nur ein Teil des Polygons sichtbar. Dieser Teil wird im folgenden als Punkt-Sichtbarkeitsabschnitt $S(G_i)$ bezeichnet und definiert sich durch die Verbindung der beiden äußersten von r_i sichtbaren Punkte p_1, p_2 auf dem Polygon G_i . Wird als Bezugspunkt jeweils der Schwerpunkt des anderen Polygons $\vec{r}_i = \overline{com}_j$ ($\vec{r}_j = \overline{com}_i$) gewählt, so erhält man über die jeweiligen Punkt-Sichtbarkeitsabschnitte $S(G_i)$ ($S(G_j)$)

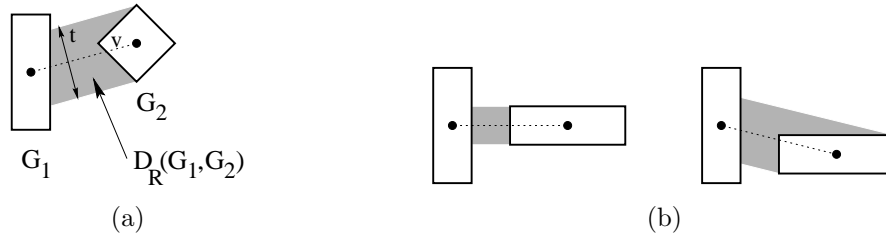


Abbildung 4.15.: Gruppierungswirkung bei der Verwendung der Richtungs-Sichtbarkeit: (a) Anschauung der Definitionen, (b) problematische Konstellation, für die bei ähnlicher Gruppierungswirkung von Nähe ein unterschiedlicher Abstand $d_{vis-dir}$ berechnet wird.

eine wechselseitige Begrenzung des Bereiches zwischen den Polygonen. Zur Definition des Abstandsmaßes zwischen zwei Polygonen G_i und G_j werden die Punkt-Sichtbarkeitsabschnitte $S(G_i)$ und $S(G_j)$ entsprechend verbunden, so daß sie den Zwischenbereich $D_P(G_i, G_j)$ von G_i und G_j umschließen (siehe Abbildung 4.14(a)). Die Distanz ist dann über die Fläche von $D_P(G_i, G_j)$ definiert:

$$d_{vis-point}(G_i, G_j) := A(D_P(G_i, G_j)) \quad (4.12)$$

Aber auch bei dieser Definition sind problematische Situationen erwarten. In Abbildung 4.14(b) ist ein Beispiel gezeigt, in dem bei beiden Konstellationen der Abstand $d_{vis-point}$ identisch aber die Gruppierungswirkung unterschiedlich ist.

Richtungs-Sichtbarkeit: Bei der Richtungs-Sichtbarkeit (Abbildung 4.15(a)) wird in Analogie zur Punkt-Sichtbarkeit vorgegangen. Hier wird jedoch der Bereich zwischen den Polygonen bezüglich seiner Breite normiert. Mit diesem Vorgehen entsteht wieder ein eindimensionales Abstandsmaß, das jedoch die Fläche zwischen den Polygonen in die Bemaßung einbezieht. Für die Berechnung wird zunächst ein Bereich $D_R(G_i, G_j)$ gewählt, dessen Begrenzung im Zwischenbereich der Polygone parallel zu der Verbindungsline v der Schwerpunkte verläuft. Die Breite t von $D_R(G_i, G_j)$ ergibt sich aus der minimalen Breite der Projektion der jeweiligen Sichtbarkeitsabschnitte auf eine zu v orthogonal aufgespannte Ebene (siehe Abbildung 4.15(a)). Mit dieser Definition läßt sich dann folgende Abstandsdefinition angeben:

$$d_{vis-dir}(G_i, G_j) := \frac{A(D_R(G_i, G_j))}{t} \quad (4.13)$$

Doch diese Definition liefert ebenfalls keine adäquate Beschreibung einer Gruppierungswirkung für Nähe. In Abbildung 4.15(b) werden beide Anordnungen als ähnlich nah empfunden, jedoch sind die Werte für $d_{vis-dir}$ signifikant unterschiedlich.

Bezüglich der verschiedenen Abstandsdefinitionen kann gesagt werden, daß keines der untersuchten Merkmale allein eine adäquate Beschreibung der Gruppierungsphänomene bei nahen Polygonen liefert. Daher ist es sinnvoll eine Kombination der verschiedenen Merkmale zu verwenden. Bevor diese Kombination jedoch in Abschnitt 4.4.3 erläutert wird, werden weitere Faktoren für die Beschreibung einer Gruppierungswirkung aufgrund von Nähe untersucht.

Nachbarschaft

Für eine Diskussion eines reinen Abstandsmaßes ist die Konzentration auf die Untersuchung *eines* Polygonpaares ausreichend. Jedoch ist in Bezug auf eine Nähewirkung auch die Umgebung der Polygone zu berücksichtigen. Insbesondere wenn zwischen zwei Polygonen noch weitere Polygone lokalisiert sind, nimmt die Gruppierungswirkung mit der Zunahme einer solchen Störung graduell ab. Doch wie kann eine solche graduelle Reduzierung der Gruppierungswirkung modelliert werden?

In Wachsmuth et al. [Wac00] wird eine ähnliche Problematik bei der Untersuchung von Sprache–Bildinteraktionen aufgeworfen. Sprachlich referenzierte Objekte („Nimm die Schraube rechts von der Leiste.“) müssen über Nachbarschaftsrelationen zwischen den Objekten im Bild identifiziert werden. In [Wac00] wird hierzu die Separierung eingeführt, die zu einer graduellen Nachbarschaftsbeziehung zwischen Projektionen von Objekten führt: Zwei Objektprojektionen bzw. deren polygonale Beschreibungen können durch eine dritte Objektprojektion vollständig oder graduell voneinander separiert sein. Dabei ist die Separierung abhängig von der Position und Ausdehnung der dritten Objektprojektion. Technisch wird dieses Problem mit folgendem Ansatz gelöst: Es wird ein Bereich zwischen den beiden Polygonen bestimmt, der an die Richtungs-Sichtbarkeit aus der Abstandsdiskussion angelehnt ist. Die Richtung wird dabei jedoch durch die kürzeste Distanz zwischen den Polygonen vorgegeben (siehe Abbildung 4.16(a)). Anschließend werden alle Objektprojektionen bestimmt, die diesen Bereich mit ihren Polygonen G_k stören. In Abhängigkeit von ihrer Überlappung mit dem Bereich $D_{sR}(G_i, G_j)$, reduzieren sie diesen Bereich um $s(G_k)$ orthogonal zu seinem Richtungsvektor $v(G_i, G_j)$ (siehe Abbildung 4.16(a)). Mit dieser Anschauung ergibt sich eine graduelle Störung der Nachbarschaft aus dem Verhältnis der ursprünglichen Breite t von D_{sR} und dem Anteil $s(G_k)$ in diesem Bereich. Werden t und $s(G_k)$ als Intervalle über D_{sR} orthogonal zu v aufgefaßt, so ergibt sich folgende Definition für die Störung der Nachbarschaft:

$$\eta_1(G_i, G_j) := \frac{\left| \bigcup_{\forall G_k: G_k \cap D_{sR} \neq \emptyset} s(G_k) \right|}{|t|} \quad (4.14)$$

Allerdings läßt sich schnell ein Beispiel konstruieren, in dem die gewählte Definition keine adäquate Beschreibung der Nachbarschaftsstörung liefert (siehe Abbildung 4.16(b)). In dieser Konstellation reduziert sich die Breite t von D_{sR} auf einen sehr kleinen Wert, wodurch plötzlich eine vollständige bzw. überhaupt keine Separierung durch andere Ob-

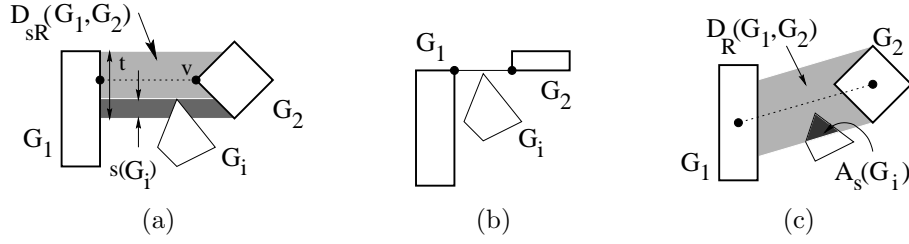


Abbildung 4.16.: Gruppierungswirkung bei Störung der Nachbarschaft: (a) Anschauung der Definition (4.14), (b) problematische Konstellation für Definition (4.14), und (c) Anschauung der Definitionen (4.15) und (4.16).

jektpolygone erfolgen kann, die offensichtlich einen Einfluß auf die Gruppierungswirkung besitzen.

Aus diesem Grund wird dieser Ansatz nicht weiter betrachtet, und es werden zwei Definitionen für eine graduelle Störung der Nachbarschaft bezüglich des Bereiches D_R aus Formel (4.13) untersucht. Eine Überlappung von G_k mit dem Bereich D_R muß nicht notwendigerweise eine vollständige Reduzierung von D_R um $s(G_k)$ orthogonal zu v hervorrufen, sondern es sollte die Ausdehnung von G_k parallel zu v berücksichtigt werden. Ein Objekt, welches nur mit einer Spitze in den Bereich D_R hineinragt, hat eine weniger starke Reduzierung der Gruppierungswirkung zur Folge, als ein Objekt, was nahezu über die gesamte Länge von D_R und genauso tief in diesen Bereich hineinragt. Mit dieser Vorstellung ergibt sich eine graduelle Nachbarschaftsstörung aus dem Verhältnis der Fläche von D_R zur anteiligen Fläche A_s des separierenden Objektes, die innerhalb von D_R liegt (siehe Abbildung 4.16(c)):

$$\eta_2(G_i, G_j) := \frac{A_s \left(\bigcup_{\forall G_k} (G_k \cap D_R) \right)}{A(D_R)} \quad (4.15)$$

Doch auch bei dieser Definition läßt sich eine Problematik feststellen. Der Störungsgrad der Nachbarschaft ist bei dieser Definition vom Abstand der Polygone abhängig: Werden zwei Polygone, bei gleichbleibender Störung durch ein anderes Objekt, weiter entfernt voneinander plziert, so nimmt mit der Distanz auch der Grad der Störung ab.

Hieraus ergibt sich das Ziel, eine Formulierung von gradueller Nachbarschaftsstörung unabhängig von dem aktuellen Abstand der Polygone zu erhalten. Dieses läßt sich mit folgender Definition erreichen:

$$\eta_3(G_i, G_j) := \frac{A_s \left(\bigcup_{\forall G_k} (G_k \cap D_R) \right)}{\frac{1}{2}(A(G_i) + A(G_j))} \quad (4.16)$$

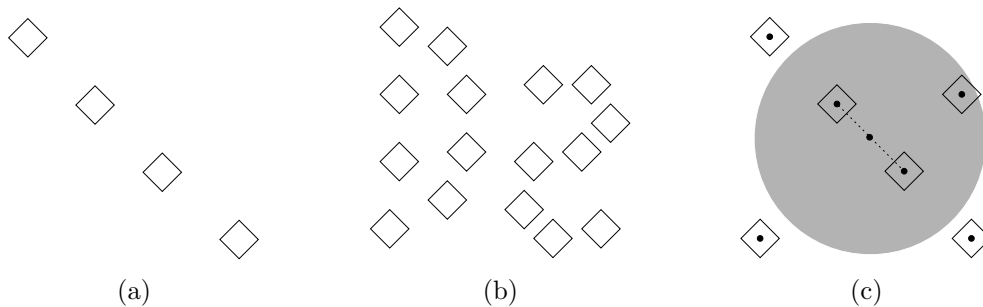


Abbildung 4.17.: Gruppierungswirkung in Abhängigkeit des Abstandes und der Dichte im lokalen Umfeld: (a) Einfache Struktur, (b) komplexere Strukturen, in denen die Struktur aus (a) enthalten ist, und (c) Bereich der lokalen Umgebung (grau schattiert) zur Bestimmung der Dichte.

Der Störungsgrad der Nachbarschaft ist damit abhängig von dem Verhältnis der Gesamtfläche der separierenden Polygone zur mittleren Fläche des betrachteten Polygonpaares. Somit gelten zwei Polygone als vollständig separiert, falls der Mittelwert ihrer Flächen kleiner als die Gesamtfläche der separierenden Polygone ist. So wird implizit auch die Ausdehnung der Polygone berücksichtigt, und eine adäquate Beschreibung für eine graduelle Nachbarschaftstörung steht zur Verfügung.

Dichte

Neben der Nachbarschaft wird nun als letzter Faktor zur Beurteilung einer Gruppierungswirkung aufgrund von Nähe die Dichte von Polygonen in einem lokalen Umfeld untersucht. Hierzu wird ein Ansatz in Anlehnung an Lowe [Low85] gewählt, der sich aus dem Prinzip des ‚*non-accidentalness*‘ (Kapitel 3.1.1) motiviert.

Zunächst wird die Projektion von unabhängig positionierten Objekten aus einer Szene betrachtet (siehe Abbildung 4.17(a),(b)). Die einzelnen Objektprojektionen seien durch Polygone in der Projektionsebene repräsentiert. Nach Lowe nimmt die Wahrscheinlichkeit, daß eine Gruppierungswirkung für ein Polygonpaar keinen kausalen Zusammenhang hat, mit der Nähe der Polygone und in Abhängigkeit der Polygondichte in einem lokalen Umfeld ab. Für eine Beurteilung der Gruppierungswirkung eines Polygonpaares muß also neben dem Abstand die Polygondichte berücksichtigt werden. Diese Polygondichte ergibt sich aus der Anzahl der Polygone pro Flächeneinheit, wobei von Lowe eine kreisförmige Umgebung um ein Polygonpaar vorgeschlagen wird, die einen Durchmesser vom zwei- bis dreifachen des Polygonabstandes besitzt. Diese Wahl ist eher als pragmatisch anzusehen, da er annimmt, daß der Mensch erst nach einer Art Texturanalyse des gesamten Bildes eine Aussage über die lokalen Dichten trifft. Dieser Vorschlag wird hier übernommen, da die Vorgaben für die Bestimmung eines Aufmerksamkeitsbereiches aus einer Stichprobe, wie in Abschnitt 4.2.1 beschrieben, nur

sehr schwer konkret formuliert werden können. Es wird daher ein Kreis um den Mittelpunkt der Verbindungslinie v der Polygonschwerpunkte als lokale Umgebung gewählt. Der Durchmesser dieses Kreises ist durch das zweieinhalbfache der Länge von v gegeben (siehe Abbildung 4.17(c)). Alle Polygone, die mit ihrem Schwerpunkt \overrightarrow{com}_i innerhalb dieses Kreises $K(G_1, G_2)$ liegen, werden zur Berechnung der Dichte ϱ herangezogen:

$$\varrho(G_i, G_j) := \frac{|\{\overrightarrow{com}_k \mid \overrightarrow{com}_k \in K(G_i, G_j)\}|}{A(K(G_i, G_j))} \quad (4.17)$$

Die Dichte berechnet sich also aus dem Verhältnis der Anzahl der Polygone innerhalb des Kreises und der Fläche des Kreises. Sie wird ebenfalls als ein Merkmal für die Quantifizierung von Nähe aufgefaßt.

Mit der Dichte liegen alle Merkmale vor, die für die Berechnung einer adäquaten Beschreibung der Gruppierungswirkung aufgrund von Nähe untersucht wurden. Wie bereits erwähnt, ist die Einschränkung der zulässigen Merkmalswerte durch empirisch bestimmte Schwellwerte nicht eindeutig möglich, was den Einsatz eines Klassifikators zur Unterscheidung zwischen *nah* und *fern* motiviert. Dieser Klassifikationsansatz zur quantitativen Formulierung der Gruppierungswirkung wird in Abschnitt 4.4.3 vorgestellt. Da für die Ähnlichkeit von Polygonen ein analoger Ansatz über einen Klassifikator gewählt wird, werden im folgenden Abschnitt zunächst die Merkmale zur Berechnung einer Ähnlichkeit diskutiert, bevor die Klassifikation beider Eigenschaften vorgestellt wird.

4.4.2. Ähnlichkeit

Eine Gruppierungswirkung aufgrund von ähnlichen Eigenschaften eines Objektes ist besonders häufig in der vom Menschen beeinflussten Umgebung zu beobachten. Gerade die Regelmäßigkeit von künstlichen Objekten führt zu einer Gruppenbildung aufgrund von ähnlich regelmäßigen Formen oder der ähnlichen Oberflächenerscheinung (z.B. Farbe, Textur) der Objekte. Für den hier verfolgten Ansatz wird nur die Form einer Objektprojektion auf Ähnlichkeit untersucht. Dies bedingt sich durch den Einsatz der Gruppierungen der **2D**-Ebene der Gruppierungshierarchie, die keine Information über das Erscheinungsbild der eingeschlossenen Fläche enthalten. Ziel ist es, analog zu dem Begriff der Nähe, eine quantitative Formulierung der Gruppierungswirkung von Polygonen G_1, G_2 aufgrund von Ähnlichkeit zu erhalten. Diese quantitative Formulierung soll den Grad der Ähnlichkeit auf eine kontinuierliche Zahl w_s abbilden:

$$w_s : (G_1, G_2) \rightarrow \mathbb{R}.$$

Ein häufiges Anwendungsgebiet für die Detektion ähnlicher Objektformen, ist die Werkstückerkennung im industriellen Umfeld, um einen Sortierungsprozeß oder eine Roboterpositionierung zu realisieren. Hierbei kommen jedoch zumeist bekannte Objektformen zum Einsatz, die im gegebenen Bild wiedergefunden werden müssen. Für diese und allgemeinere Ansätze lassen sich zwei Kategorien zur Ähnlichkeitsbestimmung von Polygonen unterscheiden:

Geometrischer Vergleich: Die geometrische Figur der Polygone wird auf ähnliche Eigenschaften untersucht. Die Untersuchung erfolgt damit im zweidimensionalen euklidischen Raum.

Funktionaler Vergleich: Die Polygone werden durch eine parametrisierte Funktion beschrieben und der Vergleich ergibt sich aus einem Abstandsmaß über den Funktionsverlauf in dem durch die Parameter aufgespannten Raum.

Auf diese Kategorien wird im folgenden im Detail eingegangen. Wie bei der Bestimmung von Nähe erfolgt auch hier eine paarweise Betrachtung der Polygone. Einen guten allgemeinen Überblick über verschiedene Methoden zur Bestimmung der Ähnlichkeit von Polygonen liefern Alt et al. in [Alt99].

Geometrischer Vergleich

Ein häufiges Vorgehen bei einem geometrischen Vergleich läßt sich folgendermaßen beschreiben.

1. Beide Polygone werden durch die Anwendung von Transformationen (Skalierung, Translation, Rotation) überlagert.
2. Es wird ein Abstand zwischen den überlagerten Polygonen berechnet.

Diese Schritte können mehrmals iteriert werden, bis ein minimaler Abstand und damit eine optimale Deckung zwischen den Polygonen gefunden ist. Es ist leicht nachzuvollziehen, daß die genaue Bestimmung eines minimalen Abstandes zu einem sehr aufwendigen Verfahren führt. In [Che93b] wird beispielsweise ein Algorithmus mit parametrischer Suche angegeben, der eine Komplexität von $O((nm)^3 \log^2(nm))$ bei m bzw. n Polygonstützpunkten aufweist. Aufgrund dieser Problematik wird häufig auf die Bestimmung eines minimalen Abstandes verzichtet und eine gute Approximation einer optimalen Lösung gesucht. Auch in dieser Arbeit wird auf das iterative Vorgehen verzichtet und eine einmalige Überlagerung zur Abstandsmessung durchgeführt.

Mit einer Referenzpunktmethode wird hier das Ziel verfolgt, zwei Polygone G_1 und G_2 möglichst gut zur Deckung zu bringen. Hierzu kommt ein Satz von Transformationen zum Einsatz:

Skalierung: Bevor zwei Polygone auf Ähnlichkeit untersucht werden, wird zunächst eine Skalierung jedes Polygons auf eine Normgröße durchgeführt. Der Normierungsfaktor N_6 wird hier auf Basis der Momente des Polygons gewählt (zur Berechnung von Momenten siehe [Bal82]):

$$N_6 := \sqrt{m_{02}(G) + m_{20}(G)}$$

Translation: Um die Polygone anschließend zur Deckung zu bringen wird eine Translation des Polygons G_1 bezüglich eines Referenzpunktes $\vec{r}(G_1)$ auf den entsprechenden Referenzpunkt $\vec{r}(G_2)$ des Polygons G_2 durchgeführt. In dieser Arbeit wird der Flächenschwerpunkt eines Polygons als Referenzpunkt gewählt, wie auch in [dB96] vorgeschlagen. Allerdings können sich in Spezialfällen andere Referenzpunkte als besser geeignet erweisen. Eine ausführliche Diskussion zur Wahl des Referenzpunktes findet sich in [Alt94].

Rotation: Vor der eigentlichen Abstandsberechnung wird abschließend eine Rotation um den Referenzpunkt mit dem Winkel der Hauptachsen der Polygone durchgeführt. Bei geringer Exzentrizität des Polygons wird statt der Hauptachse die Verbindungslinie des Schwerpunktes mit dem äußersten Stützpunkt des Polygons als Trägheitsachse des Polygons gewählt.

Mit diesen Schritten werden Polygone soweit zur Deckung gebracht, daß eine Abstandsmessung unter Verwendung verschiedener Methoden erfolgen kann.

Das Ziel einer Abstandsbestimmung zwischen zwei Polygonen soll eine quantitative Formulierung der Gruppierungswirkung aufgrund von Ähnlichkeit sein. Im Idealfall liefert die gewählte Abstandsdefinition genau die graduelle Interpretation dieser Gruppierungswirkung. Mit dieser Zielsetzung sind folgende Definitionen zwischen zwei zur Deckung gebrachten Polygonen untersucht worden:

Hausdorffdistanz: Für die *Hausdorffdistanz* wird die Menge aller Stützpunkte $\vec{p}_i \in G_i$ beider Polygone verwendet. Hinzu kommt jeweils die Punktmenge P_{ij} , die die Fußpunkte von senkrechten Projektionen der Punkte $\vec{p}_i \in G_i$ auf das Polygon G_j enthält. Die Berechnung untergliedert sich dann in die Bestimmung zweier Werte. Die einseitige *Hausdorffdistanz* ist definiert durch:

$$\tilde{d}_{haus}(G_i, G_j) := \max_{\vec{p}_i \in G_i} \min_{\vec{p}_j \in G_j \cup P_{ij}} \|\vec{p}_i - \vec{p}_j\|$$

Dieser Wert beschreibt den Maximalwert des geringsten Abstandes bei der Projektion von Punkten $\vec{p}_i \in G_i$ auf G_j . Hierbei ist zu beachten, daß diese Abstandsdefinition richtungsabhängig ist und daher die (beidseitige) Hausdorffdistanz zu verwenden ist:

$$d_{haus}(G_i, G_j) := \max(\tilde{d}_{haus}(G_i, G_j), \tilde{d}_{haus}(G_j, G_i)) \quad (4.18)$$

Mit der Hausdorffdistanz wird also eine obere Schranke für den maximal möglichen Abstand von Polygonzügen angegeben, wenn wechselseitig die Punkte über die kürzeste Entfernung projiziert werden (siehe Abbildung 4.18(a)). Der Berechnungsaufwand beträgt mit einem Algorithmus nach [HA93] $O((n+m) \log(n+m))$.

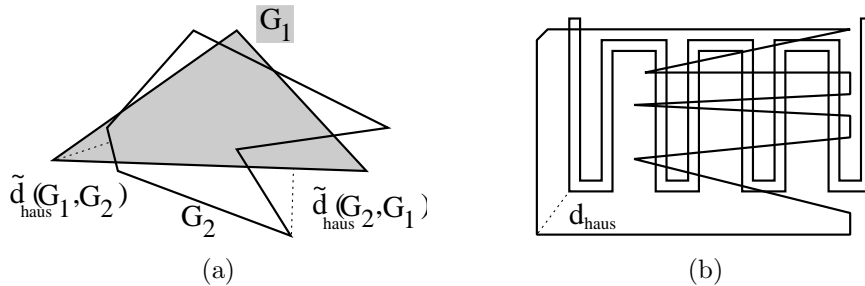


Abbildung 4.18.: Hausdorffdistanz: (a) Die einseitige Hausdorffdistanz ist durch gepunktete Linien markiert, wobei $\tilde{d}_{haus}(G_1, G_2) \neq \tilde{d}_{haus}(G_2, G_1)$. (b) Bei einer relativ kleinen Hausdorffdistanz d_{haus} ist trotzdem keine Ähnlichkeit zwischen den Polygonen vorhanden.

Die Hausdorffdistanz hat den entscheidenden Nachteil, daß sie den Verlauf der Polygone nicht berücksichtigt, und die gesuchten Extremwerte nur an den Stützpunkten eines der Polygone auftreten können. So werden kleine Werte ermittelt, obwohl zwei Polygone keine ähnliche Form besitzen (siehe Abbildung 4.18(b)).

Fréchetabstand: Mit dem *Fréchetabstand* wird die Problematik der Hausdorffdistanz aufgehoben. Er berücksichtigt den Verlauf der Polygone G_1, G_2 und ist durch folgende Idee motiviert:

Ein Mann führt seinen Hund an der Leine spazieren, wobei der Mann auf dem Weg G_1 geht und der Hund dem Weg G_2 folgt. Ihre Geschwindigkeit können beide frei wählen, sie können auch zeitweilig stehenbleiben, dürfen aber nie auf ihrem eigenen Weg zurückgehen. Der Fréchetabstand entspricht dann der kürzesten Länge der Hundeleine, die für diesen Spaziergang auf G_1 und G_2 erforderlich ist.

Der Aufwand für die Berechnung des Fréchetabstandes weist eine Komplexität von $O(mn \log^3(mn))$ auf. In einer Testphase konnten in den Ergebnissen keine gravierenden Unterschiede zur Hausdorffdistanz festgestellt werden. Deshalb wurde der Fréchetabstand aufgrund des höheren Berechnungsaufwandes nicht weiter betrachtet. Ein ausführliche Diskussion des Fréchetabstandes findet sich in [Alt95].

Symmetrische Differenz: Als ein weiteres Merkmal für einen geometrischen Vergleich wird die *symmetrische Differenz* untersucht. Bisher wurde das Ergebnis eines Vergleiches durch *einen* Maximalwert bezüglich eines Abstandsmaßes beschrieben. Nun wird der Vergleich auf der Basis eines Mengenbegriffs definiert:

$$d_{sym}(G_i, G_j) := A(G_i \cup G_j) - A(G_i \cap G_j) \quad (4.19)$$

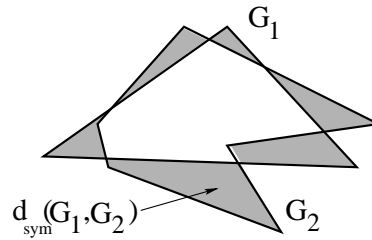


Abbildung 4.19.: Symmetrische Differenz

Die symmetrische Differenz beschreibt die Fläche der Polygone G_i und G_j , in denen sie sich nicht überdecken (siehe Abbildung 4.19). Inwieweit die symmetrische Differenz die Ähnlichkeitswahrnehmung beim Menschen widerspiegelt, ist an Beispielen schwer zu beurteilen.

Neben diesen speziellen Merkmalen werden für einen geometrischen Vergleich auch die Unterschiede bezüglich der Exzentrizität und der Kompaktheit der Polygone betrachtet:

$$d_{ecc}(G_i, G_j) := |eccentricity(G_i) - eccentricity(G_j)| \quad (4.20)$$

$$d_{compact}(G_i, G_j) := |compactness(G_i) - compactness(G_j)| \quad (4.21)$$

Zur Berechnung dieser allgemeinen Formmerkmale sei auf [Bal82] verwiesen.

Funktionaler Vergleich

Neben der rein geometrischen Betrachtung von Polygonen läßt sich auch ein Vergleich über eine parametrisierte Funktion anstellen, die den Verlauf der Polygone abbildet. Der Abstand solcher Funktionen über eine Funktionsraum-Metrik liefert dann ein Maß für das ähnliche Erscheinungsbild von Polygonen. Wird zur Abstandsberechnung die L_p -Metrik für zwei Funktionen $f : [a, b] \rightarrow \mathbb{R}$ und $g : [a, b] \rightarrow \mathbb{R}$ eingesetzt, so ergibt sich der Abstandswert zu:

$$\|f - g\|_p := \left(\int_a^b |f(x) - g(x)|^p dx \right)^{\frac{1}{p}}$$

Turning-Funktion: In dieser Arbeit wurde die von Arkin et al. [Ark91] verwendete *Turning-Funktion* zur funktionalen Beschreibung von Polygonen eingesetzt. Die Turning-Funktion $\Theta_G(b)$ beschreibt den Verlauf eines Polygons G über die Orientierung der Tangente ϕ in jedem Punkt $\vec{p} \in G$ in Abhängigkeit von der Bogenlänge

$$b : (\vec{p} \in G) \rightarrow [0, 1].$$

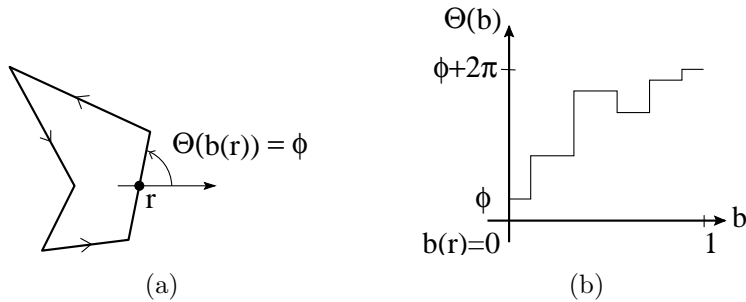


Abbildung 4.20.: Anschauung zur Definition der Turning-Funktion. (a) Beispielpolygon mit dem Referenzpunkt \vec{r} . (b) Graph der Turning-Funktion für das Polygon in (a) in Abhängigkeit von der Bogenlänge b .

Die Bogenlänge², oder auch Weglänge von G , bildet die Punkte auf G bezüglich eines Referenzpunktes $\vec{r} \in G$ auf ein normiertes Intervall ab. Zur Vereinfachung der Notation wird im folgenden für Werte von $b(\vec{p})$ nur b angegeben. Die Orientierung ϕ der Tangenten wird im mathematischen Sinn bezüglich der x-Achse angenommen (siehe Abbildung 4.20).

Bevor nun der Ähnlichkeitsabstand von Polygonen über die Turning-Funktion bestimmt wird, lassen sich noch folgende Eigenschaften von Θ_G angeben:

- a) Für jedes beliebige $s \in [0, 1]$ gilt: $\Theta_G(s + 1) = \Theta_G(s) + 2\pi$.
- b) Θ_G besitzt Sprungstellen in den Stützpunkten von G .
- c) Θ_G ist durch b translations- und skalierungsinvariant.
- d) Eine Rotation θ von G entspricht einer Verschiebung von Θ_G entlang der y-Achse: $\Theta_G(b) + \theta$
- e) Eine Verschiebung des Referenzpunktes um $t \in [0, 1]$ entspricht einer Verschiebung von Θ_G entlang der x-Achse.

Für den Vergleich von zwei Polygonen G_1 und G_2 durch ihre Turning-Funktionen sind den Punkten d) und e) besondere Beachtung zu schenken. Das Ziel, den minimalen Abstand der Turning-Funktionen zu ermitteln, ist nur von den Parametern θ und t abhängig:

$$d_{turn} := \left(\min_{\substack{\theta \in [0, 2\pi], \\ t \in [0, 1]}} \int_0^1 |\Theta_{G_1}(b+t) + \theta - \Theta_{G_2}(b)|^p db \right)^{\frac{1}{p}} \quad (4.22)$$

²Die Bogenlänge bezieht sich hier auf den Umfang $U(G)$ und nicht auf ein *Bogenmaß* in Abhängigkeit von einem Winkel.

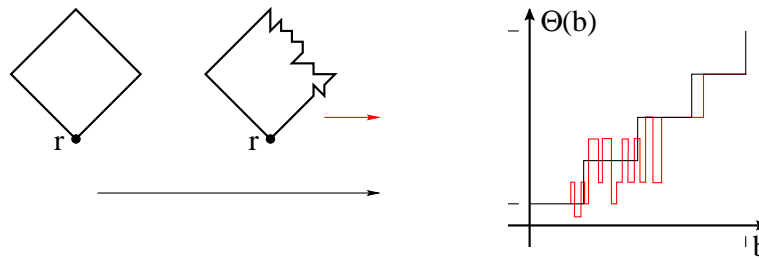


Abbildung 4.21.: Problemfall der Turning-Funktion. Der Abstand ist trotz weitgehend ähnlicher Form sehr groß.

Unter Verwendung der L_2 -Metrik zeigt Arkin et al., daß für jedes beliebige t ein optimales θ einfach zu bestimmen ist:

$$\theta^*(t) = \int_0^1 \Theta_{G_1}(b) db - \int_0^1 \Theta_{G_2}(b) db - 2\pi t. \quad (4.23)$$

Damit ergibt sich der minimale Abstand von zwei Turning-Funktionen Θ_{G_1} und Θ_{G_2} zu:

$$d_{turn} := \left(\min_{t \in [0,1]} \int_0^1 |\Theta_{G_1}(b+t) + \theta^*(t) - \Theta_{G_2}(b)|^2 db \right)^{\frac{1}{2}} \quad (4.24)$$

Aufgrund des stückweise linearen Verlaufs des zu minimierenden Terms von Formel (4.24) kann der minimale Wert von d_{turn} nur an den Stützpunkten der Polygone auftreten (siehe [Ark91]). Daraus ergibt sich ein Rechenaufwand für die Bestimmung des Polygonabstandes über Turning-Funktionen von $O(mn \log(mn))$, bei m, n Stützpunkten.

Ein entscheidender Nachteil der Turning-Funktionen ist die Anfälligkeit für ein nichtgleichverteiltes Rauschen auf den Polygonzügen. In Abbildung 4.21 ist der Umfang der beiden Figuren durch das Rauschen in dem rechten Abschnitt der rechten Figur deutlich größer. Die Abbildung des Umfangs auf $[0, 1]$ führt zu stark unterschiedlichen Turning-Funktionen, in denen gleichförmige Abschnitte der Figuren eine unterschiedliche Länge aufweisen. Dadurch entstehen große Abweichungen, und eine Ähnlichkeit über die funktionale Darstellung ist nicht mehr zu erkennen.

Mit der Turning-Funktion sind alle Merkmale diskutiert, die für die Berechnung einer Ähnlichkeit eingesetzt werden. Auch hier konnte keine eindeutige Definition vorgestellt werden, die die Gruppierungswahrnehmung des Menschen für ähnliche Formen widerspiegeln kann. Im nächsten Abschnitt wird deshalb der verwendete Klassifikationsansatz für eine quantitative Formulierung von Nähe und Ähnlichkeit zwischen Polygonen erläutert.

4.4.3. Klassifikation und Kombination

Für alle bisher vorgestellten Merkmale zur Beschreibung einer Gruppierungswirkung lassen sich Beispiele aufzeigen, die den wahrgenommenen Gruppierungsphänomenen zwischen Polygonen widersprechen. Um nun zu einer adäquaten quantitativen Formulierung der Gruppierungswirkung zu gelangen, werden die verschiedenen Faktoren jeweils in einem Merkmalsvektor für Nähe und Ähnlichkeit zusammengefaßt. Dies motiviert sich durch die Idee, daß die Kombination der Merkmale zu einer Kompensation der individuellen Schwierigkeiten führt. Für die Beurteilung der Gruppierungswirkung von zwei Polygonen aufgrund von Nähe oder Ähnlichkeit wird jeder der beiden Merkmalsvektoren mit einem entsprechend trainierten Polynomklassifikator klassifiziert. Für die Quantifizierung der Gruppierungswirkung wird dabei das Ergebnis der Klassifizierung herangezogen.

Der Aufbau eines Klassifikationssystems im allgemeinen, sowie auch die Theorie des Polynomklassifikators im speziellen, wird hier nicht weiter erläutert. Für eine ausführliche Darstellung dieser Thematiken sei auf [Nie83, Sch96] verwiesen.

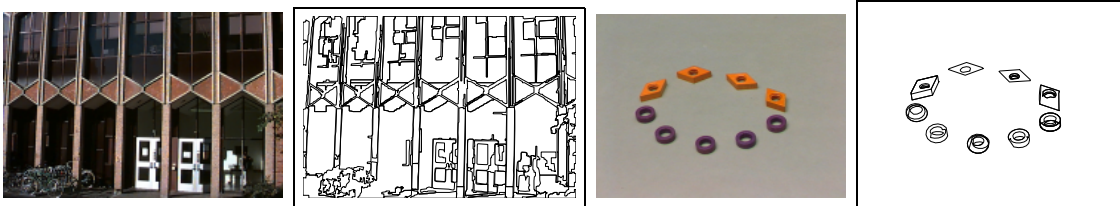


Abbildung 4.22.: Beispielbilder der Stichprobe, die zum Training von naher und ähnlicher Gruppierungswirkung verwendet wurden. Links: Polygone als Ergebnis einer Regionensegmentierung mit dem Island Growing. Rechts: Geschlossene Figuren aus dem konturbasierten Gruppieren.

Stichprobe: Für den Ansatz wurden Trainingsstichproben mit speziellen Objektkonstellationen verwendet. Für die Nähwirkung wurden die Regionen aus 12 Bildern und für Ähnlichkeit aus 8 Bildern manuell gruppiert. Dazu hat eine Versuchsperson eine Klassifizierung der Gruppierungswirkung zwischen den Objektpolygonen in zwei Klassen vorgenommen. In [Abbildung 4.22](#) sind zwei Bilder dargestellt, die in beiden Stichproben verwendet wurden. Betrachtet wurden nur paarweise disjunkte Polygone. Die Klasseneinteilung wurde wie folgt gewählt:

$$\Omega_{proximity} = \{\omega_{near}, \omega_{distant}\} \quad \Omega_{similarity} = \{\omega_{similar}, \omega_{distinct}\}$$

Diese Aufteilung entspricht der Interpretation, daß ‚Polygone gruppieren‘ (ω_{near} bzw. $\omega_{similar}$) oder ‚Polygone nicht gruppieren‘ ($\omega_{distant}$ bzw. $\omega_{distinct}$). Die Einführung einer weiteren Klasse ‚Entscheidung nicht möglich‘ wurde untersucht, jedoch

war diese Klasse in der Stichprobe stark unterrepräsentiert. Auch die Abstufung in graduelle Beurteilungen von Nähe und Ähnlichkeit durch eine höhere Anzahl von Klassen wurde in Betracht gezogen, aber aufgrund der deutlich schwierigeren Interpretation für die Versuchsperson verworfen.

Merkmalsvektoren: Nach der Markierung durch die Versuchsperson wurde aus den Stichproben für jedes klassifizierte Polygonpaar (G_i, G_j) ein entsprechender Merkmalsvektor mit folgenden Komponenten ermittelt:

$$\mathbf{c}_{proximity} = \begin{pmatrix} d_{com} \\ d_{short} \\ d_{vis-point} \\ d_{vis-dir} \\ \eta_3 \\ \varrho \end{pmatrix} \quad \mathbf{c}_{similarity} = \begin{pmatrix} d_{haus} \\ d_{sym} \\ d_{turn} \\ d_{ecc} \\ d_{compact} \end{pmatrix}$$

Diese Merkmalsvektoren dienten als Trainingsmenge für den jeweiligen Klassifikator. Der Grad des verwendeten Polynoms ist abhängig von der Anzahl der korrekt klassifizierten Stichprobenelemente einer unabhängigen Teststichprobe und wird hier zu drei gewählt. Diese Wahl wird in Zusammenhang mit den Ergebnissen in Abschnitt 7.1 diskutiert.

Zur quantitativen Formulierung von Nähe und Ähnlichkeit zwischen zwei Polygonen wird das Klassifikationsergebnis für einen Merkmalsvektor genauer untersucht: Generell ordnet ein Klassifikator einen Merkmalsvektor einer der trainierten Klassen aufgrund der Entscheidungsregel zu. Betrachtet man die hierzu verwendete Unterscheidungsfunktion d eines Polynomklassifikators, so gilt:

$$\sum_{\omega_i \in \Omega} d_{\omega_i}(\mathbf{c}) = 1$$

Dabei können die einzelnen d_{ω_i} nicht im Sinne einer genauen Wahrscheinlichkeit für die Klassenzugehörigkeit zu ω_i gedeutet werden, auch wenn die Werte sich meist im Intervall $[0, 1]$ bewegen. Es ergibt sich vielmehr folgende konkrete Situation: Bei einem Zweiklassenproblem erfolgt eine sichere Entscheidung für die Klasse ω_i , falls $d_{\omega_i}(\mathbf{c}) = 1 \pm \epsilon$ mit $\epsilon \approx 0.35$ gilt. Liegen die Werte $d_{\omega_i}(\mathbf{c})$ für alle Klassen ω_i hingegen nicht in dem Intervall $[1.0 - \epsilon, 1.0 + \epsilon]$, so trifft ein Klassifikator eine eher unsichere Entscheidung mit

$$\omega = \underset{\omega_i}{\operatorname{argmax}} d_{\omega_i}. \quad (4.25)$$

Mit dieser Charakteristik können die Komponenten der Unterscheidungsfunktion für eine quantitative Interpretation der Gruppierungswirkung eingesetzt werden. Dabei sei nochmal ausdrücklich darauf hingewiesen, daß die Komponenten d_{ω_i} hier *nicht* als Wahrscheinlichkeit für die Klassenzugehörigkeit angesehen werden, wohl aber im Sinne eines Bewertungsmaßes, auch bei einer unsicheren Entscheidung, interpretiert werden können:

graduelle Nähe: Das Ziel, eine adäquate Formulierung der wahrgenommenen Nähe von Polygonen zu erhalten, entspricht der Berechnung eines kontinuierlichen Wertes, der als Stärke für die Gruppierungswirkung interpretiert werden kann. Diese Formulierung läßt sich mit der Unterscheidungsfunktion wie folgt abbilden:

$$w_n(G_i, G_j) := \begin{cases} d_{\omega_{near}}(\mathbf{c}_{proximity}) & , \text{ falls } 0 \leq d_{\omega_{near}} \leq 1 \\ 1 & , \text{ falls } d_{\omega_{near}} > 1 \\ 0 & , \text{ sonst} \end{cases} \quad (4.26)$$

graduelle Ähnlichkeit: Für die quantitative Formulierung von Ähnlichkeit wird eine zur Nähe analoge Abbildung gewählt:

$$w_s(G_i, G_j) := \begin{cases} d_{\omega_{similar}}(\mathbf{c}_{similarity}) & , \text{ falls } 0 \leq d_{\omega_{similar}} \leq 1 \\ 1 & , \text{ falls } d_{\omega_{similar}} > 1 \\ 0 & , \text{ sonst} \end{cases} \quad (4.27)$$

Es erweist sich als günstig, für eine Gruppierung von nahen und ähnlichen Regionen in eine Struktur mit linearem Charakter, über die individuelle Beurteilung hinaus eine Kombination der graduellen Interpretation zu verwenden. Für diese Kombination wird zusätzlich eine Möglichkeit der Gewichtung $\kappa_{sn} \in [0, 1]$ der individuellen Gruppierungswirkung gewünscht. Somit kann gegebenenfalls das eine oder andere Phänomen bevorzugt werden:

$$w_{sn}(G_i, G_j) := \frac{\kappa_{sn} w_n(G_i, G_j) + (1 - \kappa_{sn}) w_s(G_i, G_j)}{2} \quad (4.28)$$

Damit ist die Beschreibung für eine quantitative Formulierung der Gruppierungswirkung von Polygonen aufgrund von Nähe und Ähnlichkeit und deren Kombination abgeschlossen. In Abschnitt 7.1 wird auf die Wahl von κ_{sn} eingegangen. Im folgenden Abschnitt wird nun die Gruppierung von Polygonen aufgrund guter Fortsetzung erläutert.

4.4.4. Gute Fortsetzung

Im letzten Abschnitt sind zwei wichtige Gruppierungsphänomene zwischen Polygonen betrachtet worden. Die daraus entstandenen Beschreibungen beruhen vorwiegend auf einer paarweisen Betrachtung der Polygone. Sie sollen im folgenden eingesetzt werden, um mehrere Polygone aufgrund guter Fortsetzung in lineare Strukturen zu gruppieren.

Wie bereits in Abschnitt 4.1 angedeutet, läßt sich eine interessante Parallelität zwischen der Ebene der Konturmodelle und der $\mathbf{N} \times \mathbf{2D}$ -Ebene der Gruppierungshierarchie (Abbildung 4.1) feststellen:

- Die Approximation von Kantenpixeln durch Geradenstücke und Ellipsenbögen entspricht einer Gruppierung *naher* und *ähnlicher* Bildprimitiva (Konturelemente)

aufgrund *guter Fortsetzung*. Dabei ist die Ähnlichkeit durch die identische Form von Kantenpixeln inhärent gegeben, die Nähe wird über die euklidische Distanz auf dem festen Raster der Bildmatrix bestimmt.

- Auch für Polygone sind mit der quantitativen Formulierung von Nähe und Ähnlichkeit zwei elementare Eigenschaften für deren Gruppierung aufgrund guter Fortsetzung gegeben. Diese Analogie motiviert die Idee, die Organisation von Regionen aufgrund guter Fortsetzung ebenfalls durch eine adaptierte parametrische Approximation zu realisieren.

In Kapitel 2.1 ist ausführlich auf die Methoden einer Kontursegmentierung eingegangen worden. Bei einer Übertragung der dort diskutierten Prinzipien entfällt für die Gruppierung von Regionen aufgrund guter Fortsetzung der Schritt der *Detektion*. Denn von einem abstrakten Standpunkt aus betrachtet, können die Regionen bereits als die elementaren Kantenpixel gedeutet werden, die es zu approximieren gilt.

Lokale Nachbarschaft

Bei einer Approximation von Kantenpixeln wird häufig eine *Verkettung* derselben in einer lokalen Nachbarschaft zu Kontursegmenten vorgenommen. Im Fall von Polygonen ist diese Nachbarschaft durch die quantitative Formulierung von Nähe und Ähnlichkeit beschrieben und läßt sich über einen Graphen modellieren. Die Repräsentation der benachbarten Polygone wird durch einen gewichteten ungerichteten Graphen $M = (\mathbb{G}, \mathbb{E})$ abgebildet. In diesem Graphen ist die Menge der Knoten \mathbb{G} durch die im Bild detektierten Geschlossenheiten oder Regionen G gegeben. Die Menge der Kanten \mathbb{E} wird in Abhängigkeit von einem Schwellwert bezüglich der Gruppierungswirkung zwischen den Polygonen aufgebaut. Sei $w_{sn}(G_i, G_j)$ das Gewicht der Kante e zwischen den Polygonen G_i und G_j , so gilt:

$$\mathbb{E} = \{e_{i,j} \mid w_{sn}(G_i, G_j) \geq \Delta_{edge}\} \quad (4.29)$$

Der Wert von Δ_{edge} schränkt die Generierung eines vollständigen Graphen ein und wird bei der Vorstellung der Ergebnisse diskutiert. Auf Basis dieses Graphen kann nun eine Verkettung und Approximation der Polygone zu Bildstrukturen mit linearer Charakteristik erfolgen.

Approximation

Zur Approximation der Strukturen wird, wie bereits für die initialen Konturmodelle, das Verfahren von Leonardis (Abschnitt 2.1.3, Seite 16) eingesetzt. Diese Wahl ist bedingt durch die Tatsache, daß bereits bei den initialen Konturmodellen eine sehr gute und korrekte Approximation der zugrundeliegenden Kantenelemente beobachtet werden konnte. Jedoch läßt sich das Verfahren nicht direkt übertragen, da die deutlich unterschiedlichen

Eingangsdaten Anpassungen an verschiedenen Stellen erfordern. Diese Anpassungen werden im folgenden dargestellt. Die hierzu verwendete Nomenklatur orientiert sich an den bisher eingeführten Symbolen und Definitionen.

Lokalisation: Die punktförmige Lokalisation von Kantenelementen wird durch die Wahl der Schwerpunkte \overrightarrow{com} der Polygone abgebildet.

Normierung: Bei der Konturapproximation liegen die Abstände zwischen Kantenelementen auf einem fest definierten Raster und variieren nur in kleinen Intervallen. Bei einer Gruppierung g von Polygonen muß deshalb jede Abstandsmessung auf eine entsprechende mittlere Entfernung zwischen den Polygonen $G_i \in GElem(g)$ bezogen werden. Als Normierungswert wird daher der mittlere interne Abstand zwischen den Polygonschwerpunkten gewählt:

$$\overline{m^2}(g) := \frac{1}{n_g - 1} \sum_{\substack{G_i \in GElem(g), \\ i < n_g}} \|\overrightarrow{com}_i - \overrightarrow{com}_{i+1}\|^2 \quad (4.30)$$

Die Menge $GElem(g) = \{G_1, \dots, G_n\}$ beschreibt hier und im weiteren die geordnete Kette der n Polygone, die für eine Approximation verwendet wird. Die Anzahl der Polygone in $GElem(g)$ wird im folgenden verkürzt mit n_g bezeichnet.

Abstandsmessung: Wie bereits bei der Diskussion der Nähe von Polygonen dargestellt (Abschnitt 4.4.1, Seite 72), müssen sich Abstände und damit auch Approximationen auf den Abbildungsmaßstab der einzelnen gruppierten Polygone beziehen. Hierbei wird nur der quadratische mittlere Abstand zwischen den Schwerpunkten berücksichtigt, da die Form und Nähe der Polygone in der Selektionsphase über die entsprechenden quantitativen Formulierungen eingehen werden. Daraus ergeben sich folgende Anpassungen bezüglich der Akzeptanz des Approximationsfehlers (vergleiche Formel (2.1)) und der Kompatibilität eines Polygons $G_i \notin GElem(g)$ für die aktuelle Approximation (vergleiche Formel (2.2)):

$$\frac{\overline{d_\psi^2}(GElem(g))}{\overline{m^2}(g)} \leq \Delta_{accept}(\ominus) \quad (4.31)$$

$$\frac{d_\psi^2(G_i)}{\overline{m^2}(g)} \leq \Delta_{compatibility}(\ominus) \quad (4.32)$$

initiale Saaten: Die Auswahl initialer Polygonmengen für eine Approximation kann ebenfalls nicht aus der Konturapproximation übernommen werden. Eine analoge Anwendung von Saatenfenstern würde die unterschiedlichen Abbildungsmaßstäbe der Polygongruppen nicht berücksichtigen. Aus diesem Grund werden die Saaten einer Approximation aus allen dreielementigen Teilmengen $\mathbb{G}' \subset \mathbb{G}$ gebildet, die einen Pfad der Länge zwei bilden und die folgenden Bedingungen erfüllen:

4. Konturbasiertes hierarchisches Gruppieren

- Die Gruppierungswirkung zwischen den Polygonen $G_1, G_2, G_3 \in \mathbb{G}'$ ist stark genug.

$$w_{sn}(G_1, G_2) \cdot w_{sn}(G_2, G_3) \geq \Delta_{w_{sn}} \quad (4.33)$$

- Die Abstände der Schwerpunkte sind ähnlich.

$$\frac{|d_{com}(G_1, G_2) - d_{com}(G_2, G_3)|}{\max(d_{com}(G_1, G_2), d_{com}(G_2, G_3))} \leq \Delta_{continuity}(\ominus) \quad (4.34)$$

- Initial liegen die drei Polygone annähernd auf einer Gerade.

$$|\phi(\overrightarrow{com}_1 - \overrightarrow{com}_2) - \phi(\overrightarrow{com}_2 - \overrightarrow{com}_3)| \leq \Delta_\phi(\ominus) \quad (4.35)$$

Mit diesen Einschränkungen bezüglich der Orientierungen ϕ kann ein initialer Parametersatz ψ für ein Geradenstück geschätzt werden.

Expansion: Wie bei der Gewinnung der initialen Saaten ist auch die Expansion der approximierten Konturmodelle anzupassen. Die Expansion erfolgt, wie im ursprünglichen Verfahren, an den aktuellen Begrenzungen $G_1, G_{n_g} \in GElem(g)$ der Gruppe. Jedoch werden kompatible Partner über die Knotenmenge \mathbb{G}' ausgewählt:

$$\mathbb{G}' = \{G_j \mid \exists e_{i,j} \in \mathbb{E} \wedge G_i \in \{G_1, G_{n_g}\} \wedge G_j \notin GElem(g)\}$$

An einem der beiden Enden von g wird dann ein Polygon G_j probeweise zur Expansion gewählt, für das Bedingung (4.34) erfüllt ist und weiterhin gilt:

$$G_j = \underset{\substack{G_k \in \mathbb{G}', \\ G_i \in \{G_1, G_{n_g}\}}}{\operatorname{argmax}} w_{sn}(G_k, G_i) \quad (4.36)$$

$$\frac{|d_{GElem(g), \psi}^2(\overrightarrow{com}_j) - \overline{m}^2(g)|}{\overline{m}^2(g)} \leq \Delta_{continuity}(\ominus) \quad (4.37)$$

Mit Bedingung (4.36) wird das Polygon G_j mit der stärksten Gruppierungswirkung zur Expansion gewählt. Die Bedingung (4.37) schränkt den Abstand der Projektion von G_j auf das aktuelle Modell bezüglich dem mittleren internen Abstand $\overline{m}(g)$ der Gruppe ein. Sind diese Bedingungen erfüllt, kann eine expandierte Approximation mit G_j erfolgen und der rekursive Prozeß der Expansion fortgeführt werden.

Selektion: Nachdem alle dreielementigen Teilmengen mit Pfaden der Länge zwei als Saaten genutzt und maximal expandiert wurden, erfolgt die Selektion der besten Approximationen. Doch auch die *ModelSelection*-Phase muß angepaßt werden. Insbesondere soll die Gruppierungswirkung aufgrund von Nähe und Ähnlichkeit in die

Beurteilung der Qualität einer Approximation einfließen. Hierzu muß zunächst ein gemeinsamer Wert für die Gruppierungswirkung einer Gruppe g definiert werden:

$$\bar{w}_{sn}(g) := \frac{1}{n_g - 1} \sum_{\substack{G_i \in GElem(g), \\ i < n_g}} w_{sn}(G_i, G_{i+1}) \quad (4.38)$$

Anschließend wird die *Objective Function* zur Auswahl der ‚besten‘ Approximationen um die Berücksichtigung dieses Wertes erweitert (vergleiche Formel (2.4)):

$$F_{\ominus}(\mathbf{m}_{\ominus}) := \kappa_1(\ominus)n(\mathbf{m}_{\ominus}) - \kappa_2(\ominus)\xi(\mathbf{m}_{\ominus}) - \kappa_3(\ominus)N(\mathbf{m}_{\ominus}) + \sum_{g_i \in \mathbf{m}_{\ominus}} \kappa_4(\ominus)\bar{w}_{sn_i} \quad (4.39)$$

$$F_{interact_{\ominus}}(\mathbf{m}_{\ominus}) := F_{\ominus}(\mathbf{m}_{\ominus}) + \sum_{g_i \in \mathbf{m}_{\ominus}} \sum_{\substack{g_j \in \mathbf{m}_{\ominus}, \\ j \neq i}} \frac{-\kappa_1(\ominus)\Gamma_{i,j}}{2} \quad (4.40)$$

In Anlehnung an Formel (2.4) bezeichnet \mathbf{m}_{\ominus} die Teilmenge aller erzeugten Gruppierungshypothesen, die Polygone in Gruppen mit linearem Charakter organisiert und Formel (4.40) maximiert. Die beeinflussende Interaktion zwischen den Gruppierungshypothesen wird hier auf die Anzahl der gemeinsam approximierten Polygone beschränkt (vergleiche Formel (2.5)). Das Vernachlässigen des Approximationsfehlers an dieser Stelle motiviert sich dadurch, daß er aufgrund der geringen Anzahl konstituierender Polygone das Ergebnis in starkem Maße beeinflussen würde.

Alle beschriebenen Anpassungen werden nun in die Algorithmen 2.1 und 2.2 zur Generierung von initialen Konturmodellen übernommen. Damit wird auf Basis der Nähe- und Ähnlichkeitsbeziehungen zwischen Polygonen das Verfahren zur Konturapproximation für die Gruppierung linear angeordneter Polygone eingesetzt. Die entsprechend Parameter werden in Abschnitt 7.1 bei der Vorstellung der Ergebnisse angegeben.

4.4.5. Rekursive Gruppierung geschlossener Figuren

Bei einer abstrakten Betrachtungsweise besteht in der Generierung dieser Gruppierungen eine Analogie zu den vorherigen Gruppierungsprinzipien: Über die Quantifizierung der Gestaltprinzipien der Nähe und Ähnlichkeit werden potentielle Gruppierungspartner bestimmt. Die Hypothesisierung von linearen Gruppen erfolgt in der *ModelRecovery* durch eine Approximation von Pfaden innerhalb eines Nachbarschaftsgraphen. Hierbei werden einschränkende Bedingungen berücksichtigt. Die datengetriebene Eliminierung von redundanten Gruppierungshypothesen erfolgt durch die *ModelSelection*. Danach stehen

4. Konturbasiertes hierarchisches Gruppieren

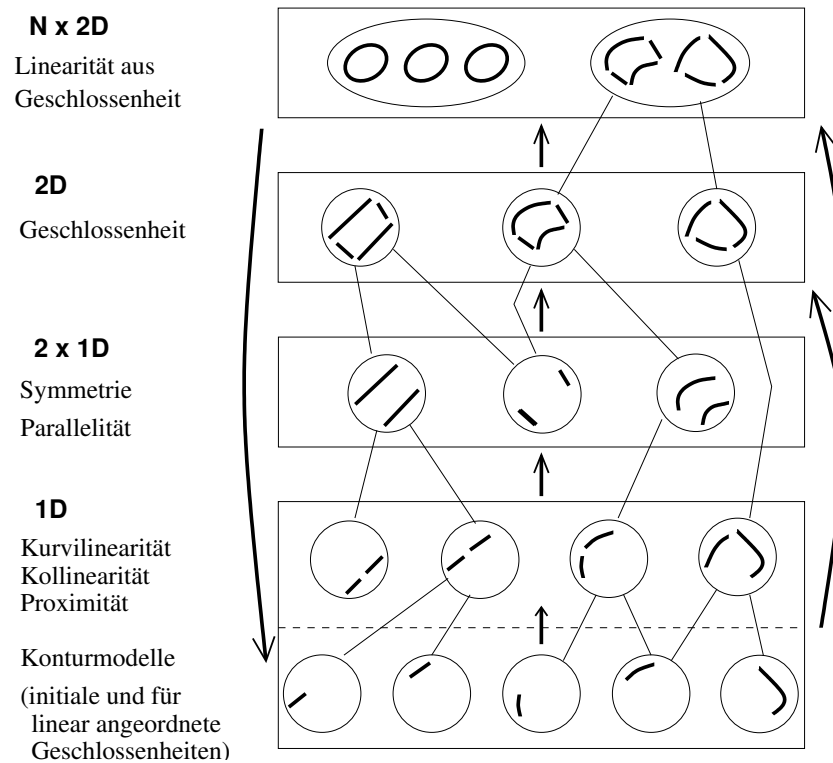


Abbildung 4.23.: Gruppierungshierarchie zur Modellierung rekursiver konturbasierter Gruppierungsphänomene. Die Konturmodelle für linear angeordnete Geschlossenheiten werden rekursiv in die Konturmodellebene eingefügt, und nur diese Modelle werden erneut gemäß der Hierarchie strukturell organisiert. Dicke Pfeile markieren den Datenfluß zwischen den Ebenen, während dünne Linien die Bestandteilsrelationen darstellen.

Konturmodelle zur Verfügung, die den Bildinhalt auf einer sehr hohen und abstrakten Ebene organisieren. Zur weiteren Verwendung können diese Gruppierungen wiederum in der Konturmodellebene der Gruppierungshierarchie eingesetzt werden und erneut gruppiert werden (vergleiche Abbildung 4.23). Damit ergibt sich ein rekursives Gruppierungsverfahren, das zunehmend stärker strukturierende Gruppierungen generieren kann. Dieser rekursive Gruppierungsprozeß endet, wenn auf der **N**×**2D**-Ebene keine Gruppierungen mehr erzeugt werden. In diesem Status liegen keine neuen abstrakten Bildprimitiva vor und der höchstmögliche Strukturierungsgrad für ein Bild ist erreicht. In Abschnitt 7.1 wird an Beispielen dieser rekursive Strukturierungsprozeß dargestellt. Allerdings muß darauf hingewiesen werden, daß möglicherweise eine Anpassung der Skalierung der Suchbereiche für Gruppen der **1D**-Ebene erfolgen muß, da die ursprünglichen Masken für deutlich lokalere Prozesse auf Bildpunktebene ermittelt wurden. In der Organisation von Konturmodellen für linear angeordnete Geschlossenheiten sollen jedoch

zunehmend globalere Strukturen detektiert werden. Die Untersuchung dieser Einflüsse wird im Rahmen der vorliegenden Arbeit nicht durchgeführt. Vielmehr werden die Ergebnisse in Abschnitt 7.1 unter Verwendung der in Abschnitt 4.2.1 diskutierten Bereiche perzeptiver Aufmerksamkeit generiert.

Ein weiterer Vorteil dieses Verfahrens ist, daß der Gruppierungsprozeß nicht auf die Gruppierung konturbasierter Geschlossenheiten eingeschränkt ist. Mit der Wahl von Polygonen als beschreibendes Bildprimitivum können mit diesem Verfahren verschiedenste polygonal beschriebene geschlossene Figuren organisiert werden. Auch dies wird in Abschnitt 7.1 gezeigt.

4. Konturbasiertes hierarchisches Gruppieren

5. Integration regionen- und konturorientierter Beschreibung

Bei der Vorstellung der verschiedenen Bildsegmentierungsverfahren in Kapitel 2 ist deutlich geworden, daß diese unterschiedlichen Ansätze ihre jeweiligen Vor- und Nachteile besitzen. Bezogen auf die konturbasierte Segmentierung wurde in Kapitel 4 ein Ansatz zur Gruppierung von Konturen vorgestellt, der Schwächen einer konturbasierten Interpretation (Fragmentierung) reduziert und gleichzeitig deren Vorteile für eine Strukturierung des Bildinhaltes in komplexere Einheiten ausnutzt. Dennoch bleiben problematische Konstellationen bestehen, die eine Organisation der Konturen in abstraktere Einheiten verhindern. In diesem Kapitel wird nun ein Ansatz für eine Integration der verschiedenen Interpretationskanäle vorgestellt, der zum einen die konturbasierte Problematik auflösen soll und zum anderen Hinweise auf eine fehlerhafte Interpretation eines regionenbasierten Ansatzes bereitstellt.

Zunächst wird im folgenden Abschnitt ein kurzer Überblick integrierender Ansätze gegeben und daraus die Motivation für das hier vorgeschlagene Verfahren abgeleitet. Anschließend wird die entwickelte Technik zur Kombination von Regionen und Konturgruppen erläutert, bevor dann der Einsatz des Verfahrens zur Detektion von konturbasierten geschlossenen Figuren vorgestellt wird (siehe auch [Sch98, Pos98]). Abschließend folgt eine Darstellung der Integration von konturbasierten Geschlossenheiten in ein regionenbasiertes Objekterkennungssystem (vergleiche [Sch00a]).

5.1. Motivation

Das visuelle System des Menschen nutzt verschiedene Arten von visuellen Reizen für eine Bildinterpretation. Die Perzeption von Bereichen mit homogenem Erscheinungsbild (gleiche Farbe, Textur, etc.) oder von Konturen, die diese Bereiche begrenzen und genau lokalisieren, sind nur zwei der wichtigsten Komponenten. Unter anderem ist auch die Detektion von Bewegung von Bereichen oder von Konturen ein wichtiger Aspekt für die Perzeptionsleistung. Eine solche Segmentierung über die Zeit liefert wichtige Informationen für eine quantitative Beurteilung der Interpretation von Einzelbildern und trägt zu einer Verifikation der Bildanalyse bei. Für den Menschen ist eine Objekterkennung allein aufgrund charakteristischer Objektkonturen oder aufgrund des Oberflächenerscheinungsbildes deutlich schwieriger, als eine Erkennung unter Ausnutzung aller Interpretationskanäle. Häufig führt sie sogar zu Fehlinterpretationen. Mit diesen Vorstellungen

motiviert sich der Ansatz verschiedene Interpretationskanäle auch in der Bildanalyse zu kombinieren.

Bezugnehmend auf die automatische Bildanalyse diskutieren beispielsweise Sanocki et. al in [San98] die Frage „*Are edges sufficient for object recognition?*“ In ihrer Diskussion orientieren sie sich an den Ergebnissen von Biederman [Bie87b, Bie87a], in denen aufgrund von Experimenten mit Strichzeichnungen die hohe Signifikanz von Konturen für eine Objekterkennung postuliert wird. Auf Basis dieser Experimente sieht Biederman Konturen mit ihren Eigenschaften (Krümmung, lineare Fortsetzung, Symmetrie, Kreuzungen) als hinreichende Informationsquelle für eine Objekterkennung, und viele Ansätze sind dieser Vorstellung gefolgt. Sanocki et al. argumentieren jedoch, daß in der automatischen Bildanalyse mit Konturen zunächst nur lokale datengetriebene Information zur Verfügung steht, die ohne eine Figur-Hintergrund-Trennung, d.h. ohne eine gewisse Vorerwartung oder Weltwissen über das Betrachtete, unzureichend für eine Objekterkennung ist. Konturen per se sind keine Strichzeichnungen – „*A line drawing is a high-level rendering of an object in which noise is filtered out and essential properties are abstracted and included*“ ([San98]). Die Problematik einer Bildinterpretation kann nach ihrer Meinung nur durch die Berücksichtigung höherer oder weiterer Informationen (Textur, Farbe, homogene Bereiche) gelöst werden. Diese Informationen liefern einschränkende Bedingungen bei der Generierung der jeweiligen Segmentierungsergebnisse und führen zu einer Verschränkung von struktureller und globaler Interpretation mit datengetriebenen Informationen (siehe auch [Pal94]). Darüber hinaus schlagen sie vor, auch domänenspezifische Informationen einzusetzen, die jedoch die Allgemeinheit eines Ansatzes in hohem Maße einschränken können. Abschließend sehen sie den Einsatz von Informationen der mittleren Interpretationsebene (Struktur, etc.) als die zunächst vielversprechendste Komponente für eine Kombination in der automatischen Bildanalyse.

Verschiedene Ansätze sind entwickelt worden, um Kanäle unterschiedlicher Informationen miteinander zu kombinieren und damit eine robustere Bildsegmentierung zu erreichen. Pavlidis und Liow [Pav90] führen beispielsweise zunächst ein Regionenwachstum mit einem Split-and-Merge Ansatz durch. Dabei wählen sie die Parametrisierung so, daß eine Übersegmentierung des Bildes erfolgt. Anschließend werden Regionengrenzen entfernt oder modifiziert, die Einschränkungen bezüglich des Kontrastes und der Variationen im Gradienten entlang einer Regionengrenze nicht erfüllen. Abschließend werden dann noch Artefakte eliminiert, die aufgrund der eingesetzten Quadtree-Technik entstanden sind. Dies erfolgt durch Berücksichtigung von Glattheitsbedingungen in Bezug auf die Krümmung des Konturverlaufs. In ähnlicher Weise wird in [Xua95, Xia92] eine Integration von Regionen- und Konturdetektion für die Segmentierung von MR-Bildern vorgenommen. In [Koh95] wird Konturinformation in eine Regionensegmentierung integriert, um die Detektion von Regionen zu verifizieren und letztendlich eine robuste 3D-Lokalisierung von Objekten zu erreichen.

Im Gegensatz zu den bisherigen Techniken, eine Regionensegmentierung durch Gradienteninformation einzuschränken oder zu verifizieren, wird in [Pau96] ein stochasti-

sches Regionenwachstum verwendet, um die Fragmentierung von bereits detektierten Konturen zu reduzieren. Auch in [Ben92] werden die Ergebnisse einer Konturdetektion verwendet, um letztendlich geeignete Startpunkte für ein Regionenwachstumsverfahren zu erhalten.

Eine neuere Arbeit wird von Malik et al. in [Mal99, Mal00] vorgestellt. Aufbauend auf der Methode aus [Leu98], die eine Integration von Kantengruppierung mit linearem Charakter und einer Regionensegmentierung vorschlägt, wird eine Bildbeschreibung von natürlichen Szenen auf Basis von Texturen und Konturen vorgeschlagen. Hierzu werden diese bereits bei ihrer Segmentierung kombiniert. Ein ähnlicher Ansatz gegenseitig beeinflussender Segmentierungstechniken wird in [Cha99] vorgeschlagen. Motiviert durch Entscheidungstechniken aus der Spieltheorie wird eine wechselseitig beeinflussende Regionensegmentierung mit Markov Random Fields und eine parametrisierte Konturapproximation mit Fourierkoeffizienten präsentiert.

In dieser Diskussion fällt auf, daß alle Ansätze bereits bei der Bildsegmentierung eine Integration der verschiedenen Verfahren vorschlagen. Im Gegensatz dazu wird in dem hier entwickelten Verfahren eine Kombination der verschiedenen Segmentierungskanäle erst auf der Ebene der entsprechenden Beschreibungsform (Regionen, *Konturgruppen*) vorgeschlagen¹. Dies motiviert sich aus der Tatsache, daß auf der Seite der Konturgruppen bereits eine stärkere Strukturierung des Bildinhaltes stattgefunden hat, und diese unabhängig von den Homogenitätsbedingungen einer regionenbasierten Interpretation generiert wurde. So werden mögliche Fehlsegmentierungen vermieden, die bei einer frühzeitigen Kombination entstehen können: Regionen werden wie bisher unabhängig von Konturen generiert und beruhen nur auf ihrem Homogenitätskriterium, wie auch Konturbeschreibungen nur auf die Diskontinuitäten im Intensitätskanal des Bildes zurückzuführen sind.

5.2. Kombination von Konturgruppen und Regionen

Die im letzten Abschnitt vorgestellten Ansätze besitzen die gemeinsame Eigenschaft, daß alle die Kombination bereits in der Segmentierungsebene eines Bildverarbeitungssystems modellieren und daher auf einer pixelorientierten diskreten Beschreibungsform basieren. In diesem Abschnitt wird nun der in dieser Arbeit entwickelte Ansatz vorgestellt. Hierzu wird zunächst aus charakteristischen Situationen einer fehlerhaften Segmentierung die grundlegende Konzeption einer Kombination dargestellt, bevor im Anschluß das Verfahren im Detail beschrieben wird.

¹ Mit dem Begriff *Konturgruppen* seien in Bezug auf die Kombination mit Regionen nur lineare Gruppen und Geschlossenheiten bezeichnet.

5.2.1. Allgemeine Konzeption einer Zuordnung

Ein prototypischer Fehler bei der Segmentierung von Regionen entsteht, wenn die Projektion einer homogenen Objektoberfläche in zwei Regionen zerfällt, weil ein anderes Objekt die Oberfläche zum Teil verdeckt. Abbildung 5.1 zeigt ein Beispiel, in dem jede Regionensegmentierung für die von den Würfeln verdeckte Leiste vollständig getrennte Regionen ermittelt. Diese problematische Situation kann durch das Verfahren zur konturbasierten Gruppierung aufgelöst werden: Für die fragmentierten Konturen der Leistenoberfläche werden kollineare Gruppen gebildet, die anschließend zu einer Hypothesisierung einer geschlossenen Figur eingesetzt werden (Abbildung 5.1(d)). Diese geschlossene konturbasierte Region stellt dann eine alternative (korrekte) Beschreibung der Leistenoberfläche dar. Für eine integrierte Objekterkennung fehlt allerdings die Information, daß diese geschlossene Figur die getrennten Holzregionen vereinigt. Zu diesem Zweck wird hier das Konzept von *Matches* eingeführt. Ein Match beschreibt die Zuordnung einer Konturgruppierung zu einer Region in dem Sinne, daß beide Beschreibungen aufgrund der selben kausalen Ursache im Bild entstanden sind, beispielsweise der Projektion einer Objektoberfläche. Die Matches werden dabei in zwei Arten unterschieden:

umrissbeschreibende Zuordnung: Eine Konturgruppierung beschreibt im wesentlichen die Begrenzung einer homogenen Region (Abbildung 5.1(e)).

strukturbeschreibende Zuordnung: Eine Konturgruppierung beschreibt eine Struktur oder Textur innerhalb des homogenen Bereiches einer Region (Abbildung 5.1(f)).

Mit diesem Konzept steht die Information zur Verfügung, daß in dem Beispiel aus Abbildung 5.1 Elemente einer geschlossenen Figur (die fragmentierten Konturen) den einzelnen Regionen als umrissbeschreibend zugeordnet sind. Durch diese Zuordnung kann nun eine Beziehung zwischen den Regionen und der konturbasierten Beschreibung für die Leistenoberfläche etabliert werden und für eine Objekterkennung genutzt werden.

Ein anderes Problem bei einer Regionensegmentierung ist die Wahl der einschränkenden Bedingungen für das verwendete Homogenitätskriterium. In Abbildung 5.1(c) ist die Trennung der beiden holzfarbenen Leisten nicht gelungen, da bei dem hier verwendeten Verfahren aus Abschnitt 2.2.1 allein die Farbe als Homogenitätskriterium eingesetzt wird. Auch andere Regionensegmentierungsverfahren erzielen hier aufgrund schwach ausgeprägter Inhomogenitäten keine Trennung. Das vorgeschlagene Konzept vermittelt einer Objekterkennung jedoch die Information, daß Konturgruppierungen die Strukturen innerhalb der verschmolzenen Regionen beschreiben. Damit kann eine Erkennung der beiden Leisten aus alternativen konturbasierten Regionen (Geschlossenheiten) erfolgen.

Ein charakteristischer Nachteil von Kontursegmentierungsverfahren ist die Auswertung nur lokaler Information bei der Detektion von Diskontinuitäten im Bild. Eine Kontursegmentierung mißlingt daher in Bildbereichen, in denen der Gradientenbetrag an den Kanten eines Objektes nur schwach ausgeprägt ist. Ein Beispiel ist in Abbildung 5.2 gezeigt. Im oberen Bereich der Leisten können aufgrund des fehlenden Kontrastes im

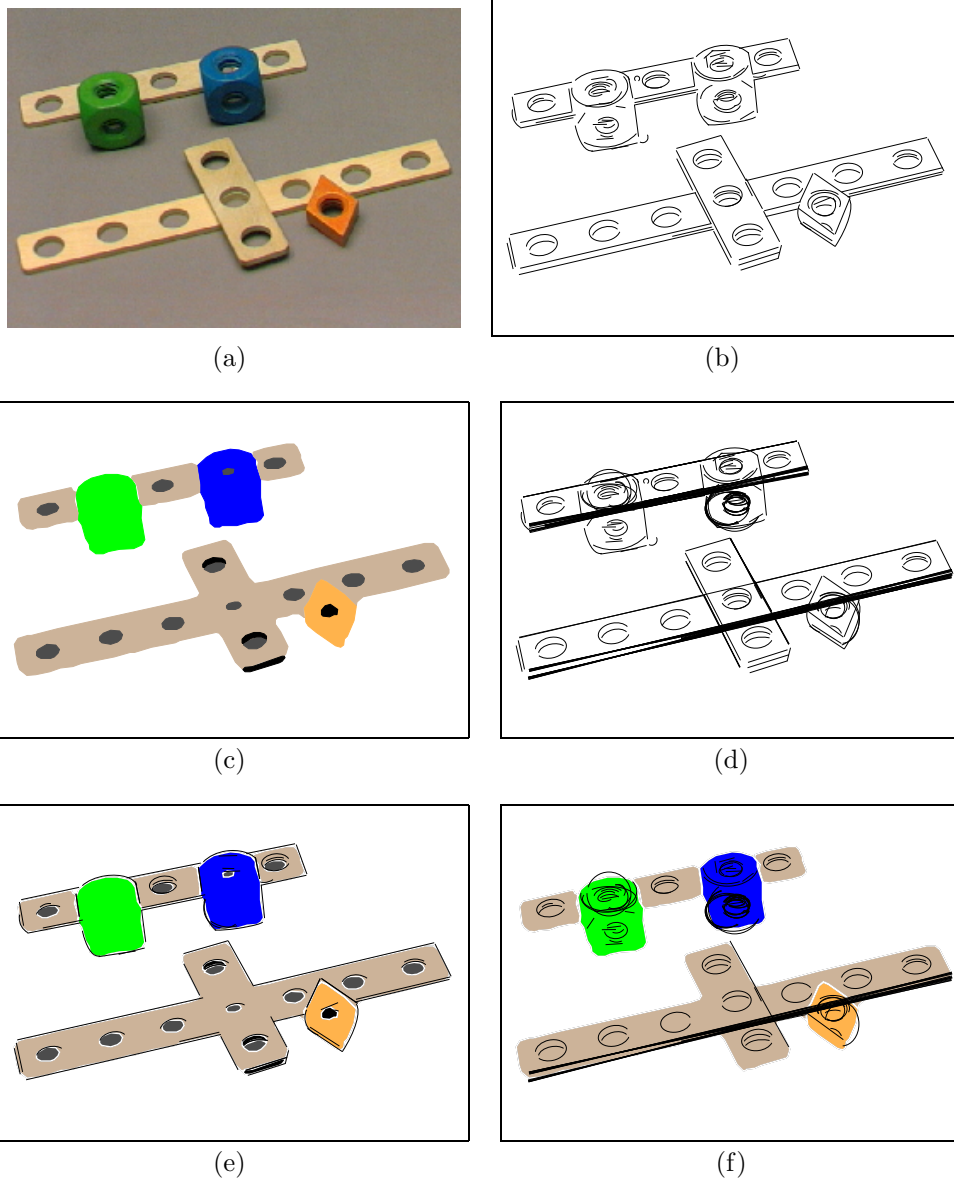


Abbildung 5.1.: Beispielbild zur Anschauung des Konzeptes von Matches zur Unterstützung regionenbasierter Objekterkennungsverfahren: (a) Originalbild, (b) detektierte Konturmodelle, (c) Regionensegmentierung mit pixelweiser Klassifikation, (d) lineare Gruppen die aus (b) generiert wurden, (e) umrissbeschreibende Zuordnung der linearen Gruppen zu den Regionen und (f) strukturbeschreibende Zuordnungen. Mit den linearen Gruppen in (d) kann problemlos eine geschlossene Figur für die verdeckte Leiste generiert werden. Die geschlossenen Figuren für die überlagerten Leisten lassen sich ebenfalls generieren.

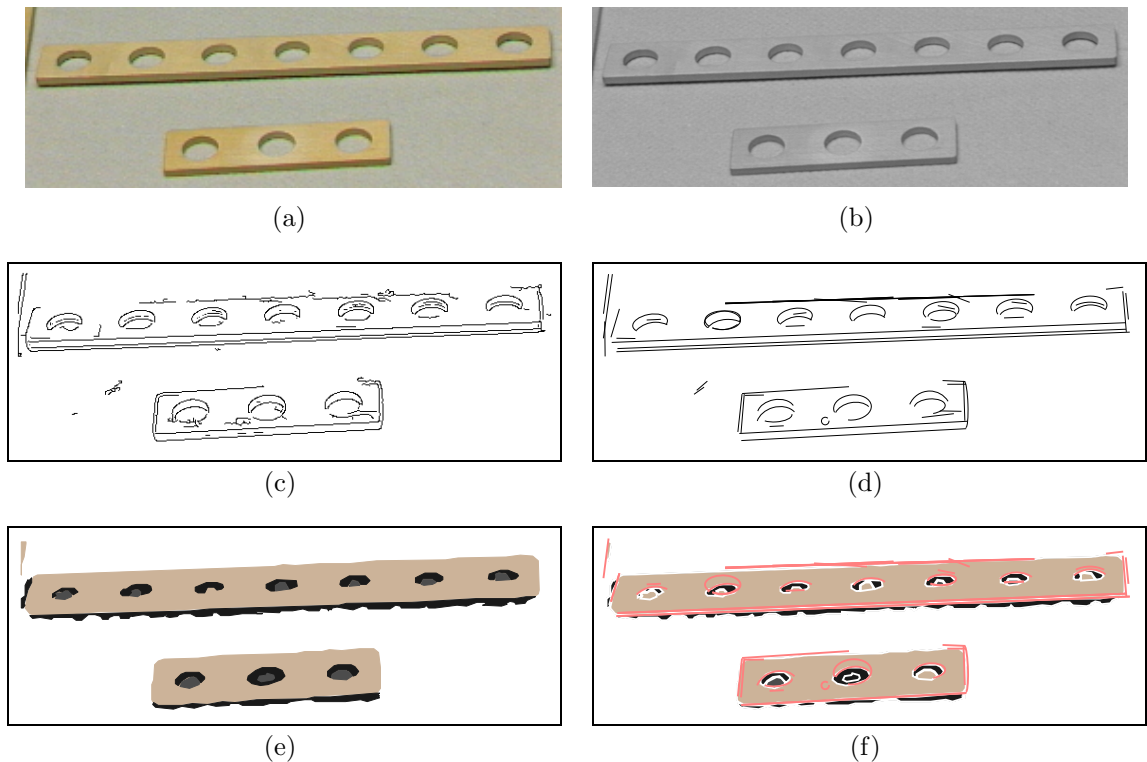


Abbildung 5.2.: Beispielbild zur Anschauung des Konzeptes von Matches zur Unterstützung beim konturbasierten Gruppieren: (a) Originalbild, (b) Intensitätskanal des Bildes, (c) verdünntes Gradientenbetragsbild, (d) lineare Gruppen die aus (c) generiert wurden, (e) Regionensegmentierung mit pixelweiser Klassifikation und (f) umrissbeschreibende Zuordnung der linearen Gruppen zu den Regionen (lineare Gruppen sind farbig hervorgehoben). Mit den Regionen in (e) kann problemlos eine geschlossene Figur für die fragmentierten Konturen der Leisten generiert werden.

Intensitätskanal des Bildes (Abbildung 5.2(b)) keine oder nur geringe Diskontinuitäten ermittelt werden. Eine Regionensegmentierung kann jedoch homogene Regionen für die Objektflächen detektieren, da sie für die Bestimmung der Regionen globalere Information einsetzt (Abbildung 5.2(e)). Mit dem oben vorgeschlagenen Konzept können nun die ermittelten linearen Gruppen den Regionen als umrissbeschreibend zugeordnet werden (Abbildung 5.2(f)). Damit kann die Regionengrenze im Bereich des schwachen Kontrastes als Hinweis genutzt werden, um eine Gruppierung zwischen den nächstgelegenen zugeordneten Gruppen zu hypothetisieren. Mit dieser regionenbasierten Konturgruppierung wird dann die Generierung einer konturbasierten Geschlossenheit unterstützt, die die Objektfläche korrekt beschreibt. Neben dieser unterstützenden Wirkung in problematischen Bildbereichen können die Matches auch genutzt werden, um eine Bear-

beitungsreihenfolge im Generierungsprozeß der Gruppierungen vorzugeben. Bevor diese Anwendungsmöglichkeiten im Detail dargestellt werden, wird im folgenden das Verfahren zur Erzeugung der Matches auf Basis einer einheitlichen geometrischen Repräsentation beschrieben.

5.2.2. Einheitliche Repräsentation

Für die Kombination von Konturgruppen und Regionen ist zunächst die Form zu betrachten, in denen die jeweiligen Ergebnisse beschrieben werden. Obwohl beide das gleiche Ereignis im Bild beschreiben, liegen ihre Ergebnisse in unterschiedlicher Form vor. Für eine Integration ist es jedoch vorteilhafter, eine einheitliche Repräsentation der Ergebnisse einzusetzen, da dies die Berechnung verschiedener Merkmale unterstützt.

Jede Region wird in der Regel durch die Pixelkette ihrer Randpunkte beschrieben. Diese Kette wird in den hier eingesetzten Segmentierungsverfahren aus speichertechnischen Gründen und wegen der effizienteren Handhabung durch eine geschlossene polygonale Approximation $poly(r)$ angenähert. Hierzu wird ein *Splitting*-Algorithmus in Kombination mit einem *Least-Square*-Verfahren eingesetzt. Für ähnliche Verfahren sei auf [Pav74, Wal84, Ray95] verwiesen. Die Grenzen zu inhomogenen Bereichen oder anderen homogenen Regionen innerhalb einer Region werden nicht berücksichtigt. In Abbildung 5.3(b) ist eine Approximation einer synthetischen Region gezeigt.

Eine Konturgruppierung hingegen ist durch parametrische Konturmodelle beschrieben. Diese lassen sich durch offene Polygonzüge $poly(g)$ abbilden: Bei Geradenstücken und Kollinearitäten erfolgt diese Abbildung durch die Angabe von Start- und Endpunkt. Für Ellipsenbögen und Kurvilinearitäten wird der Verlauf der Kurve näherungsweise durch entsprechende Stützpunkte und die Start- und Endpunkte approximiert (siehe Abbildung 5.3(b)). Bei Geschlossenheiten werden geschlossene Polygonzüge erzeugt, wobei Proximitäten durch Geradenstücke zwischen den einzelnen polygonalen Approximationen der linearen Gruppen modelliert sind. Um den Polygonverlauf von geschlossenen Figuren überschneidungsfrei zu halten (siehe Beispiel in Abbildung 4.10(b)), wird im Falle einer Kreuzung von zwei nahen linearen Gruppen nur der Kreuzungspunkt als Stützpunkt in die polygonale Beschreibung aufgenommen. Die eigentlichen Endpunkte werden nicht berücksichtigt. Dies ist eine zulässige Vereinfachung, da bereits bei der Generierung der Geschlossenheiten nur geringfügige Überschneidungen zugelassen werden. Parallelitäten und Gruppen von geschlossenen Figuren finden keine Berücksichtigung in diesem Integrationsansatz, da sie bereits durch die entsprechenden konstituierenden Elemente repräsentiert sind. Damit liegt eine einheitliche polygonale Repräsentation aller Bildprimitiva vor. Im folgenden wird das Symbol P für die geordnete Menge $\{p_1, \dots, p_n\}$ der n Stützpunkte eines Polygons $poly$ verwendet. Bei Regionen oder Geschlossenheiten wird die Ordnung der Menge als zyklisch angenommen.

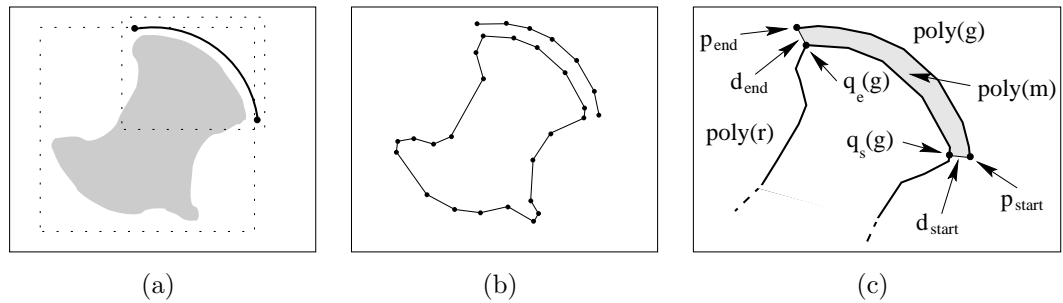


Abbildung 5.3.: Beispiel zur Repräsentation von Regionen und Gruppierungen und zur Merkmalsberechnung für Matches. (a) Synthetisches Beispiel mit den umschreibenden Rechtecken zur Auswahl potentieller Kombinationspartner, (b) Repräsentation beider Bildprimitiva als Polygone, (c) Merkmale zur Klassifikation eines Matches (siehe Text).

5.2.3. Merkmale und Klassifikation einer Zuordnung

Die Zuordnung einer Konturgruppierung zu einer Region soll nach dem vorgeschlagenen Konzept erfolgen, wenn beide Bildprimitiva mit hoher Wahrscheinlichkeit die Projektion desselben Objektes bzw. Ereignisses in der Szene beschreiben. Zu diesem Zweck werden im folgenden verschiedene Merkmale vorgestellt, mit denen eine Entscheidung getroffen werden kann, ob dasselbe Ereignis kausale Ursache ist und in welcher Beziehung die Kontur und die Region zueinander stehen. Da Konturgruppen und Regionen in demselben Bild ermittelt werden, kann zunächst der Suchbereich eingeschränkt werden, in dem potentiell kombinierbare Bildprimitiva lokalisiert sein können. Hierzu werden die im folgenden vorgestellten Merkmale nur für Paare von Regionen und Konturgruppierungen berechnet, die in einer gegenseitigen Nachbarschaft lokalisiert sind. Zur Bestimmung der Nachbarschaft wird zunächst für die Region r und für die Kontur g jeweils ein um 10 Pixel expandiertes umschließendes Rechteck bestimmt (siehe Abbildung 5.3(a)). Eine Region und eine Konturgruppierung sind demnach benachbart, wenn ihre umschließenden Rechtecke einen nicht leeren Schnitt bilden. Für entsprechend benachbarte Bildprimitiva werden dann folgende Merkmale berechnet (siehe Abbildung 5.3(c)):

Endpunktabstand: Der Endpunktabstand ergibt sich aus dem euklidischen Abstand der Endpunkte einer Gruppierung zu dem Polygon der Region:

$$d_{start}(g, r) := \min d(p_{start}, poly(r)) \quad (5.1)$$

$$d_{end}(g, r) := \min d(p_{end}, poly(r)) \quad (5.2)$$

Hierbei beschreibt $d(p, poly)$ allgemein den kürzesten Abstand eines Punktes p zu der polygonalen Repräsentation $poly$ eines Bildprimitivums. Bei Geschlossenheiten wird der Stützpunkt als identischer Start und Endpunkt gewählt, der den

geringsten Abstand besitzt: $p_{start} = p_{end} = \underset{p \in P(g)}{\operatorname{argmin}} d(p, \operatorname{poly}(r))$

flächenbasierter Abstand: Zur Bestimmung eines mittleren Abstandes zwischen einer Konturgruppierung und einer Region wird ein flächenbasierter Abstand gewählt. In einem ersten Schritt werden die Endpunkte der Gruppierung über die kürzeste Distanz auf das Polygon der Region projiziert. Mit dieser Projektion wird die Menge $P(r)$ um zwei Punkte erweitert, die die Fußpunkte $q(g)$ der Projektion beschreiben. Es entsteht eine neue Menge $P_g(r)$ für das Regionenpolygon, in der die Fußpunkte $q_s(g), q_e(g)$ entsprechend der Ordnung der Stützpunkte von $P(r)$ einsortiert sind:

$$P_g(r) = \{p_1, \dots, p_i, q_s(g), p_j, \dots, p_k, q_e(g), p_l, \dots, p_n\}$$

Nun kann der Bereich zwischen der Region und der Kontur durch ein Polygon $\operatorname{poly}(m)$ beschrieben werden (siehe Abbildung 5.3(c)), das sich aus der folgenden Menge von Stützpunkten ergibt:

$$P(m) = \{q_s(g), p_j, \dots, p_k, q_e(g), p_{end}, \dots, p_{end}\},$$

wobei $\{q_s(g), p_j, \dots, p_k, q_e(g)\} \subseteq P_g(r)$ und $\{p_{end}, \dots, p_{end}\} = P(g)$ gilt. Hierbei wird garantiert, daß $\operatorname{poly}(m)$ ein *einfaches* Polygon ist, also überschneidungsfrei gehalten wird. Um nun eine Schätzung des mittleren Abstandes zwischen der Region und der Kontur zu ermitteln, wird die Fläche dieses Polygons $\operatorname{poly}(m)$ mit der analytischen Länge der Konturgruppierung normiert:

$$d_{\text{avg-area}}(g, r) := \frac{A(\operatorname{poly}(m))}{l(g)}$$

Überlappung: Die unabhängige Generierung von Regionen und Konturen kann zu sehr deutlichen Abweichungen in der Beschreibung ein und desselben Ereignisses führen. Für eine Quantifizierung, in welchem Maße die Kontur zu der entsprechenden Region gehört, wird der Anteil $l_r(\operatorname{poly}(g), \operatorname{poly}(r))$ des Konturpolygons ermittelt, der innerhalb des Regionenpolygons verläuft. Dieser Wert wird dann ins Verhältnis zu der Länge der Gruppe g gesetzt:

$$l_{\text{over}}(g, r) := \frac{l_r(\operatorname{poly}(g), \operatorname{poly}(r))}{l(g)}$$

Mit diesen Merkmalen wird dann ein Test durchgeführt, der entscheidet ob beide Bildprimitiva dasselbe Ereignis als kausale Ursache besitzen. Ähnlich wie bei der Gruppierung von geschlossenen Figuren, liefert keines der Merkmale hinreichende Information für eine klare Entscheidung, und so wird auch hier ein Polynomklassifikator zur Unterscheidung eingesetzt. Zu diesem Zweck werden die oben beschriebenen Merkmale in einen Merkmalsvektor zusammengefaßt. Ein entsprechend trainierter Klassifikator unterscheidet dann die Zuordnung einer Gruppe zu einer benachbarten Region in umrissbeschreibend oder strukturbeschreibend.

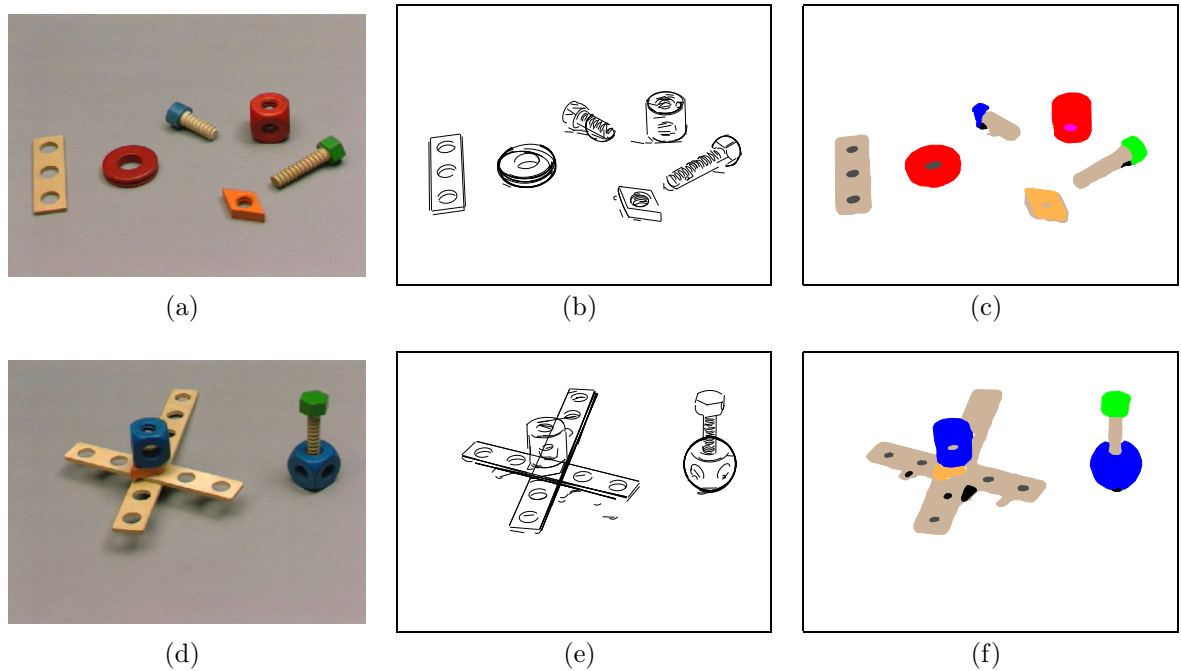


Abbildung 5.4.: Auswahl von Bildern aus der Stichprobe zur Klassifikation von Matches: linke Spalte: Originalbilder, mittlere Spalte: lineare Gruppen, rechte Spalte: Regionensegmentierung mit pixelweiser Klassifikation.

Stichprobe: Zur Unterscheidung zwischen einer umrissbeschreibenden und einer strukturbeschreibenden Kontur wurde eine Stichprobe von 11 Testbildern mit prototypischen Objektkonstellationen aus der *baufix*[®] Domäne gewählt. In [Abbildung 5.4](#) sind zwei der Testbilder mit den generierten linearen Gruppen und den entsprechenden Regionen dargestellt. Von einer Versuchsperson wurden in diesen Testbildern die Konturen als umrissbeschreibende oder strukturbeschreibende Matches den Regionen zugeordnet. Dabei wurde eine Unterscheidung in zwei Klassen gefordert:

$$\Omega_{match} = \{\omega_{boundary}, \omega_{structural}\}$$

Die Einführung einer weiteren Klasse ω_{noise} , die eine vorhandene Kontur als Rauschen interpretiert, wurde untersucht, aber aufgrund der starken Unterrepräsentation in der Stichprobe verworfen. Entsprechende Konturen sind daher für das Training nicht berücksichtigt worden.

Merkmalvektor: Für die markierten Paare aus linearen Gruppen und Regionen wird der

Merkmalsvektor für einen Match wie folgt definiert:

$$\mathbf{c}_{match} = \begin{pmatrix} d_{start} \\ d_{end} \\ d_{avg-area} \end{pmatrix}$$

Der Wert l_{over} wurde zunächst in den Merkmalsvektor aufgenommen, jedoch hat sich in einer Analyse der erzielten Klassifikationsergebnisse herausgestellt, daß dieses Merkmal keinen signifikanten Einfluß auf das Klassifikationsergebnis ausübt. Aus diesem Grund wurde der Wert für l_{over} in dem Merkmalsvektor nicht weiter berücksichtigt. Mit diesen Merkmalsvektoren aus der klassifizierten Stichprobe wurde anschließend ein Polynomklassifikator mit einem Polynom sechsten Grades trainiert. Zur Wahl des Polynomgrades sei hier auf das Ergebniskapitel 7.1, Seite 165 verwiesen.

Für ein Zwei-Klassen-Problem wurden die Eigenschaften eines Polynomklassifikators bereits in Abschnitt 4.4.3 dargestellt. Im Gegensatz zu der Anwendung bei der Beurteilung der Gruppierungswirkung von linear angeordneten Geschlossenheiten wird hier keine quantitative Aussage aus dem Klassifikationsergebnis abgeleitet. Alleiniges Ziel ist die Zuordnung eines Merkmalsvektors in eine der beiden Klassen $\omega_{boundary}$ und $\omega_{structural}$. Um auch hier die Sicherheit der Entscheidung für eine Klasse ω_i gemäß Gleichung (4.25) (Seite 88) in der Anwendung zu gewährleisten, wird eine Entscheidung nur akzeptiert, wenn der Wert der Unterscheidungsfunktion $d_{\omega_i}(\mathbf{c})$ im Intervall $]0.65, 1.25[$ liegt. Alle anderen Werte deuten auf eine Zuordnung einer Kontur zu einer Region hin, die mit dem trainierten Klassifikator nicht klassifiziert werden kann bzw. wo die Entscheidung eher unsicher ist. In diesen Fällen wird dann *kein* Match generiert.

Trotz einer klassifizierten Stichprobe von 1559 Merkmalsvektoren, kann es zu offensichtlichen Fehlklassifikationen durch den Klassifikator kommen. Aus diesem Grund wird nach einer Entscheidung für eine strukturbeschreibende Zuordnung die Bedingung definiert, daß für die Akzeptanz dieser Entscheidung eine minimale Überlappung $l_{over} > \Delta_{struct-over}$ der Kontur mit der Region vorliegen muß. Der Wert für $\Delta_{struct-over}$ wurde zu 0.5 gewählt. So müssen also Konturen mit ihrer polygonalen Approximation zumindest zu 50% innerhalb der Regionengrenze verlaufen, um als eine Strukturapproximation für eine Region akzeptiert zu werden.

Algorithmisch wird der Prozeß einer Kombination für die vollständigen Mengen von Konturgruppen und Regionen eines Bildes folgendermaßen durchgeführt: Zunächst werden die Regionen entsprechend ihrer Größe sortiert. Anschließend wird beginnend mit der kleinsten Region jeweils ein Test gegen alle vorhandenen Konturgruppen durchgeführt und entsprechende Zuordnungen generiert. Wird dabei einer Region r eine Konturgruppe g als umrissbeschreibend zugeordnet, so darf diese Gruppe g im weiteren Prozeß nicht mehr als Strukturbeschreibung verwendet werden. Damit wird sichergestellt, daß Konturgruppen, die bereits den Umriß einer kleineren Region approximieren, nicht als Strukturbeschreibung für größere Regionen (z.B. die Hintergrundregion) interpretiert

werden. Prinzipiell ist diese Konturgruppe naturgemäß eine Strukturapproximation innerhalb der größeren Region, allerdings soll hier möglichst eine eindeutige Zuordnung für alle Regionen erfolgen. Diese Einschränkung gilt jedoch nicht für umrissbeschreibende Zuordnungen. So kann *eine* Konturgruppe selbstverständlich als Approximation des Umrisses von *zwei* aneinandergrenzenden Regionen verwendet werden.

Damit ist die Beschreibung des Verfahrens zur Erzeugung von Zuordnungen zwischen Regionen und konturbasierten Gruppen abgeschlossen. In den folgenden Abschnitten werden nun Einsatzmöglichkeiten für die Matches vorgestellt und diskutiert.

5.3. Konturbasiertes Gruppieren mit Regionenintegration

Nachdem im vorherigen Abschnitt die Generierung von Zuordnungen zwischen Konturgruppen und Regionen diskutiert wurde, wird nun deren Verwendung als unterstützende Information in der Hypothesisierung von konturbasierten Gruppen beschrieben. Zum einen unterstützen die Zuordnungen die Auswahl der zu untersuchenden Konturmodelle auf Basis von Regioneninformation (Abschnitt 5.3.1) und zum anderen resultiert diese Auswahl in einer inkrementellen Verarbeitungsstrategie des gesamten Gruppierungsprozesses (Abschnitt 5.3.2). In Abschnitt 5.3.3 wird ein Verfahren vorgestellt, das die Generierung von **2D** Gruppierungen durch Regioneninformation unterstützt.

5.3.1. Bearbeitungsreihenfolge gruppierender Konturmodelle

In Abschnitt 4.2.2 wurde bei der Hypothesisierung von Gruppierungen der **1D** und **2×1D** Ebene auf die Frage hingewiesen, in welcher Reihenfolge Konturmodelle initial auf ihre Gruppierungswirkung mit anderen Konturmodellen ausgewählt werden sollen. Es wurde vorgeschlagen, daß die Auswahl über die Länge $l(g)$ der jeweiligen linearen Gruppen g getroffen wird, da lange Konturmodelle eine hohe Evidenz für eine kausale Ursache im Bild beinhalten. Soweit keine anderen Evidenzen für die Bestimmung einer Bearbeitungsreihenfolge vorliegen, ist die Wahl dieses Kriteriums naheliegend.

Mit der Zuordnung von Konturen zu Regionen liegt nun eine weitere Information für eine Bearbeitungsreihenfolge vor. Das Ziel der Matches ist es, Konturen und Regionen einander zuzuordnen, wenn sie das Segmentierungsergebnis desselben Ereignisses im Bild darstellen. Unter der Annahme, daß diese Zielsetzung erfüllt wird, werden lineare Gruppen g nun regionenweise für die Untersuchung auf Gruppierungseigenschaften betrachtet. Hierzu werden zunächst die initialen Konturmodelle als umrissbeschreibende *und* strukturbeschreibende Matches den Regionen zugeordnet. Anschließend wird für jede Region r_i die Menge O_i aller zugeordneten Konturen bestimmt. Für jede dieser Mengen O_i werden dann die enthaltenen Konturmodelle $g \in O_i$ auf die in der **1D** und **2×1D** Ebene definierten Gruppierungseigenschaften hin untersucht. Dabei wird innerhalb einer Menge O_i die Bearbeitungsreihenfolge wiederum durch die analytische Länge

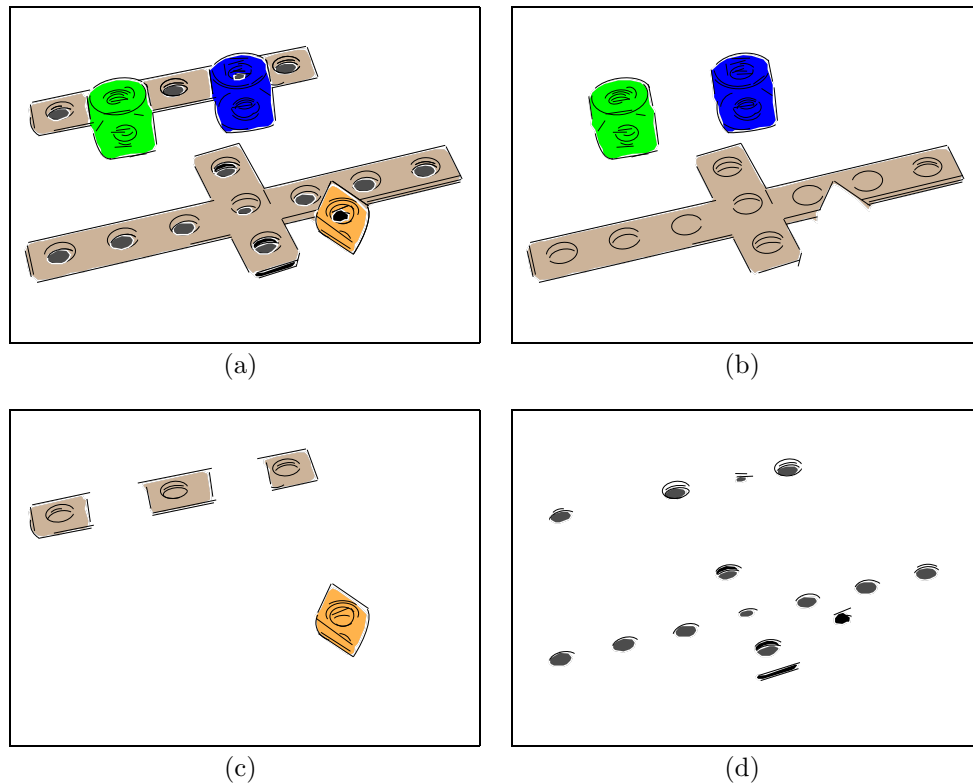


Abbildung 5.5.: Selektion der zu gruppierenden Konturmodelle für die $1D$ und $2 \times 1D$ -Ebenen. Für das Beispiel aus Abbildung 5.1 wurden hier die initialen Konturen den Regionen zugeordnet. Die Bearbeitungsreihenfolge ergibt sich in diesem Beispiel aus der Regionengröße (Anzahl der homogenen Bildpunkte innerhalb der Region). (a) Alle zugeordneten Konturmodelle, (b)–(d) Zuordnung zu den einzelnen Regionen (siehe Text).

$l(g)$ der Konturmodelle vorgegeben. Somit wird das Bild durch die Regionen in *Regions Of Interest* (ROI) eingeteilt, und die Konturen werden zunächst nur in diesen Bereichen gruppiert. Potentielle Gruppierungspartner g_k eines aktuell betrachteten Konturmodells $g_{cur} \in O_i$ müssen jedoch nicht notwendigerweise Element der aktuell Menge O_i sein. Dies motiviert sich durch die Tatsache, daß Konturen die einer Region zugeordnet sind durchaus zu globaleren Strukturen beitragen können. Zudem ist keine Information über die Korrektheit einer Region vorhanden. Sie kann auch durch eine Fehlsegmentierung oder Fehlinterpretation des Bildinhaltes entstanden sein. Demgegenüber haben die korrekten konturbasierten Strukturen im Bild möglicherweise eine größere räumliche Ausdehnung als die der Region.

Aufmerksamkeit von Regionen

Neben der regionenbasierten Einteilung der zu untersuchenden Konturen, fehlt nun noch die Angabe einer Reihenfolge, in der die einzelnen Mengen O_i abgearbeitet werden sollen. In Abbildung 5.5 ist eine Verarbeitungsreihenfolge für das Beispiel aus Abbildung 5.1 dargestellt. Hier erfolgt die Auswahl aufgrund der Regionengröße. Zuerst werden Konturen für die gekreuzten Holzleisten in Kollinearitäten, Kurvilinearitäten, Proximitäten und Parallelitäten gruppiert. Daran anschließend erfolgt die Betrachtung der Konturen für den grünen Würfel und der Konturen für den blauen Würfel (siehe Abbildung 5.5(b)). Abschließend werden die Konturen für die Raute, für die Teilregionen der verdeckten Leiste und für die Löcher in den Leisten betrachtet. Mit dieser Reihenfolge wird also der Fokus der Hypothesisierung sukzessive auf dominante und weniger dominante Regionen im Bild gelegt.

Eine weitere Art die Reihenfolge der zu bearbeitenden Mengen O_i festzulegen, kann durch die Berechnung eines Aufmerksamkeitsfaktors für jede Region erfolgen. In [Os98] wird beispielsweise für jede Region r_i ein Aufmerksamkeitsfaktor aus fünf Einzelmerkmalen berechnet. Die Wahl der einzelnen Merkmale motiviert sich dabei aus Resultaten von Fokussierungsexperimenten. Für jede Region werden Eigenschaften wie Kontrast zu benachbarten Regionen, Größe, Formfaktoren, Lokalisation im Bild und Beitrag zum Hintergrund ermittelt und zu einem gemeinsamen Wert kombiniert, der wiederum eine Aussage über die Prägnanz der Region treffen soll. Auf Basis dieses Wertes könnte so eine Reihenfolge der zu bearbeitenden Mengen O_i bestimmt werden.

Bei den zuvor diskutierten Auswahlstrategien wird die Reihenfolge der zu bearbeitenden Konturmodelle speziell durch die Ergebnisse einer Regionensegmentierung geleitet. Mit dem Konzept von Matches kann aber auch eine Unterstützung durch Regionen erreicht werden, die aus höheren Analyseschritten zur Verfügung gestellt werden. So kann die Identifikation von Objektbereichen durch eine Objekterkennung ebenfalls als Hinweis genutzt werden, zuerst Konturmodelle innerhalb dieser Bereiche zu gruppieren. Eine ähnliche Möglichkeit besteht, wenn durch Interaktion Bereiche perzeptiver Aufmerksamkeit durch einen Menschen vorgegeben werden. In [Rae00, Rae99] wird ein solches System beschrieben, dessen Bereiche perzeptiver Aufmerksamkeit hier eingesetzt werden könnten. In der vorliegenden Arbeit wird jedoch die Auswahl über die Regionengröße bestimmt.

5.3.2. Inkrementelles perzeptives Gruppieren

Mit dem Verfahren zur Selektion von Konturmodellen aufgrund einer Regionenzuordnung kann in idealer Weise die Fokussierung des Generierungsprozesses zunächst auf potentiell relevante Strukturen ausgerichtet werden. Dieses wird in dem hier beschriebenen Gesamtansatz eingesetzt, um bei einer Verarbeitung von Bildsequenzen möglichst frühzeitig Gruppierungshypothesen von prägnanten Strukturen im Bild für nachgelagerte Analyseprozesse zur Verfügung zu stellen. Da eine regionenbasierte Segmentierung mit einer pixelweisen Klassifikation wesentlich schneller für ein Bild zur Verfügung steht, als die Segmentierung und Approximation von Konturmodellen, wird hier insbesondere

die Information aus der Regionensegmentierung für eine Fokussierung genutzt.

1D und 2×1D Gruppierungen: Mit dem im letzten Abschnitt beschriebenen Verfahren können nun initiale Konturmodelle in Bildbereichen homogener Farbe für die Generierung von Gruppierungshypothesen der **1D** und **2×1D** Ebene sukzessive entsprechend der Bereichsgröße gewählt werden. Prinzipiell erfolgt die Generierung wie in Abschnitt 4.2.2 beschrieben, jedoch wird der Prozeß auf die jeweiligen homogenen Bereiche begrenzt. Sobald die Hypothesisierung für einen Bereich abgeschlossen ist, werden alle generierten Gruppen an das nächste Gruppierungsmodul weitergeleitet. Weniger wichtige Strukturen werden sukzessive generiert, bis der gesamte Bildbereich bearbeitet ist. Dabei bleiben bis zur vollständigen Bearbeitung des Bildes, alle bisher generierten Hypothesen erhalten, da auch aus anderen, später zu bearbeitenden Bereichen des Bildes noch Gruppierungen mit bereits vorher generierten linearen Gruppen entstehen können.

2D Gruppierungen: Im Modul zur Generierung von geschlossenen Figuren wird in ähnlicher Weise verfahren. Stehen neue Gruppierungen der **1D** und **2×1D** Ebenen für einen Bereich zur Verfügung, werden diese sofort in den Proximitätsgraphen eingefügt und eine neue Pfadsuche beginnt (siehe Abschnitt 4.3). Eine günstige Kantenwahl als Startpunkt für eine neue Suche ermöglicht, daß nicht wiederholt der gesamte Proximitätsgraph nach Pfaden durchsucht werden muß und die Gefahr von identischen Pfaden vermieden werden kann. Eine neue Pfadsuche wird hierzu jeweils nur von den neu hinzugekommenen Gruppierungs- oder nun auch von den Proximitätskanten aus gestartet. So können nur Pfade entstehen, die bisher noch nicht detektiert wurden. Das generelle Problem der Redundanz von einzelnen Pfaden wird damit zwar nicht verringert (das Ergebnis für das gesamte Bild bleibt identisch), aber geschlossene Figuren, die bereits vorher entstanden sind, haben eine etwas höhere Evidenz eine korrekte Interpretation zu sein, da sie auf Elementen aus einem oder mehreren vorher bearbeiteten Bereichen beruhen.

Dieses inkrementelle Vorgehen kann einfache Strukturen sehr schnell zur Verfügung stellen, wohingegen komplexere Strukturen (größere, mehrere Bereiche umfassende Strukturen) erst nach und nach erzeugt werden. Höheren Verarbeitungsmodulen steht damit schneller Teilinformation zur Verfügung, mit der sie ihrerseits Teilergebnisse erzeugen können. Die Gruppierungen der **N×2D** Ebene werden in dieser inkrementellen Arbeitsweise noch nicht berücksichtigt. Prinzipiell verschließen sie sich diesem Vorgehen aber nicht. Allerdings muß dabei beachtet werden, daß sie in der Regel auf einem sehr hohen Strukturierungsgrad aufsetzen, was häufig mit einer bereits vollständig abgeschlossenen Analyse der vorgeschalteten Module gleichzusetzen ist. Gleiches gilt für den in Abschnitt 4.4.5 angesprochenen Ansatz zur rekursiven Gruppierung komplexer Strukturen.

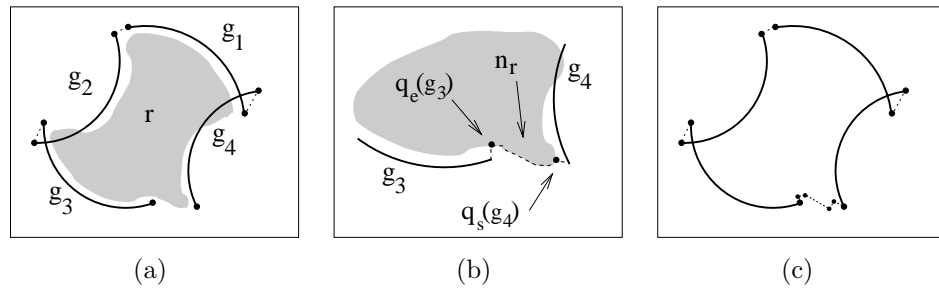


Abbildung 5.6.: Illustration der Integration von Regioneninformation in den **2D** Gruppierungsprozeß. (a) Synthetisches Beispiel, für das die Generierung einer geschlossenen Figur mißlingt. (b) Ausschnittvergrößerung im Bereich der Projektion der Gruppierungsendpunkte von g_3, g_4 auf die Region. Die Verwendung der Matchinformation führt zu einer Verbindung der linearen Gruppen g_3, g_4 durch eine regionenbasierte Proximität n_r . (c) Geschlossene Figur, die aufgrund der Nutzung von n_r generiert wird.

5.3.3. Regionenbasierte Proximitäten

Grundlegende Voraussetzung für den Erfolg einer konturbasierten Bildinterpretation ist die Detektion von hinreichend vorhandenen und entsprechend ausgeprägten Diskontinuitäten. Mit dem Verfahren zum perzeptiven Gruppieren können Fragmentierungen, die auf fehlende Diskontinuitäten zurückgeführt werden können, häufig aufgelöst werden. Wie bereits im Beispiel in Abbildung 5.2 angedeutet, kann es jedoch zu Fragmentierungen kommen, die von ihrer Ausdehnung her zu groß sind, als daß sie vom Gruppierungsprozeß überbrückt werden könnten. Im folgenden wird im Detail diskutiert, wie diese problematischen Bereiche durch die Unterstützung von Regioneninformation trotzdem für den Gruppierungsprozeß nutzbar gemacht werden.

In der **1D** Ebene der Gruppierungshierarchie werden Konturmodelle aufgrund von Nähe und guter Fortsetzung gruppiert. Die hierzu eingesetzten Aufmerksamkeitsbereiche sind aus einer Stichprobe ermittelt worden, um ein aus der menschlichen Wahrnehmung motiviertes Gruppierungsverhalten zu modellieren und verfahrensbedingte Fragmentierungen und mögliche Verdeckungen auflösen zu können. In der **2D** Ebene werden die dabei entstandenen linearen Gruppen und Proximitäten genutzt, um geschlossene Figuren zu beschreiben. Die Ermittlung dieser Geschlossenheiten wird jedoch verhindert, wenn eine zu große Fragmentierung entstanden ist und keine Proximitätsgruppierungen generiert werden konnten. In Abbildung 5.6(a) ist für ein synthetisches Beispiel eine solche Situation dargestellt. Für die linearen Gruppen werden die Proximitäten $g_4 \cdot g_1, g_1 \cdot g_2, g_2 \cdot g_3$ im **1D** Gruppierungsprozeß generiert. Die Nähegruppierung $g_3 \cdot g_4$ wird nicht generiert, da die Gruppierung g_3 die eigentliche Kontur nicht vollständig beschreibt. Ursache sind beispielsweise schwach ausgeprägte Diskontinuitäten. Aufgrund

der Orientierung und Größe der Proximitäts-Suchbereiche für g_3 und g_4 können die entsprechenden Endpunkte von g_3 und g_4 gegenseitig nicht als nah gefunden werden. Resultierend daraus scheitert die Suche nach einem geschlossenen Pfad für $g_1 \circ g_2 \circ g_3 \circ g_4$. Im Inneren der Figur ist die Region r dargestellt, die mit einer Regionensegmentierung generiert wird. Alle vier Gruppierungen werden mit dem Konzept der Matches dieser Region als umrissbeschreibend zugeordnet. Mit dieser Konstellation kann folgende Strategie verfolgt werden, um die offensichtliche geschlossene Figur zu ermitteln.

Im Verfahren zur Generierung von **2D** Gruppierungen (siehe Abschnitt 4.3) wird beim Aufbau des Proximitätsgraphen nun auch die Match-Information einer Gruppierung berücksichtigt. Hierzu wird für jede Region r_i eine Liste O_i aufgebaut, in der die zu r_i zugeordneten umrissbeschreibenden linearen Gruppen einsortiert sind. Die Reihenfolge der linearen Gruppen in O_i entspricht einem Umlauf gegen den Uhrzeigersinn um die Region. Für diese Sortierung werden die senkrechten Projektionen der Gruppierungsendpunkte $q_{s,e}(g_j)$ auf das Regionenspolygon $poly(r_i)$ verwendet. Eine Abfolge der Punkte $q_s(g_j), q_e(g_j)$ auf dem Regionenspolygon im Sinne der Umlaufrichtung wird dabei gewährleistet (vergleiche Abbildung 5.6(b)). Nachdem alle Gruppierungs- und Proximitätskanten in den Proximitätsgraphen eingefügt sind, wird für jede Gruppierungskante überprüft, ob fehlende Proximitätsbeziehungen zu benachbarten linearen Gruppen durch Ausnutzung der Regioneninformation generiert werden können. Hierzu wird jede Liste O_i betrachtet, in der eine lineare Gruppe g_j enthalten ist. Ausgehend von der Gruppe g_j werden die nächsten benachbarten linearen Gruppen $g_n \in O_i$ auf eine bereits vorhandene Gruppierungswirkung untersucht. Folgende Bedingungen werden evaluiert:

- Eine Proximitätsbeziehung $g_j \cdot g_n$ aus dem **1D** Gruppierungsprozeß zwischen den direkt benachbarten Endpunkten von g_j, g_n besteht nicht.
- Die Gruppierungen g_j, g_n sind nicht Bestandteile *einer* linearen Gruppe.
- Die Gruppierungen g_j, g_n verbindet keine Bestandteilbeziehung:

$$g_j \notin GElem(g_n) \quad \wedge \quad g_n \notin GElem(g_j) \quad (5.3)$$

Sind diese Bedingungen erfüllt, wird über die am nächsten benachbarten Gruppierungsendpunkte und deren Projektionspunkte eine Proximitätsrelation zwischen den Gruppierungen entlang der Regionengrenze generiert. Die Relation wird dabei immer zwischen dem Projektionsendpunkt q_e der Gruppierung g_j und dem Projektionsstartpunkt q_s der nächsten Gruppierung g_n in O_i etabliert. Die entstehende Proximität wird im weiteren als eine *regionenbasierte Proximität* n_r bezeichnet (siehe Abbildung 5.6(b)). Um übermäßig lange und inkorrekte regionenbasierte Proximitäten zu vermeiden, wird zusätzlich die Länge $l(n_r)$ folgendermaßen eingeschränkt:

- Bilden die Projektionsabschnitte der verbundenen Gruppierungen auf der Region r_i einen nicht leeren Schnitt

$$q_{intersect} = [q_s(g_j), q_e(g_j)] \cap [q_s(g_n), q_e(g_n)] \neq \emptyset,$$

darf eine regionenbasierte Proximität nicht länger als der maximale Durchmesser $d_{max}(l(g))$ einer skalierten Suchbereichsmaske für Proximitäten sein. Die Skalierung $1/\tau'$ des Durchmessers erfolgt analog zu der Skalierung des Suchbereichs im **1D** Gruppierungsprozeß (siehe Abschnitt 4.2.2, Formel (4.3)), wobei die längere der beiden verbundenen linearen Gruppen als Parametrisierung gewählt wird.

$$l(n_r) < \frac{1}{\tau'} d_{max}(\max(l(g_j), l(g_n))) \text{ falls } q_{intersect} \neq \emptyset \quad (5.4)$$

Diese Einschränkung motiviert sich durch die Annahme, daß die überlappenden Gruppen im **1D** Gruppierungsprozeß möglicherweise nur von einer der beiden linearen Gruppen als potentielle Gruppierungspartner für eine Proximität gefunden werden (vergleiche Abschnitt 4.2.2, Seite 64). Nun liegt mit der Regionengrenze zusätzliche Evidenz für die Korrektheit einer solchen Gruppierung vor. Dies gilt jedoch nur, wenn die Länge der regionenbasierten Proximität die Ausdehnung der ursprünglichen Proximitätsmaske nicht überschreitet.

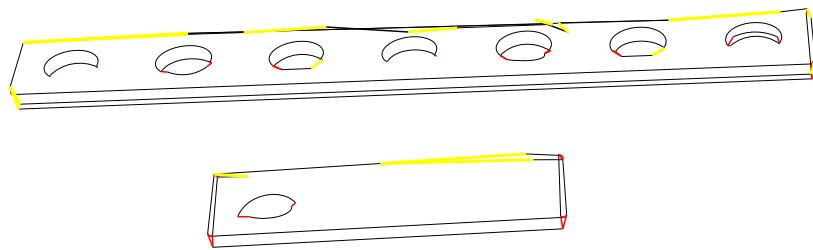
- Liegt keine Überlappung der Projektionsintervalle von benachbarten Gruppierungen vor, so wird $l(n_r)$ durch den Umfang $U(r)$ von r eingeschränkt:

$$l(n_r) < \Delta_{reg-prox} U(r) \text{ falls } q_{intersect} = \emptyset \quad (5.5)$$

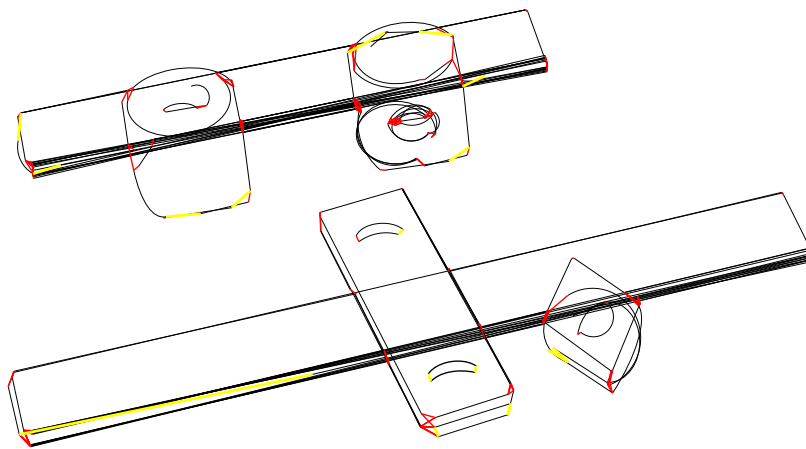
Der Parameter $\Delta_{reg-prox}$ wurde empirisch zu 0.15 ermittelt. Damit skaliert die Einschränkung der Ausdehnung von n_r zwischen linearen Gruppen mit dem Umfang der Region.

Die unterschiedliche Betrachtung der einschränkenden Längen zwischen überlappenden und nicht überlappenden Projektionsintervallen motiviert sich aus der Anschauung: Überlappende Projektionsbereiche deuten im Wesentlichen daraufhin, daß Orientierung und Skalierung der Suchbereichsmasken die Hypothesierung einer Proximitätsgruppierung im **1D** Gruppierungsprozeß verhindern. Diese kann aber nun aufgrund der zusätzlichen Evidenz durch die Regionengrenze generiert werden. Nicht überlappende Bereiche lassen auf Fehler in der Segmentierung schließen, die auch vom Gruppierungsprozeß nicht aufgehoben werden können. Eine Regionensegmentierung hingegen führt zu einer korrekten Interpretation. Somit kann die Regionengrenze bis zu einem gewissen Grad als ergänzende Information eingesetzt werden. Diese Wahl der einschränkenden Bedingungen hat sich in Experimenten als eine gute Modellierung für die Beschränkung der Ausdehnung von regionenbasierten Proximitäten bewährt.

Die Generierung von regionenbasierten Proximitäten erfolgt *nicht* zwischen allen zugeordneten linearen Gruppen einer Region. Es werden ausgehend von g_j nur solange die nächsten Nachbarn aus O_i untersucht, bis auf der jeweiligen Seite von g_j eine Gruppierung g_n gefunden wird, für die eine der eingangs erwähnten Bedingungen nicht erfüllt ist, oder bis zwischen g_j und g_n eine regionenbasierte Proximität generiert werden kann. Diese Verfahrensweise ist letztendlich der Schlüssel, um übermäßig lange und inkorrekte



(a)



(b)

Abbildung 5.7.: Initiale Proximitätsgraphen zur Hypothesierung von geschlossenen Figuren für die Beispiele aus (a) Abbildung 5.2 und (b) Abbildung 5.1. Lineare Gruppen sind schwarz dargestellt. Proximitätsgruppierungen aus dem **1D** Gruppierungsprozeß sind rot markiert, während die regionenbasierten Nähegruppen gelb markiert sind.

regionenbasierte Proximitäten zu vermeiden. In Abbildung 5.6(c) ist die so eingefügte Proximität und die generierte Geschlossenheit für das Beispiel abgebildet.

In Abbildung 5.7 sind die initialen Proximitätsgraphen für die Beispiele aus Abbildung 5.2 und Abbildung 5.1 mit eingefügten regionenbasierten Proximitäten dargestellt. In Abbildung 5.7(a) ist gut ersichtlich, daß die problematischen Konstellationen aus Abbildung 5.2 im Bereich der oberen Leistenbegrenzungen durch die Integration der Regioneninformation aufgelöst werden können. Für beide Leisten können geschlossene

Figuren generiert werden. Auch für das Beispiel aus Abbildung 5.1 können zusätzliche regionenbasierte Proximitäten ein deutlich verbessertes Gruppierungsverhalten erzielen (siehe Abbildung 5.7(b)). Im Bereich der Würfel unterstützen sie die Hypothesisierung von Geschlossenheiten für die Würfel, die sonst nicht möglich waren, und auch für die Rautenmutter im Vordergrund können korrekte zusätzliche Geschlossenheitshypothesen aufgestellt werden. Ein weiterer interessanter Aspekt ist in Abbildung 5.7(b) zu beobachten. Im Bereich der Strukturen für die gekreuzten Holzleisten werden im unteren Teil der kürzeren Leiste regionenbasierte Proximitäten für die beiden kleinen parallel gruppierten Konturstücke gebildet (gelb markiert). Hier unterstützt die regionenbasierte Information die Hypothesisierung einer geschlossenen Figur für rein symmetrische Strukturen (zwei parallele kurze Geradenstücke), die zwischen ihren Bestandteilsendpunkten keine Näheinformation besitzen. Das gleiche Phänomen läßt sich oberhalb davon auch bei zwei symmetrischen Ellipsenbögen innerhalb der Struktur für die kürzere Leiste beobachten.

Mit der Beschreibung dieser Beispiele, ist die realisierte Integration von Regioninformation in den konturbasierten Gruppierungsprozeß vollständig dargestellt. Neben der Unterstützung in der Auswahl der zu betrachtenden Konturmodelle und der damit verbundenen Möglichkeit zur inkrementellen Verarbeitung, konnte insbesondere dargelegt werden, daß mit der Integration von Regioninformation die Hypothesisierung zusätzlicher konturbasierter Gruppierungen ermöglicht wird. Im folgenden Abschnitt wird nun die Unterstützung der Konturinformation für eine hybride Objekterkennung vorgestellt.

5.4. Objekterkennung durch Integration

Ein Ziel der Integration von konturbasierten Gruppierungen und Regioninformation ist die Unterstützung und Verbesserung einer Objekterkennungsleistung in einem automatischen Bildanalyse-System. Nach verbreiteter Expertenmeinung ist die Fokussierung auf nur eine Segmentierungsstrategie für eine Objekterkennung nicht ausreichend. Dies wurde bereits im einleitenden Abschnitt dieses Kapitels dargestellt.

In diesem Abschnitt wird nun eine Strategie vorgestellt, die in einem hybrid angelegten Objekterkennungssystem durch Berücksichtigung von Konturinformation zu robusteren Erkennungsleistungen führt. Zunächst wird ein kurzer Überblick über das hybride Erkennungssystem gegeben, bevor beschrieben wird, wie die Integration von Kontur- und Regioninformation die Erkennung verbessert.

5.4.1. Hybride Objekterkennung

Der im folgenden beschriebene hybride Objekterkenner wurde in [Kum97, Kum98b] zur Erkennung von *baufix*[®] Objekten vorgestellt. In Abbildung 5.8 ist der Systemaufbau dargestellt, dessen zentrale Idee auf der Verwendung eines hybriden semantischen Netzwerkes beruht.

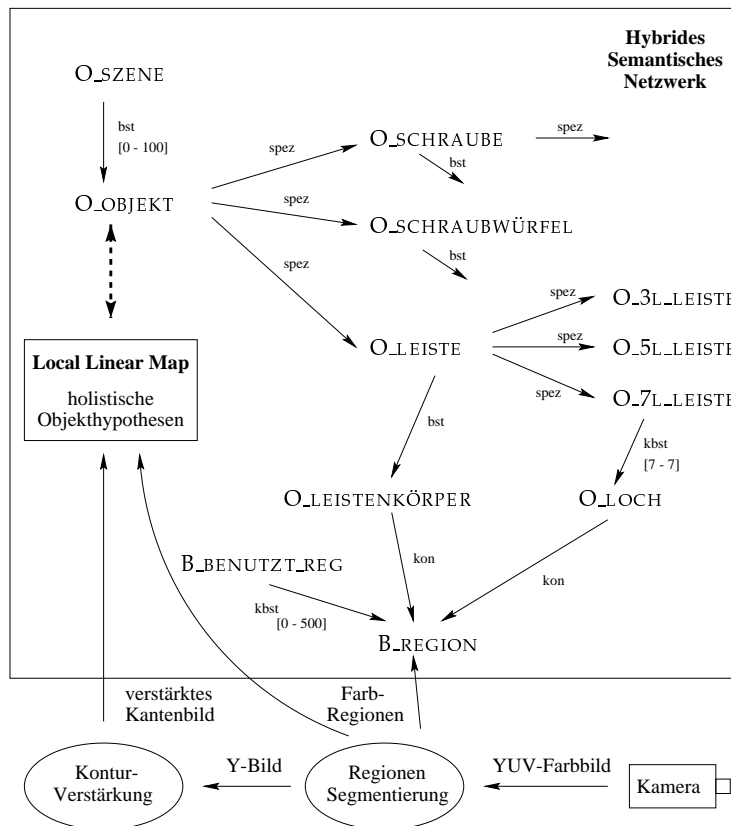


Abbildung 5.8.: Systemaufbau der 2D-Objekterkennung, in Anlehnung an [Kum97].

Als primäre Eingangsdaten für das System werden die Regionen aus der Segmentierung mit pixelweiser Klassifikation (vergleiche Abschnitt 2.2.1) verwendet. Auf Basis dieser Regionen, die im Wesentlichen vollständige Objektregionen für die *baufix*[®] Domäne beschreiben, werden mit Hilfe eines neuronalen Netzansatzes, den ‘Local Linear Maps’ (LLM), 2D-Objekthypothesen aufgestellt (für Details hierzu siehe [Hei98]). Die grundlegende Idee dieses Systemteils beruht auf der Berechnung eines 16-dimensionalen Merkmalsvektors an den Schwerpunkten der Region, der dann als Eingabe für das LLM dient. Hierzu wird aus dem Intensitätskanal des Bildes durch Laplacefilterung und durch eine logarithmische Intensitätstransformation ein kantenverstärktes Bild gewonnen, welches an den Regionenschwerpunkten mit 16 Gaborfiltern skalar multipliziert wird. Als Ausgabe liefert das LLM eine holistische Objekthypothese für den Merkmalsvektor oder weist die Region als *unbekanntes* Objekt zurück.

Den Rahmen für die Realisierung einer hybriden Objektrepräsentation bildet in diesem System die semantische Netzwerksprache ERNEST [Kum98a]. Der folgende kurze Überblick lehnt sich an die Darstellung in [Hei99] an. Die Repräsentation von Objektklassen wird in ERNEST durch den Netzknontyp *Konzept* abgebildet. Diesen Netz-

knoten werden Attribute zugeordnet, die qualitative und quantitative Merkmale und deren Einschränkungen für das zu modellierende Konzept beschreiben. Im **baufix**[®] Szenario sind dies im Wesentlichen die Formeigenschaften der Farbregionen die ein Objekt repräsentieren. Aus diesen *Konzept-Knoten* können *Instanz-Knoten* werden, wenn Instanzen zu allen erforderlichen Teilen eines Konzeptes aus den Eingangsdaten ermittelt worden sind. Mögliche Zwischenzustände entstehen, wenn aufgrund fehlender einzelner Teile noch keine vollständige Instantiierung erfolgen kann. Diese Zwischenzustände werden durch den Knotentyp *modifiziertes Konzept* repräsentiert, der die Eigenschaften eines Konzeptes bereits aufgrund vorhandener Daten einschränkt, wobei die bisherige Analyse jedoch noch nicht ausreicht, um eine vollständige Instantiierung zu ermöglichen. Die Modellierung einer Wissensbasis in ERNEST erfolgt über die Verbindung von Konzepten durch *Bestandteils-*, *Spezialisierungs-* und *Konkretisierungskanten*. Mit Hilfe einer auf dem A*-Algorithmus basierenden problemunabhängigen Kontrollstrategie wird die so definierte Wissensbasis zur Objekterkennung angewendet. Im folgenden wird in Anlehnung an [Kum97] beispielhaft der Ablauf einer Analyse beschrieben.

Zu Beginn einer Analyse wird eine holistische Instantiierung von O_OBJEKT² durchgeführt, wodurch eine Aktivierung des LLM-Netzwerks ausgelöst wird. Dieser holistische Erkenner liefert für den ersten Regionenschwerpunkt bis zu drei konkurrierende Objekthypothesen, die in jeweils einer holistischen Instanz O_OBJEKT_i^(H) abgelegt werden. In Abhängigkeit vom Typ der holistischen Objekthypothese wird daraufhin eine Spezialisierung auf ein bestimmtes Baufixobjekt durchgeführt. Zur Verifikation der erzeugten holistischen Objekthypothesen wird das strukturelle Wissen des semantischen Netzwerks aktiviert. Dies führt zu modellgetriebenen Bindungen, die sich aus der Repräsentation eines Objektes durch Bestandteile und Konkretisierungen ergeben. Es entsteht ein Suchbaumknoten des A*-Algorithmus, der beispielsweise zur Modellierung einer Leiste eine Bestandteilkante zu einem Leistenkörper hat, der dann wiederum durch eine Region eine Konkretisierung beinhaltet (siehe Abbildung 5.9(a)). Damit muß nun eine Region zur Instantiierung gewählt werden, deren Schwerpunkt im Bereich einer LLM-Hypothese liegt, holzfarben ist und den Einschränkungen bezüglich der Formparameter eines Leistenkörpers genügt. Darüber hinaus darf sie für kein anderes Objekt verwendet worden sein. Um dies zu gewährleisten, wird nach der Instantiierung von B_REGION₁^(M) auch eine neue Instanz des Konzeptes B_BENUTZT_REG generiert, wodurch alle bereits benutzten Regionen durch Instanzen dieses Konzeptes gesammelt und gespeichert werden. Resultierend aus der Instantiierung von B_REGION₁^(M) ist nun eine Instantiierung von O_LEISTENKÖRPER₁^(M) und O_LEISTE₁^(M) und eine Spezialisierung auf Drei-, Fünf- bzw. Siebenlochleiste möglich. Eine Verifikation dieser partiellen Instanzen erfolgt modellbasiert, indem passende Regionen für die Löcher instantiiert werden (siehe Abbildung 5.9(b) für eine Instanz zu Dreilochleiste). Nach der vollständigen Instantiierung eines 2D-Objekts wird dieses als Bestandteil von O_SZENE etabliert. Konkurrierende In-

²Konzepte der 2D-Objektebene sind durch das Präfix O_ und Konzepte der Bildebene durch das Präfix B_ gekennzeichnet.

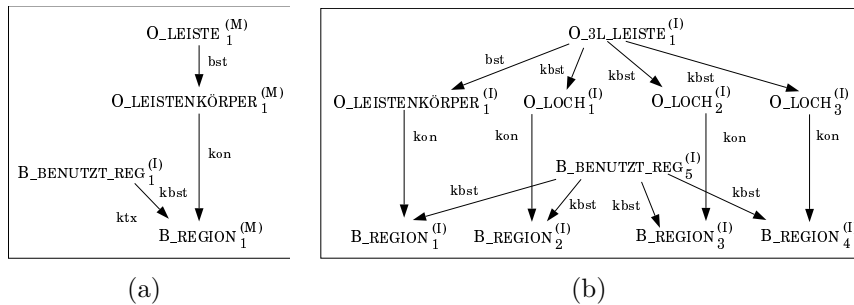


Abbildung 5.9.: Netzstruktur des hybriden semantischen Netzes während einer Analyse (vergleiche [Kum97]): (a) Suchbaumknoten, der aus der Aktivierung des LLM-Netzwerks entsteht, und zu einer Instantiierung einer Bildregion für das Konzept Leistenkörper führen soll. (b) Netzstruktur nach der Instantiierung eines Konzepts Drei-Lochleiste mit entsprechenden Bestandteilen und Spezialisierungen.

stanzen von Objekten werden aufgrund einer Bewertung aus der strukturellen Analyse eliminiert, und es bleiben nur die bestbewerteten Objekthypothesen für die Interpretation erhalten. Regionen, deren zugehörige LLM-Hypothese vom semantischen Netz nicht verifiziert werden konnte oder zu denen keine LLM-Hypothese generiert wurde, werden abschließend rein wissensbasiert analysiert.

Dieses System ist von der Einzelbildverarbeitung auf eine Analyse in Bildfolgen erweitert worden. Mit dieser Erweiterung ergibt sich die Möglichkeit eine temporale Zuordnung von Objekthypothesen über die Zeit zu generieren, und resultierend daraus eine Effizienzsteigerung und eine Detektion von Aktionen oder Handlungen abzuleiten. Die *schritthaltende* Verarbeitungsstrategie beruht im Wesentlichen darauf, daß Ergebnisse zum Zeitpunkt $t - 1$ eine Erwartungshaltung für die Analyse zum Zeitpunkt t aufbauen. Hierzu wird zunächst ein Regionengedächtnis eingesetzt, das aufgrund von Einschränkungen bezüglich Abweichungen des Regionenschwerpunktes, der Farbe und der Formparameter eine Identifizierung von identischen Regionen über eine Bildfolge hinweg leistet. Durch diese eindeutige Identifizierung der Eingangsdaten über die Zeit kann die zu früheren Zeitpunkten aufgebaute Instanzstruktur des hybriden semantischen Netzes übernommen werden, wobei die Regionen vom Zeitpunkt t ihre veralteten identischen Regionen vom Zeitpunkt $t - 1$ in den Instanzen $B_REGION_i^{(I)}$ ersetzen. Mit dieser Vorgehensweise können durch Ausnutzung der vorhandenen Netzstruktur sofort Objekthypothesen zum Zeitpunkt t berechnet bzw. verifiziert werden. Für weitere Details zu dieser schritthaltenden Erkennung sei auf [Kum97, Kum98b] verwiesen.

5.4.2. **Objektfokussierte Kontur-Integration**

Im letzten Abschnitt wurde die rein regionenbasierte Objekterkennung auf Bildsequenzen mit einem hybriden semantischen Netz vorgestellt. Dieses Verfahren schlägt in Situationen fehl, in denen strukturelle Information über den Bildinhalt für eine Interpretation erforderlich ist. Solche Situationen sind bereits zu Beginn von Abschnitt 5.2.1 ausführlich an Beispielen diskutiert worden (Verdeckungen oder Verschmelzung gleichfarbiger Objekte). Das Ziel einer objektfokussierten Kontur-Integration ist es nun, die datengetriebene strukturelle Information, die in dem konturbasierten Gruppierungsansatz gewonnen wird, dem Objekterkennungssystem als ergänzende Information zur Verfügung zu stellen. Mit dem Prinzip der Matches existiert hierzu ein Verfahren, das einen entscheidenden Schritt in diese Richtung leistet, da es auch für die Zuordnung von Konturgruppierungen zu Objektregionen genutzt werden kann. Die Integration der strukturellen Information kann auf zwei Arten erfolgen:

- Bevor die Regionen in dem hybriden Objekterkennungssystem verwendet werden, werden ihnen die entsprechenden umriss- und strukturbeschreibenden Konturgruppierungen zugeordnet. In dem semantischen Netz werden zusätzliche Attribute an den Netzknoten eingetragen, die diese zusätzliche Konturinformation auswerten und Einschränkungen bezüglich der Analysestrategie (Instantiierung) ableiten. Dieses Vorgehen entspricht einer Erweiterung der Wissensbasis, die in dem semantischen Netz kodiert ist. Diese Integrationsstrategie wird zukünftig in das bestehende System aufgenommen werden.
- Ein anderer Ansatz verfolgt die Strategie, die Konturgruppierungen der **2D** Gruppierungsebene (geschlossene Figuren) als konturbasierte Regionen zu interpretieren und diese dem Objekterkennungssystem als alternative Objektregionen zur Verfügung zu stellen. Hierzu wird eine *Top-Down* Strategie eingesetzt, die aus den Informationen des höheren Interpretationsmoduls eine fokussierte Bearbeitung der Bilddaten in den unteren Ebenen unterstützt. Dieser Ansatz erweitert die schrittweise Objekterkennung auf Bildsequenzen und wird im folgenden detailliert beschrieben.

Konturbasierte Regionen für unbekannte Objekte

Der hybride Objekterkennungssystem ist für die Erkennung von einzelnen, überwiegend vollständig sichtbaren Objekten ausgelegt. Verdeckungen von Objekten können nur bis zu einem gewissen Grad, Verschmelzungen von Objektprojektionen können hingegen überhaupt nicht aufgelöst werden. Lassen sich aus einem dieser Gründe keine Objekthypothesen für eine Farbregion generieren, so interpretiert das Erkennungssystem diese Region als *unbekanntes* Objekt. Genau an dieser Stelle setzt das hier entwickelte und in Abbildung 5.10 skizzierte Top-Down Integrationschema an (siehe auch [Sch00a]).

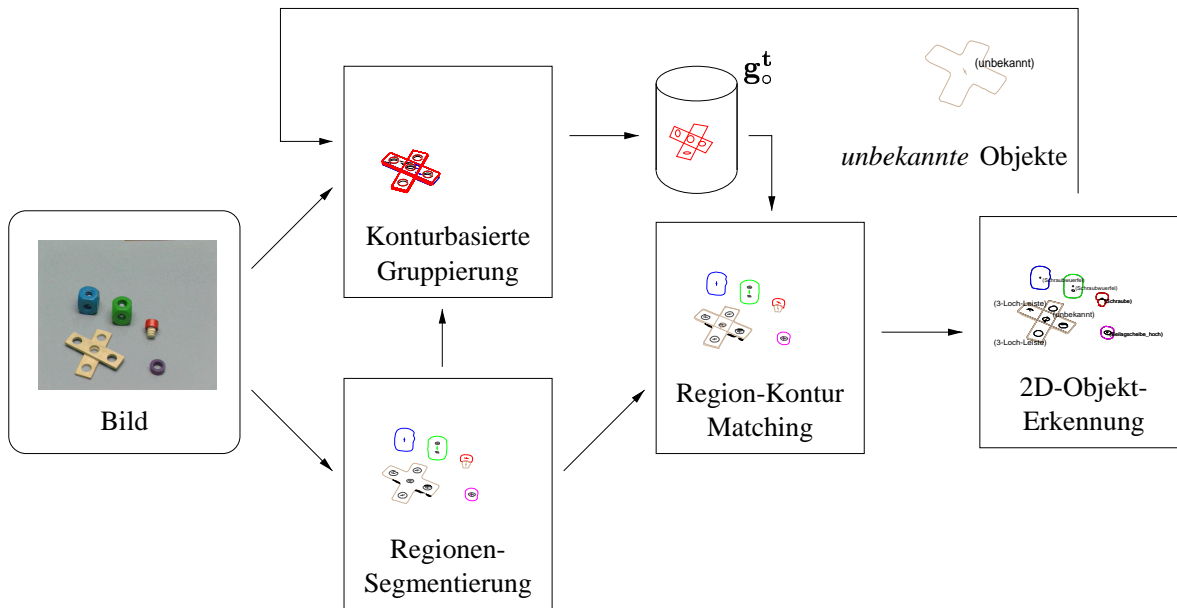


Abbildung 5.10.: Integrationsschema für die Erkennung von unbekanntem Objekten durch den Einsatz konturbasierter Geschlossenheiten und des Konzeptes der Matches.

Das Erkennungssystem liefert für einen Großteil der Farbreionen im Bild korrekte Objekthypothesen (Erkennungsrate $\approx 90\%$, [Kum98b]). Daher ist eine genaue strukturelle Untersuchung nur für Farbreionen erforderlich, die als unbekanntes Objekte klassifiziert werden. Als Konsequenz daraus wird die konturbasierte Interpretation nur für diese Objektbereiche im Bild vorgenommen. Hierzu werden die Regionen aus dem Erkennen dem Prozeß zur konturbasierten Interpretation zugeführt. Dort werden zunächst mit Hilfe einer Dilatation die unbekanntes Objektregionen etwas vergrößert, um die Projektion des Objektes im Bild vollständig zu bewahren. Anschließend werden sie zur Maskierung des Intensitätsbildes eingesetzt, um unnötige Berechnungen in bereits interpretierten Teilen des Bildes zu unterbinden. Nach einer initialen Kontursegmentierung werden die Konturmodelle mit dem perzeptiven Gruppieren strukturell organisiert, und mit dem in Kapitel 6 vorgestellten Verfahren bewertet³. Für den Gruppierungsprozeß werden dabei auch die in Abschnitt 5.3 beschriebenen Ansätze eingesetzt, die die Generierung der Gruppierungen durch Regioneninformation unterstützen. Die als strukturell signifikant bewerteten **2D** Gruppierungen werden abschließend den Farbreionen als struktur- oder umrissbeschreibend zugeordnet. Diese zugeordneten konturbasierten Regionen werden mit der Menge der Farbreionen aus der pixelweisen Klassifikation vereinigt und dem

³ Die Gruppierungen bleiben von ihrem Erscheinungsbild her identisch, ihnen wird nur ein Bewertungsmaß zur Unterstützung einer Auswahl hinzugefügt.

hybriden semantischen Netz zur Analyse übergeben. Da die konturbasierte Bearbeitung nur im Intensitätskanal des Bildes erfolgt, das Erkennungsmodul aber auch das Merkmal Farbe einer Region auswertet, wird die Farbe für die konturbasierten Regionen aus den zugeordneten Farbregionen übernommen. Die darüber hinaus erforderlichen Formparameter werden hingegen direkt auf der polygonalen Beschreibung der Geschlossenheiten berechnet. Damit stehen für die Objekterkennung Bildbeschreibungen aus zwei vollständig unabhängigen Interpretationskanälen auf Basis einer einheitlichen Repräsentation durch Polygone zur Verfügung. Eine Behandlung konkurrierender Regionen aus der Farbsegmentierung und dem Gruppierungsprozeß wird aktuell noch nicht durchgeführt. Dieses läßt sich aber in das Objekterkennungsmodul integrieren, indem beispielsweise eine Bewertung des Erkennungsergebnisses eingeführt wird und die Auswahl der bestbewerteten Objekthypothese zu einer Redundanzauflösung führt..

Ein entscheidendes Problem bei diesem Integrationschema sind die deutlich unterschiedlichen Bearbeitungszeiten der initialen Segmentierungsverfahren bzw. der zusätzlichen Gruppierung. Auf relativ einfachen Szenen beträgt das Verhältnis der Laufzeiten von Regionensegmentierung zum konturbasierten Gruppieren ca. 1 : 15 Zeitschritte. Jedoch wird auch mit dem integrierten System weiterhin eine schnelle schritthaltende Objekterkennung im Takt der Regionensegmentierung (ca. 200-800ms/Bild) angestrebt. Dieses kann nur erreicht werden, wenn die Module zur schritthaltende Objekterkennung nicht durch das Modul zum konturbasierten Gruppieren ‚ausgebremst‘ werden. Um diese Problematik in den Griff zu bekommen, ist daher folgende Strategie entwickelt worden: Zunächst wird das im letzten Abschnitt erwähnte Regionengedächtnis zur Identifikation identischer Regionen um ein weiteres Leistungsmerkmal ergänzt. Neben der Identifikation identischer Regionen auf einer Bildsequenz, wird nun für jede Region ein Stabilitätszeitpunkt bezüglich Form und Lokalisierung angegeben. Dieser Stabilitätszeitpunkt t_{stable} entspricht dem Aufnahmezeitpunkt des Bildes, in dem eine Region *zum ersten mal* mit ähnlicher Form und ähnlichem Schwerpunkt detektiert worden ist. Sobald sich diese Merkmale für eine Region stark ändern (Bewegung in der Szene) wird der Stabilitätszeitpunkt für die betreffende Region im Regionengedächtnis neu gesetzt. Neben dieser Erweiterung auf der Regionenseite, wird für jede Gruppierung der Aufnahmezeitpunkt t_{grab} des Bildes mitgeführt, welches zur Kontursegmentierung und Generierung der Gruppierung verwendet wurde. Mit diesen zusätzlichen Informationen können beide Interpretationskanäle die Bilder in dem durch die jeweilige Verarbeitungsdauer vorgegebenen Takt analysieren. Erst in dem Modul zur Zuordnung der Geschlossenheiten zu den Farbregionen wird diese Information genutzt, um eine Kombination von Bildprimitiva zu vermeiden, die völlig unterschiedliche, zeitlich versetzte Ereignisse im Bild beschreiben: Eine Zuordnung ist demnach nur zulässig, wenn der Stabilitätszeitpunkt einer aktuellen Region r älter ist, als der Bildaufnahme-Zeitpunkt der aktuell zuzuordnenden Geschlossenheit g :

$$t_{stable}(r) < t_{grab}(g) \quad (5.6)$$

Im Umkehrschluß bedeutet dies, daß Regionen, die während der Berechnung der Grup-

pierungen *neu* entstanden sind oder durch Manipulation in der Szene sich stark *verändert* haben, nicht für die Zuordnung zu den Gruppierungen herangezogen werden.

Mit obiger Einschränkung wird vermieden, daß Bildprimitiva einander zugeordnet werden, die unterschiedliche Ereignisse in der Szene beschreiben. Allerdings können mit dieser Vorgehensweise dem Objekterkennungssystem die ergänzenden strukturellen Informationen (konturbasierte Regionen) aufgrund der stark unterschiedlichen Verarbeitungszeiten nur ca. alle 15 Zeitschritte einmalig zur Verfügung gestellt werden. Um nun die Information auch während des Zeitraumes bereitstellen zu können, in dem keine aktuellen Gruppierungen vorliegen, wohl aber aktuelle Regionen berechnet werden, wird folgende Verfahrensweise vorgeschlagen: Es werden zunächst jeweils die zuletzt generierten Gruppierungen g^t vom Zeitpunkt t in dem Zuordnungsmodul zwischengespeichert \mathbf{g}_o^t (siehe Abbildung 5.10). Während des Zeitabschnittes $[t, t + n]$ (n bezeichne die Anzahl der Zeitschritte, die für die konturbasierte Analyse eines Bildes benötigt werden) werden dann diese Gruppierungen aus \mathbf{g}_o^t den jeweils aktuell berechneten Regionen zugeordnet, wobei die Bedingung (5.6) überprüft wird. Mit diesem Vorgehen können alternative konturbasierte Regionen für unbekannte Objekte solange schritthaltend generiert werden, bis neue aktuellere Gruppierungen vom Zeitpunkt $t + n$ berechnet sind.

Dieses Integrationsschema realisiert einen vielfach diskutierten Ansatz, möglichst einfach und schnell zu berechnende Interpretationskanäle zu verwenden, und nur in komplexeren Teilausschnitten zusätzliche Ressourcen bereitzustellen und entsprechend aufwendigere Analysen durchzuführen. Darüber hinaus wird eine Top-Down Verarbeitung realisiert, die aus der Objekterkennung eine fokussierte Bearbeitung komplexer Bildausschnitte durch das konturbasierte Gruppieren anstößt. Für das verwendete Erkennungssystem konnte so eine signifikante Steigerung der Erkennungsleistung bei komplexen Objektkonstellationen erreicht werden.

In Abschnitt 7.3 werden Ergebnisse dieser Integrationsstrategie auf einigen Beispielen vorgestellt und diskutiert. Zunächst wird jedoch im nächsten Kapitel die Bewertung von konturbasierten Gruppierungshypothesen mit einem Markov Random Field vorgestellt.

6. Bewertung und Selektion mit Markov Random Fields

In den vorangegangenen Kapiteln wurden Methoden vorgestellt, die eine Auflösung von fehlerhaften oder ungenügenden Interpretationen der Bilddaten unterstützen. Jedoch wird in dem dafür entwickelten konturbasierten Gruppieren zunächst nur eine relativ lokale Betrachtung der Konturmodelle und -gruppen sowie der unterstützenden Information durch Regionen vorgenommen. In diesem Kapitel wird nun eine Bewertung von konturbasierten Gruppierungshypothesen in einem globaleren Zusammenhang vorgestellt, so daß ein Hypothesensatz gefunden werden kann, der eine global konsistente konturbasierte Interpretation der Bilddaten bereitstellt. Diese Bewertung von Hypothesen erfordert jedoch die Kodierung und Anwendung von Wissen über die Relationen zwischen den verschiedenen Gruppierungstypen und entsprechenden Gruppierungshypothesen.

Eine häufig eingesetzte Form zur Abbildung von Wissen über Kontextabhängigkeiten innerhalb eines Systems ist ein Netzwerk. In Abschnitt 5.4.1 wurde bereits auf die Verwendung eines semantischen Netzwerkes zur Wissensrepräsentation für die **baufix**[®] Objektwelt kurz eingegangen. Weitere eingesetzte Netzwerke zur Modellierung von Wissen sind beispielsweise Bayes-Netzwerke [Pea88] und *Markov Random Fields* (MRFs) [Che93a, Li95]. In der vorliegenden Arbeit wird zur Modellierung des Wissens über Kontextabhängigkeiten und Relationen zwischen den Gruppierungshypothesen die Technik der MRFs eingesetzt. Diese Wahl begründet sich im Wesentlichen dadurch, daß in MRFs die Wechselwirkungen zwischen den Gruppierungshypothesen richtungsunabhängig modelliert werden können. Damit wirken Veränderungen instantan auf *alle* Elemente in einem lokalen Kontext und so zunehmend auch auf den globalen Kontext, in dem die Gruppierungshypothesen zueinander stehen.

Zunächst werden im folgenden Aufbau und Eigenschaften von MRFs und notwendige Optimierungsverfahren vorgestellt. Daran anschließend folgt ein kurzer Überblick der Einsatzgebiete von MRFs in der Bildanalyse, bevor die Modellierung des MRFs zur Bewertung von Gruppierungshypothesen erläutert wird. Die allgemeine Darstellung von MRFs und die Beschreibung des MRFs zur Bewertung von Gruppierungshypothesen ist an die Präsentationen in [Pos97, Sch95] angelehnt. Ein weiteres Einsatzgebiet für MRFs ist die Analyse von Bildfolgen. In Abschnitt 6.3 wird ein Verfahren untersucht, das mit Hilfe eines MRFs eine zeitliche Zuordnung von Kontursegmenten anstrebt. Damit sollen Konturmodelle identifiziert werden, die die Projektion *einer* realen Kontur in einer Bildfolge beschreiben. Ist eine solche Zuordnung möglich, so ergeben sich weitere Aspekte,

die eine schritthaltende Generierung *und* Bewertung von Gruppierungshypothesen auf Bildsequenzen in Aussicht stellen.

6.1. Markov Random Fields

Markov Random Fields bieten einen fundierten wahrscheinlichkeitstheoretischen Formalismus zur Modellierung von Kontextabhängigkeit zwischen Entitäten eines Systems. Motiviert durch Bemühungen um eine Modellbildung für physikalische ferro-magnetische Prozesse, entstanden sie als eine Modellierung für das Verhalten des *Ising Modells* [Kin80]. Auch auf dem Gebiet der Bildanalyse werden sie als Repräsentationsformalismus für verschiedenste Problemstellungen eingesetzt.

6.1.1. Aufbau und Eigenschaften

Allgemein läßt sich die Abbildung einer Problemstellung auf ein MRF in folgender Weise beschreiben: Zunächst wird eine Menge

$$\mathcal{S} = \{s_1, \dots, s_n\}$$

von *Stellen* definiert. Diese können beispielsweise den Pixeln oder auch schon den vollständig segmentierten Regionen in einem Bild entsprechen. Jeder dieser Stellen soll nun ein *Label* aus einer Labelmenge

$$\mathcal{L} = \{l_1, \dots, l_q\}$$

zugewiesen werden, wobei die einzelnen l_m einen Zustand oder eine Interpretation der Stelle s_i beschreiben. Der Wertebereich dieser Label kann numerisch, vektoriell oder auch symbolisch sein und sei hier als diskret und abzählbar angenommen. Entsprechen die Stellen beispielsweise den Regionen in einem Bild, so könnten die einzelnen Label eine Interpretation im Sinne von „Straße“, „Haus“ oder „Wasser“ vornehmen. Um nun jeder Stelle Werte aus \mathcal{L} zuweisen zu können, wird den Stellen s_i eine Zufallsvariable X_i zugeordnet. Die Menge aller Zufallsvariablen X_i definiert dann ein *Random Field*:

$$\mathcal{X} = \{X_1, \dots, X_n\}$$

Die Belegung aller X_i mit Werten $\omega_i \in \mathcal{L}$ wird als eine *Konfiguration* ω dieses Random Fields bezeichnet. Alle möglichen Konfigurationen werden in einer weiteren Menge zusammengefaßt:

$$\Omega = \{\omega = (\omega_1, \dots, \omega_n) \mid \omega_i \in \mathcal{L}, i = 1, \dots, n\}$$

Für ein Markov Random Field muß darüber hinaus ein *Nachbarschaftssystem* \mathcal{N} definiert sein, das die gegenseitigen Kontextabhängigkeiten der einzelnen Stellen beschreibt. Dieses kann z.B. die Bildmatrix sein, in der die Pixel in einer 4er oder 8er Nachbarschaft

zueinander stehen können. Die Modellierung dieser Nachbarschaften erfolgt allgemein in einem ungerichteten Graphen $\mathbf{G} = \{\mathcal{S}, \mathbb{E}\}$, in dem die Stellen \mathcal{S} als *Knoten* und die Abhängigkeiten durch *Kanten* $\mathbb{E} = \{(s_i, s_j) \mid s_i, s_j \in \mathcal{S}\}$ eingetragen sind. Die Nachbarschaft N_i an der Stelle s_i ist dabei durch folgende Beziehungen definiert:

$$\begin{aligned} s_i &\notin N_i, \\ s_j &\in N_i, \quad \text{wenn } \exists (s_i, s_j) \in \mathbb{E} \end{aligned}$$

Eng verbunden mit dem Begriff des Nachbarschaftssystem ist in der Graphentheorie das Konzept von *Cliquen*. Eine Clique \mathbf{c} beschreibt eine Menge von Knoten, in der jeder Knoten mit allen anderen Knoten in \mathbf{c} paarweise benachbart ist. Im folgenden wird die Menge aller Cliques für einen Graphen G mit \mathcal{C} notiert:

$$\begin{aligned} \mathcal{C}_1 &:= \{\{s_i\} \mid s_i \in \mathcal{S}\} \\ \mathcal{C}_2 &:= \{\{s_i, s_j\} \mid s_i, s_j \in \mathcal{S} \wedge (s_i, s_j) \in \mathbb{E}\} \\ \mathcal{C}_3 &:= \{\{s_i, s_j, s_k\} \mid s_i, s_j, s_k \in \mathcal{S} \wedge (s_i, s_j), (s_i, s_k), (s_j, s_k) \in \mathbb{E}\} \\ &\vdots \\ \mathcal{C} &:= \{\mathcal{C}_1 \cup \mathcal{C}_2 \cup \dots\} \end{aligned}$$

Mit den obigen Definitionen läßt sich nun ein *Markov Random Field* folgendermaßen charakterisieren: Ein Random Field \mathcal{X} mit einem Nachbarschaftssystem \mathcal{N} ist ein Markov Random Field, falls folgende Bedingungen erfüllt sind:

$$P(\mathcal{X} = \boldsymbol{\omega}) > 0, \quad \forall \boldsymbol{\omega} \in \Omega \quad (6.1)$$

$$P(X_i = \omega_i \mid X_j = \omega_j, s_j \in N_i) = P(X_i = \omega_i \mid X_j = \omega_j, \forall s_j \in \mathcal{S} \wedge s_j \neq s_i) \quad (6.2)$$

Die Bedingung (6.1) setzt die Positivität der Wahrscheinlichkeit für eine Konfiguration voraus und ist eher von technischer Natur. Entscheidend für ein MRF ist die Bedingung (6.2). Sie beschreibt die Markov Eigenschaft des Random Fields und dessen *lokale Charakteristik*, daß die Wahrscheinlichkeit für den Zustand eines Knoten s_i *nur* von der Wahrscheinlichkeit der Zustände der lokalen Nachbarn $s_j \in N_i$ abhängt.

In [Bes74] wird gezeigt, daß diese lokale Charakteristik eines MRFs auch in eindeutiger Weise deren Verbundwahrscheinlichkeit $P(\mathcal{X} = \boldsymbol{\omega})$ bestimmt. Jedoch ist genau diese nicht einfach aus den bedingten Wahrscheinlichkeiten zu errechnen. Und selbst für die bedingten Wahrscheinlichkeiten ist es schwierig entsprechende Funktionen für die lokale Charakterisierung zu finden, die dann auch tatsächlich bedingte Wahrscheinlichkeiten berechnen. Diese Schwierigkeiten können vermieden werden, wenn die Charakterisierung von Markov Random Fields durch eine Energiefunktion in Verbindung mit einer *Gibbs-Verteilung* erfolgt. So wird in [Bes74] ebenfalls bewiesen, daß die Konfigurationen von *Gibbs Random Fields* durch Gibbs-Verteilungen eindeutig bestimmt sind. Zudem beschreibt ein Theorem von Hammersley und Clifford genau die Äquivalenz von MRFs

und Gibbs-Verteilungen (siehe auch [Li95]). Demnach kann die Verbundwahrscheinlichkeit eines MRFs durch eine Gibbs-Verteilung folgendermaßen ausgedrückt werden:

$$P(\mathcal{X} = \boldsymbol{\omega}) = \frac{1}{Z} e^{-\frac{1}{T}U(\boldsymbol{\omega})} \quad (6.3)$$

Hier beschreibt Z die *Partition Function*, die im Wesentlichen eine Normierungskonstante darstellt:

$$Z := \sum_{\forall \boldsymbol{\omega} \in \Omega} e^{-\frac{1}{T}U(\boldsymbol{\omega})} \quad (6.4)$$

Da die Berechnung von Z sehr komplex ist und im folgenden nur die maximale Wahrscheinlichkeit einer Konfiguration von Interesse ist, kann ihre genaue Bestimmung vernachlässigt werden und ihr Wert als konstant angenommen werden. Der Parameter T beschreibt die Temperatur des Systems und wird hier zu $T = 1$ angenommen. Er motiviert sich ursprünglich aus physikalischen Prozessen bezüglich des Einschwingverhaltens von Systemen, und beeinflusst hier die Glattheit der Energiefunktion. Die Energiefunktion $U(\boldsymbol{\omega})$ berechnet wiederum die Energie für eine aktuelle Konfiguration des Random Fields aus den *Clique-Potentialen* $V_c(\boldsymbol{\omega})$ der Cliques, die sich aus \mathcal{N} ergeben:

$$U(\boldsymbol{\omega}) := \sum_{c \in \mathcal{C}} V_c(\boldsymbol{\omega}) \quad (6.5)$$

Die Clique-Potentiale modellieren dabei das *a priori* Wissen über die Abhängigkeiten zwischen den Zuständen einzelner Knoten innerhalb einer Clique. Sie können durch beliebige Funktionen kodiert werden, die nur von den Zuständen der Elemente einer Clique abhängen. In Summe kodieren sie das gesamte a priori Wissen über die Problemstellung, die in einem MRF abgebildet werden soll. Mit der Äquivalenz zu Gibbs-Verteilungen ergibt sich damit eine einfache Kodierung von Wissen über Kontextabhängigkeiten in MRFs durch Clique-Potentiale.

Das Ziel einer Bildanalyse ist es, neben der Kontextabhängigkeit zwischen den Entitäten eines Systems auch eine Abhängigkeit von den aktuellen Bilddaten zu modellieren. Eine Konfiguration soll also nicht die a priori Wahrscheinlichkeit maximieren, sondern die a posteriori Wahrscheinlichkeit unter gegebenen Bilddaten. Eine solche Konfiguration entspricht dann genau der Interpretation, die die Knoten s_i im MRF im Kontext *und* mit den zugrundeliegenden Daten d_i am besten beschreibt. Beschreibe \mathbf{d} den Vektor der Daten d_i für alle Knoten s_i , so läßt sich unter Anwendung der Bayes-Regel folgende Beziehung aufstellen:

$$\begin{aligned} P(\boldsymbol{\omega}|\mathbf{d}) &= \frac{P(\mathbf{d}|\boldsymbol{\omega})P(\boldsymbol{\omega})}{P(\mathbf{d})} \\ &= \frac{1}{P(\mathbf{d})} \frac{e^{-\frac{1}{T}U(\boldsymbol{\omega})}}{Z} P(\mathbf{d}|\boldsymbol{\omega}) \\ &= \frac{1}{P(\mathbf{d})} \frac{e^{-\frac{1}{T}(U(\boldsymbol{\omega}) - T \ln P(\mathbf{d}|\boldsymbol{\omega}))}}{Z} \end{aligned} \quad (6.6)$$

Auch diese Gleichung beschreibt wieder eine Gibbs-Verteilung, wenn sich der Term $\ln P(\mathbf{d}|\boldsymbol{\omega})$ in ähnlicher Weise wie die a priori Energie $U(\boldsymbol{\omega})$ durch eine Summe über alle Cliques beschreiben läßt. Wird die statistische Unabhängigkeit der Daten d_i vorausgesetzt, so folgt daraus, daß dann

$$\ln P(\mathbf{d}|\boldsymbol{\omega}) = \sum_{i=1,\dots,n} \ln P(d_i|\boldsymbol{\omega})$$

gilt. Somit kann die Datenabhängigkeit als zusätzlicher Summand in der Potentialberechnung der einelementigen Cliques berücksichtigt werden. Damit wird ein *a posteriori* Markov Random Field beschrieben, das durch die a posteriori Energiefunktion einer entsprechenden Gibbs-Verteilung modelliert ist:

$$U(\boldsymbol{\omega}|\mathbf{d}) := U(\boldsymbol{\omega}) - T \sum_{i=1,\dots,n} \ln P(d_i|\boldsymbol{\omega}) \quad (6.7)$$

Mit diesen Zusammenhängen sind nun die Voraussetzungen geschaffen, einen optimalen Zustand eines gegebenen Gesamtsystems durch eine *maximum a posteriori* (MAP) Formulierung der Wahrscheinlichkeiten zu erreichen. Bei gegebenen Bilddaten \mathbf{d} muß hierzu nun die Konfiguration $\boldsymbol{\omega}$ gewählt werden, die die minimale a posteriori Energie $U(\boldsymbol{\omega}|\mathbf{d})$ und damit die maximale a posteriori Wahrscheinlichkeit liefert.

6.1.2. Energieminimierung

Die Energieminimierung eines a posteriori MRF stellt ein Suchproblem nach einer optimalen Zustandkonfiguration des Feldes dar. Dabei ist aufgrund der Größe des Konfigurationsraumes Ω eine vollständige Suche nach der Konfiguration mit minimaler Energie häufig nicht möglich. Ein Überblick und Diskussion verschiedener Verfahren findet sich beispielsweise in [Li95].

Von der Idee her bauen viele Verfahren auf einem *stochastischen Relaxationsprozeß* auf und liefern eine Approximation des globalen Energieminimums. Ursprünglich motiviert sich das Verfahren aus physikalischen Prozessen, die zur Stabilisierung des Atomgitters bestimmter Stoffe eingesetzt werden. Hierzu wird zunächst eine Erhitzung der Stoffe durchgeführt, wodurch die Atome zu schwingen beginnen. Anschließend wird die Temperatur wieder langsam reduziert. Als Resultat wird das Einschwingen der Atome in energetisch stabilere Zustände erwartet. Bei der stochastischen Relaxation wird ähnlich vorgegangen. Zunächst wird eine beliebige Startkonfiguration gewählt und die a posteriori Energie bestimmt. Anschließend wird eine leicht veränderte Konfiguration eingesetzt, indem das Label an *einer* Stelle geändert und erneut eine Energie berechnet wird. Ist diese Energie geringer, so wird die entsprechende Konfiguration übernommen. Erhöht sich hingegen die Energie, so wird die Konfiguration nur mit einer gewissen Wahrscheinlichkeit in Abhängigkeit von der Energiedifferenz übernommen. Diese Schritte werden iteriert, bis eine gute Approximation des globalen Energieminimums erreicht ist.

Zur Einschränkung des Minimierungsprozesses wird beispielsweise beim Verfahren des *Simulated Annealing* in jedem Iterationsschritt die Temperatur in einer bestimmten Schrittweite verringert, bis sie gegen 0 konvergiert. Damit ist das Minimierungsergebnis jedoch entscheidend von der gewählten Schrittweite der Temperaturreduzierung abhängig. Bei genügend kleiner Schrittweite kann so eine perfekte Approximation des globalen Energieminimums erreicht werden. Hingegen führt eine zu schnelle Abkühlung zu einer Konfiguration, die ein lokales Minimum der Energie eines MRF darstellt. Der Gibbs-Sampler geht in analoger Weise vor, jedoch wird bei einer Änderung der Konfiguration gezielt eine Stelle ausgewählt. Diese Auswahl wird durch die lokale Charakteristik des MRFs bestimmt (siehe hierzu auch [Gem84]). Beide Verfahren können zu einem globalen Energieminimum führen, jedoch sind sie nicht deterministisch. Der Aufwand, der mit diesen Vorgehensweisen verbunden ist, ist häufig inakzeptabel hoch.

Highest Confidence First

In dieser Arbeit wird das Verfahren *Highest Confidence First* (HCF) nach [Cho90, Cho93] eingesetzt. Diese Methode versucht während einer Minimierung gezielt Knoten für eine Konfigurationsänderung zu wählen. Hierzu signalisieren die Knoten, in wie weit eine Änderung zu einer Reduzierung einer *lokalen Energie* führt und damit stückweise auch die Gesamtenergie des MRF reduziert wird.

Im HCF wird zunächst ein Label für den Zustand *unentschieden* ($l_{uncommitted}$) eines Knotens eingeführt. Dieses Label ist nicht Element der Labelmenge $l_{uncommitted} \notin \mathcal{L}$ und markiert für einen Knoten, daß noch keine Entscheidung für ein Label aus \mathcal{L} getroffen wurde. Zu Beginn eines Minimierungsprozesses wird das gesamte Random Field \mathcal{X} mit diesem Label initialisiert. Verläßt ein Knoten während einer Minimierung diesen unentschiedenen Zustand, weil ihm ein Label aus der Menge \mathcal{L} zugewiesen wird, kann er nicht wieder in den unentschiedenen Zustand zurückkehren. Er kann nur noch andere Zustände aus \mathcal{L} annehmen.

Die grundlegende Strategie des HCF ist es, im Minimierungsprozeß iterativ immer den Knoten auszuwählen, der das höchste Vertrauen genießt, eine maximale Reduzierung der globalen a posteriori Energie zu gewährleisten. Hierzu wird ein *Stabilitätsmaß* definiert, das unter Berücksichtigung der a posteriori Energie dieses Vertrauen für jeden Knoten quantifiziert. Zur Berechnung der Stabilität wird zunächst noch der Begriff der *lokalen Energie* $U_i(\omega_i|\mathbf{d}, \boldsymbol{\omega})$ eingeführt. Sie beschreibt den anteiligen Energiewert, den der Knoten s_i bei einer festen Konfiguration $\boldsymbol{\omega}$ durch das Label ω_i zur globalen Energie beiträgt.

$$U_i(\omega_i|\mathbf{d}, \boldsymbol{\omega}) := \sum_{\mathbf{c} \in \mathcal{C}: s_i \in \mathbf{c}} V'_{\mathbf{c}}(\boldsymbol{\omega}) - T \ln P(d_i|\omega_i) \quad (6.8)$$

Dabei definiert $V'(\mathbf{c})$ ein Clique-Potential, das im HCF zusätzlich den unentschiedenen

Zustand $l_{uncommitted}$ berücksichtigt:

$$V_{\mathbf{c}}'(\boldsymbol{\omega}) := \begin{cases} V_{\mathbf{c}}(\boldsymbol{\omega}) & , \text{ wenn } \forall s_j \in \mathbf{c} : \omega_j \in \mathcal{L} \\ 0 & , \text{ sonst} \end{cases} \quad (6.9)$$

Ein Potential beeinflußt demnach nur die lokale Energie, wenn bereits alle Knoten in den jeweiligen Cliques den unentschiedenen Zustand verlassen haben. Resultierend daraus reduziert sich dann die initiale lokale Energie für einen Knoten zu der a posteriori Energie für das Label ω_i unter den Daten \mathbf{d} .

Die Bestimmung der lokalen *Stabilität* $S_i(\boldsymbol{\omega})$ für einen Knoten s_i erfolgt nun durch eine Untersuchung der lokalen Energie für verschiedene Belegungen von X_i mit $\hat{\omega} \in \mathcal{L}$ (ω_i bezeichnet die *aktuelle* Belegung des Knoten s_i in der Konfiguration $\boldsymbol{\omega}$).

$$S_i(\boldsymbol{\omega}) := \begin{cases} \min_{\hat{\omega} \in \mathcal{L}, \hat{\omega} \neq \omega_i} U_i(\hat{\omega} | \mathbf{d}, \boldsymbol{\omega}) - U_i(\omega_i | \mathbf{d}, \boldsymbol{\omega}) & , \text{ wenn } \omega_i \in \mathcal{L} \\ \min_{\omega' \in \mathcal{L}, \omega' \neq \omega_i} U_i(\hat{\omega} | \mathbf{d}, \boldsymbol{\omega}) - U_i(\omega' | \mathbf{d}, \boldsymbol{\omega}) & , \text{ wenn } \omega_i = l_{uncommitted} \text{ und} \\ & \hat{\omega} = \operatorname{argmin}_{\tilde{\omega} \in \mathcal{L}} U_i(\tilde{\omega} | \mathbf{d}, \boldsymbol{\omega}) \end{cases} \quad (6.10)$$

Die Stabilität $S_i(\boldsymbol{\omega})$ beschreibt also den maximal erreichbaren Energieverlust, wenn der aktuelle Zustand ω_i eines Knoten unter Berücksichtigung der aktuellen Konfiguration $\boldsymbol{\omega}$ in den Zustand $\hat{\omega}$ geändert wird. Befindet sich ein Knoten im unentschiedenen Zustand, so nimmt seine lokale Stabilität dann große negative Werte an, wenn die lokale Energie für einen Zustand $\hat{\omega}$ sehr gering gegenüber der lokalen Energie für alle anderen Zuständen $\omega \in \mathcal{L}$ ist. Dies bedeutet, daß Knoten im unentschiedenen Zustand, die eine starke lokale Evidenz für ein Label besitzen, mit ihrer Zustandsänderung eine starke Reduzierung der globalen Energie in Aussicht stellen.

Das gesamte Verfahren des HCF stellt sich mit diesen Definitionen wie folgt dar: Nach der initialen Belegung des Feldes \mathcal{X} mit dem Label $l_{uncommitted}$, wird für jeden Knoten seine lokale Stabilität bestimmt. Anschließend wird iterativ immer der Knoten für eine Zustandsänderung gewählt, der die maximale negative Stabilität hat. Diesem Knoten wird dann der Zustand zugewiesen, der den maximalen Energieverlust realisiert. Nach jeder Zustandsänderung an einem Knoten wird für alle benachbarten Knoten eine neue lokale Stabilität ermittelt. Dieser Vorgang iteriert, bis alle Knoten eine positive Stabilität aufweisen. Das MRF befindet sich dann in einem lokalen Energieminimum, da allen Knoten ein Zustand aus \mathcal{L} zugewiesen wurde und an keiner Stelle mehr durch eine Zustandsänderung eine Reduzierung der lokalen Energie zu erzielen ist.

Der HCF-Algorithmus ist deterministisch, unabhängig von einer geeigneten Startkonfiguration und liefert eine Lösung, die eine sehr gute Annäherung an eine MAP-Lösung darstellt. In [Cho90] wird das HCF-Verfahren mit anderen Minimierungsansätzen auf konkreten Beispieldaten verglichen. Danach liefert das Verfahren qualitativ bessere Ergebnisse als andere Verfahren und resultiert zudem in niedrigeren Energieniveaus, als die gegenübergestellten Verfahren. Dies unterstreicht die Wahl dieses Verfahrens zur Minimierung von Markov Random Fields.

6.1.3. Einsatzgebiete in der Bildanalyse

Der Formalismus von Markov Random Fields ist in der Bildanalyse schon mehrfach mit Erfolg eingesetzt worden (für einen Überblick siehe [Che93a, Li95]). Dabei ist das hauptsächliche Anwendungsgebiet der sogenannte *low-level*-Bereich der Bildverarbeitung. Sie werden unter anderem zur Bildrestaurierung, Kantendetektion (in restaurierten Bildern), Texturanalyse und für die Berechnung des optischen Flusses eingesetzt. Zunehmend wird das Anwendungsfeld auch auf Ansätze ausgedehnt, deren Ziel eher eine Interpretation des Bildinhaltes wie etwa eine Objekterkennung ist. Im folgenden wird ein kurzer Überblick über die verschiedenen prinzipiellen Einsatzgebiete gegeben.

Im Bereich der signalnahen Anwendungen ist der Einsatz von MRFs sehr verbreitet, da durch die Bildmatrix eine feste Struktur der Nachbarschaft implizit vorgegeben wird. Je nach Anwendung sind die Stellen des MRFs dabei auf den Bildpunkten, oder aber zwischen den Pixeln lokalisiert. Die Nachbarschaft wird dann überwiegend über eine 8-er oder 4-er Nachbarschaft auf der Bildmatrix definiert.

Auf einem so definierten MRF werden zum Beispiel Ansätze zur Rauschunterdrückung und Bildrekonstruktion realisiert [Gem84, Boy98, Lu00]. Dabei wird das Ziel verfolgt die ideale Bildfunktion f anhand der vorhandenen Beobachtungen d unter Annahme bestimmter Kontextabhängigkeiten zu rekonstruieren. Für die Bildrekonstruktion wird beispielsweise ein konstanter oder kontinuierlicher Intensitätsverlauf in einer lokalen Nachbarschaft erwartet. Dieses a priori Wissen wird in entsprechenden Potentialfunktionen für mehrelementige Cliques abgebildet. Die Abhängigkeit von den Daten wird in der Regel durch Funktionen realisiert, die auf der ersten oder zweiten Ableitung der Bildfunktion basieren. In Zusammenhang mit einer Rauschunterdrückung kann zusätzlich eine Kantendetektion realisiert werden [Gem84, Gei91, Lei97]. Weitere Verfahren setzen sich ausschließlich mit der Kantendetektion und deren Verbesserung auseinander [Mar89, Mil92, Ura95, Nad96].

Ein weiteres Anwendungsgebiet für MRFs im Bereich der Bildsegmentierung ist die Detektion von Bereichen mit einer vorgegebenen Homogenitätscharakteristik. Hierfür sind viele Ansätze vorgeschlagen worden, die eine Regionen- oder Textursegmentierung für verschiedenste Anwendungsgebiete realisieren [Kim94, Kri95, Pag00]. Auch für die Analyse von Bildsequenzen, beispielsweise zur Berechnung des optischen Flusses [Hei93, Lut94, Roy00] oder zur robusten Verfolgung von Bildmerkmalen [Kam00], wird der MRF-Formalismus eingesetzt.

Das grundlegende Prinzip für alle diese Verfahren ist, daß sie die einschränkenden Bedingungen ihrer jeweiligen Anwendung, deren Wirkung überwiegend räumlich begrenzt ist, in den Potentialfunktionen des MRFs abbilden. Gelingt diese Modellierung in einer angemessenen Form, so entsteht aus der lokalen Interaktion bei der Energieminimierung eine global konsistente Interpretation. Dabei kommen zum Teil auch mehrere gekoppelte MRFs zum Einsatz, die unterschiedliche Modalitäten interpretieren und erst in ihrer gemeinsamen Betrachtung eine konsistente Interpretation im Sinne der Anwendung realisieren (siehe beispielsweise [Gem84, Li95]).

Im Gegensatz zu den signalnahen Anwendungen sind nur verhältnismäßig wenige Ansätze vorgeschlagen worden, die den MRF-Formalismus auf einer höheren Interpretationsstufe einsetzen. So werden beispielsweise in [Mod93, Li94, Fri97, Boy99b] Anwendungen zur Objekterkennung vorgestellt. Hierzu werden in [Mod93] Regionen als Stellen im MRF aufgefaßt und aneinandergrenzende Regionen stehen durch eine Kante in Relation zueinander. Als Labelmenge werden Objektidentifizierungen („Himmel“, „Straße“, „Wasser“) verwendet, die den Regionen im Minimierungsprozeß zugewiesen werden. Einschränkungen für die Labelzuweisung ergeben sich aus den Nachbarschaftsbeziehungen und Merkmalen der Regionen, und sind in entsprechenden Potentialfunktionen kodiert. Auf Basis dieses Ansatzes wird in [Kim93] zusätzlich ein Verfahren präsentiert, das eine Verbindung von Neuronalen Netzen und MRFs zum Lernen von Clique-Potentialen vorschlägt. Neben diesem Verfahren existieren noch diverse andere Arbeiten, die sich mit der Schätzung von Parametern für MRFs in den jeweiligen Anwendungsgebieten befassen [SCZ98, Des99, Bor99].

Die Objekterkennungssysteme nach [Li94, Fri97, Boy99b, Sch00b] beruhen prinzipiell auf der Idee jedes Objekt durch *ein* MRF zu repräsentieren. Die Objekte werden in mehrere Bestandteile oder Eigenschaftskomponenten (räumlich oder symbolischer Art) zerlegt, die dann als Stellen für das MRF verwendet werden. Die Nachbarschaft wird so gewählt, daß bekannte Kontextabhängigkeiten zwischen den einzelnen Stellen mit entsprechenden Potentialfunktionen modelliert werden können. Als Labelmenge werden in der Regel die detektierten Bildprimitiva oder Bildmerkmale verwendet. Die Objekterkennung wird dann durch eine Abbildung der Energiezustände einzelner Konfigurationen auf entsprechende Objektmodelle realisiert.

In Abschnitt 3.2.1 wurden bereits einige Ansätze erwähnt, die Markov Random Fields zur Realisierung perzeptiven Gruppierens einsetzen. Im folgenden Abschnitt erfolgt nun die Beschreibung des in der vorliegenden Arbeit entwickelten Verfahrens den MRF-Formalismus zur Bewertung konturbasierter perzeptiver Gruppen zu verwenden. Die Motivation den MRF-Formalismus für diese Aufgabe einzusetzen begründet sich durch die Eigenschaft von MRFs, ungerichtete Wechselwirkungen zwischen den Stellen zu modellieren. Dies ist insbesondere für die Auflösung von widersprüchlichen Strukturierungen durch Gruppierungshypothesen von entscheidender Bedeutung.

6.2. MRF zur Bewertung von Gruppierungshypothesen

Die im Kapitel 4 und in Abschnitt 5.3 vorgestellte Generierung von konturbasierten Gruppierungshypothesen ist ein rein datengetriebener Prozeß. Dieser Prozeß trifft unter Berücksichtigung lokaler Eigenschaften eine Entscheidung über das Gruppierungsverhalten von Konturbeschreibungen, ohne diese Entscheidung in einen globalen Kontext mit anderen Entscheidungen zu setzen. Eine Bewertung jeder dieser Gruppierungshypothesen ist daher bisher nur unter ihrem lokalen Charakter möglich: wie gut erfüllt eine Gruppierungshypothese die einschränkenden Bedingungen ihres zugrundeliegenden

Gestaltgesetzes. Zielsetzung ist jedoch eine Bewertung aller Gruppierungshypothesen, so daß ein Hypothesensatz gefunden wird, der eine global konsistente Interpretation der Daten darstellt. Diese *globale* Bewertung erfordert zusätzliches Wissen über Relationen und Kontextabhängigkeiten zwischen den verschiedenen Gruppierungstypen.

Zur Formulierung des Bewertungsproblems durch ein MRF wird im folgenden der Gruppierungsgraph zur Modellierung eines Nachbarschaftssystems zwischen den Gruppierungshypothesen definiert [Sch95, Maß96, Maß97, Ack97, Pos98]. Im Vorfeld dieser Definition wird die hier eingesetzte Labelmenge vorgestellt und die damit verfolgte Interpretation für eine Hypothese beschrieben. Anschließend erfolgt die Formulierung des a priori Wissens über die Kontextabhängigkeiten zwischen den Gruppierungshypothesen durch entsprechende Clique-Potentiale. Danach wird die Modellierung der Datenabhängigkeit dargestellt, bevor dann die Energieminimierung mit dem HCF-Algorithmus beschrieben wird.

In der folgenden Darstellung des Verfahrens zur Bewertung von Gruppierungshypothesen werden die kol- und kurvilinearen Gruppierungen von geschlossenen Figuren aus Abschnitt 4.4 nicht berücksichtigt. Dies begründet sich aus der Tatsache heraus, daß die Implementation aus zeitlichen Gründen im Rahmen dieser Arbeit nicht mehr realisiert werden konnte. Generell ist das Konzept jedoch für die Integration einer zusätzlichen Hierarchieebene und auch für die Berücksichtigung eines rekursiven, zunehmend abstrakteren Gruppierungsprozesses offen.

6.2.1. Gruppierungsgraph und Labelmenge

Mit den Markov Random Fields ist in Abschnitt 6.1 eine Methode vorgestellt worden, die bezüglich eines Nachbarschaftssystems die Kodierung von Wissen über Kontextabhängigkeiten zwischen den Gruppierungshypothesen unterstützt. Wird die Gruppierungshierarchie aus Abbildung 4.1 auf Seite 52 betrachtet, so motiviert sich sofort die Abbildung der Gruppierungshypothesen auf ein Markov Random Field: Die initialen Konturmodelle mit den daraus entstandenen Gruppierungen g_i stellen die Menge der Knoten $\mathcal{S} = \{g_1, \dots, g_n\}$ eines Gruppierungs-MRFs (G-MRF) dar. Im weiteren werden nun auch die initialen Konturmodelle als primitive Gruppierungen unter dem Begriff der Gruppierungshypothesen subsumiert, da die Kontursegmentierung in gewisser Weise auch ein Prozeß einer Hypothesenbildung ist. Mit dieser Modellierung hebt sich der Ansatz gegenüber vielen Anwendungen von MRFs in der Bildanalyse ab. Diese Knoten interpretieren hier keine disjunkten Bildbereiche, sondern beschreiben vielmehr den jeweiligen Strukturierungsgrad der teilweise *gleichen* Bilddaten auf unterschiedlichem Niveau. So führen zwei konstituierende Geradenstücke g_2, g_3 einer kollinearen Gruppierung $g_2 - g_3$, die wiederum Teil einer Parallelität $g_1 \parallel (g_2 - g_3)$ ist, zu vier Knoten im G-MRF, die alle die gleichen Bilddaten g_2, g_3 in unterschiedlicher Weise interpretieren (siehe Abbildung 6.1).

Labelmenge

Aus der Zielsetzung der Bewertung von Gruppierungshypothesen motiviert sich hier auch eine andere Vorstellung über die Interpretation oder den Zustand eines Knotens. Die Konturmodelle und deren Gruppierungen, die beispielsweise Objektstrukturen beschreiben, sind für eine Bildanalyse signifikanter, als Gruppierungshypothesen die auf Hintergrundstrukturen basieren. So wird der Labelmenge hier aber keine symbolische Interpretation wie {Objektstruktur, Hintergrundstruktur} oder {Haus, Rasen, ...} zugeschrieben, sondern das zugewiesene Label trifft eine Aussage über die *Signifikanz* der entsprechenden Gruppierungshypothese für eine Strukturierung des Bildinhaltes. Der Signifikanzwert ist dabei ein Maß für die lokale Evidenz, die eine Hypothese durch ihre Daten erfährt, und markiert so die Sicherheit der entsprechend vorgenommenen Interpretation (wie gut erfüllen die Daten einer Hypothese das ihr zugrundeliegende Gestaltgesetz). Darüber hinaus spiegelt der Signifikanzwert den Grad einer konsistenten Interpretation im Kontext mit anderen Hypothesen wider. Die Signifikanz kann Werte aus dem diskretisierten Intervall $[0, 1]$ annehmen, wobei 1 eine hohe Signifikanz und 0 keine Signifikanz beschreibt:

$$\mathcal{L} := \{0, \Delta_{sig}, 2\Delta_{sig}, \dots, 1\} \quad , \text{ mit } \Delta_{sig} = 0.05 \quad (6.11)$$

Nachbarschaftssystem

Die Definition eines Nachbarschaftssystems für ein Gruppierungs-MRF motiviert sich in naheliegender Weise ebenfalls aus der Gruppierungshierarchie in Abbildung 4.1. Mit den Gruppierungshypothesen und der Gruppierungshierarchie läßt sich ein sogenannter *Gruppierungsgraph* definieren, in dem die Gruppierungshypothesen durch die Knoten und die Relationen zwischen den Gruppierungen durch die Kanten repräsentiert sind. Diese Kanten beschreiben explizit *keine* räumliche Nachbarschaft zwischen den Hypothesen, sondern definieren zwei wechselseitige Kontextbezüge, die für eine global konsistente Interpretation der Daten erforderlich sind und sich aus der Gruppierungshierarchie ableiten lassen.

Unterstützende Relationen bestehen zwischen Gruppierungshypothesen, die in einer Bestandteilsbeziehung zueinander stehen. Eine Hypothese g_i ist Bestandteil einer anderen Hypothese g_k , wenn erstere zur Gruppenbildung der zweiten Hypothese eingesetzt wird: $g_i \in GElem(g_k)$. Die unterstützende Wirkung ist dabei *nicht* richtungsabhängig. Eine lineare Gruppe $g_2 - g_3$ unterstützt die Hypothese einer Parallelität $g_1 \parallel (g_2 - g_3)$, und $g_2 - g_3$ wiederum erfährt eine Unterstützung, weil sie zur Bildung von $g_1 \parallel (g_2 - g_3)$ eingesetzt wird und so eine korrekte Interpretation zu sein scheint (vergleiche Abbildung 6.1). Unterstützung wirkt also allgemein von einer Gruppierung „höheren Strukturierungsgrades“ auf ihre Bestandteile und umgekehrt.

Neben den unterstützenden Relationen werden *konkurrierende* Relationen zwischen den Gruppierungshypothesen definiert. Die konkurrierenden Relationen können *nur* zwischen Gruppierungshypothesen entstehen, denen das gleiche Gruppierungsprinzip zu-

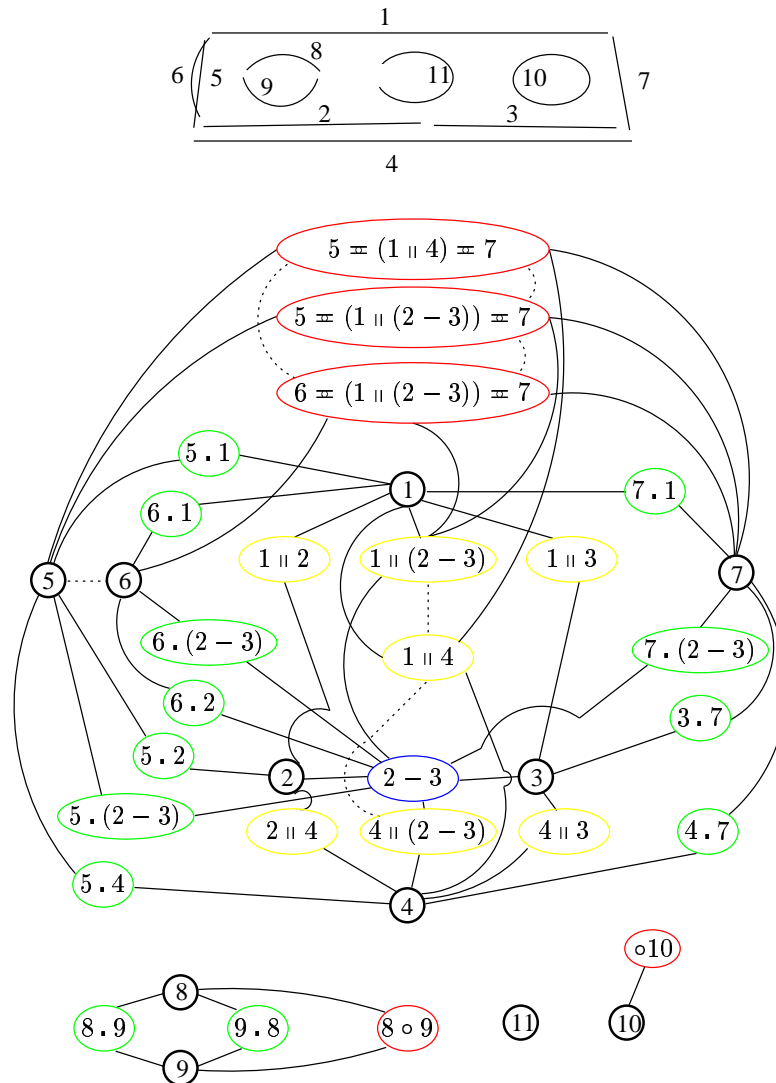


Abbildung 6.1.: Ein synthetisches Beispiel aus der *baufix*[®] Domäne (Drei-Loch-Leiste) mit seinem Gruppierungsgraphen. Initiale Konturmodelle sind in schwarzen, Proximitäten in grünen, Kollinearitäten in blauen, Parallelitäten in gelben und Geschlossenheiten in roten Kreisen dargestellt. Unterstützende Relationen sind mit durchgezogenen und konkurrierende Relationen durch gepunktete Linien markiert. Es ist deutlich sichtbar, daß ein Gruppierungsgraph in disjunkte Teilgraphen zerfallen kann, weil einzelne lokale Strukturen nicht zu globaleren Gruppierungen beitragen. Darüber hinaus illustriert dieses Beispiel, daß bereits bei relativ einfachen Strukturen ein komplexes Nachbarschaftssystem entstehen kann. Die Positionierung der Knoten ist im Bereich der entsprechenden Gruppierung erfolgt, soweit dies möglich ist.

grunde liegt und die eine sich widersprechende Interpretation des Bildinhaltes vornehmen (vergleiche Konturmodelle g_5 vs. g_6 in Abbildung 6.1). Aufgrund unterschiedlicher Gruppierungsprinzipien kann also auch eine Kollinearität nicht in Konkurrenz zu einer Parallelität stehen. Bereits bei der Generierung der Gruppierungshypothesen in den Abschnitten 4.2.2 und 4.3 sind offensichtlich konkurrierende (redundante) Gruppierungen eliminiert worden. An dieser Stelle soll nun jedoch keine weitere Redundanzauflösung in diesem Sinne erfolgen, sondern es soll vielmehr die konträre Strukturierung des Bildinhaltes in einem globaleren Kontext Berücksichtigung finden und zu einer global konsistenten Interpretation führen. Im folgenden werden die Bedingungen, die zu einer konkurrierenden Relation zwischen zwei Gruppierungshypothesen führen, in Abhängigkeit des zugrundeliegenden Gruppierungsprinzips vorgestellt:

Konturmodelle: Zwei Konturmodelle g_i, g_k werden als konkurrierend definiert, wenn beide Modelle teilweise die identischen Kantenelemente P im Bild approximieren. Als konstituierende Bedingung für eine Konkurrenz muß gelten, daß mindestens ein Prozentsatz $\Delta_{overlap}(\cdot)$ der Kantenelemente des kürzeren Modells auch von dem längeren Modell approximiert werden:

$$\frac{|P(g_i) \cap P(g_k)|}{\min(|P(g_i)|, |P(g_k)|)} > \Delta_{overlap}(\cdot) \quad (6.12)$$

Kollinearitäten/Kurvilinearitäten: Kollinearitäten und Kurvilinearitäten konkurrieren miteinander, wenn zwei Hypothesen g_i, g_k mindestens ein identisches Bestandteilelement enthalten:

$$GElem(g_i) \cap GElem(g_k) \neq \emptyset$$

Proximitäten: Nähe-Gruppierungen konkurrieren *nicht* miteinander. Diese Gruppierungen enthalten in zwei Hypothesen notwendigerweise identische lineare Gruppen, damit beispielsweise eine Geschlossenheit aufgrund der Proximitäten hypothetisiert werden kann. Sie stellen daher mehr die Interpretation einer Relation zwischen den linearen Gruppen, als eine echte Strukturierung des Bildinhaltes dar.

Parallelität: Die Konkurrenz von Parallelitätshypothesen motiviert sich durch das Beispiel in Abbildung 6.2(a). Zwischen den dargestellten linearen Gruppen können folgende Parallelitätshypothesen gebildet werden:

$$p_1: g_1 \parallel g_2, \quad p_2: g_1 \parallel g_3, \quad p_3: g_3 \parallel g_2$$

Dabei steht die Hypothese p_1 jeweils in einer Konkurrenzrelationen zu den anderen beiden Hypothesen p_2, p_3 , weil p_1 die folgende Bedingungen erfüllt: Allgemein wird eine Konkurrenz zwischen zwei Parallelitäten definiert, wenn beide Hypothesen p_1, p_2 eine identische lineare Gruppe g_1 gruppieren:

$$GElem(p_1) \cap GElem(p_2) \neq \emptyset$$

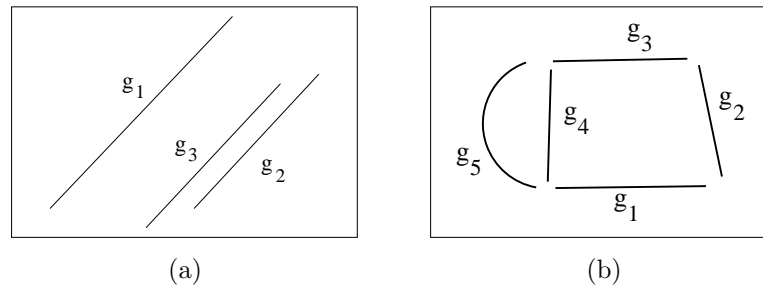


Abbildung 6.2.: Motivation zur Konkurrenz zwischen (a) parallelen Gruppierungshypothesen und (b) Geschlossenheitshypothesen (siehe Text).

Als weitere Bedingung muß eine Parallelitätshypothese p_1 den Bildbereich der anderen Hypothese p_2 überdecken. Eine solche Überdeckung entsteht, wenn die nicht identischen linearen Gruppen g_2, g_3 auf der gleichen Seite von g_1 lokalisiert sind. Letztere Bedingung verhindert deshalb auch eine Konkurrenzrelation zwischen p_2, p_3 , da die linearen Gruppen g_1 und g_2 auf unterschiedlichen Seiten von g_3 positioniert sind.

Geschlossenheit: Geschlossenheiten werden als konkurrierend definiert, wenn die Bestandteilelemente einen nicht leeren Schnitt bilden und sich die eingegrenzten Flächen von beiden Geschlossenheitshypothesen überlagern:

$$GElem(G_1) \cap GElem(G_2) \neq \emptyset \quad \wedge \quad A(G_1) \cap A(G_2) \neq \emptyset$$

In Abbildung 6.2(b) ist ein Beispiel gezeigt, in dem die Gruppierungshypothesen $G_1 = g_2 = (g_3 \parallel g_1) = g_4$ und $G_2 = g_2 = (g_3 \parallel g_1) = g_5$ konkurrieren.

Generierung des Gruppierungsgraphen

Mit der Definition von unterstützenden und konkurrierenden Relationen zwischen Hypothesen ist das Nachbarschaftssystem \mathcal{N} und die daraus entstehende Menge der Cliques \mathcal{C} für ein G-MRF vollständig spezifiziert. Die Generierung des Gruppierungsgraphen für das G-MRF erfolgt in einem Top-Down Verfahren entlang der Gruppierungshierarchie. Ausgehend von den Gruppierungen, die in keiner anderen Gruppierung als konstituierendes Element enthalten sind, werden sukzessive Knoten für die entsprechende Bestandteilshypothesen eingetragen, bis die Ebene der Konturmodelle für eine Gruppierung vollständig eingetragen ist. Während dieser Knotengenerierung wird sichergestellt, daß Knoten nicht doppelt eingetragen werden, und daß entsprechende unterstützende Kanten aufgrund einer Bestandteilsbeziehung erzeugt werden. Nach der Generierung eines Knoten werden gegebenenfalls konkurrierende Kanten zu anderen bereits vorhandenen Knoten des gleichen Gruppierungstyps eingetragen, wenn die oben spezifizierten

Bedingungen für eine Konkurrenz erfüllt sind. Diese Vorgehensweise wird wiederholt, bis alle Gruppierungshypothesen eingetragen sind. In Abbildung 6.1 ist für ein synthetisches Beispiel der Gruppierungsgraph dargestellt.

6.2.2. Kontextbezüge zwischen Gruppierungen

In der formalen Darstellung von MRFs in Abschnitt 6.1.1 wurde erläutert, daß das gesamte a priori Wissen über die Kontextabhängigkeiten zwischen den Zuständen der Knoten (hier zwischen den Signifikanzwerten der Gruppierungshypothesen) in den Clique-Potentialen $V_c(\omega)$ kodiert wird. Im folgenden werden die Clique-Potentiale zur Abbildung des a priori Wissens für das hier entwickelte G-MRF vorgestellt.

Die Clique-Potentiale $V_c(\omega_i)$ der *einelementigen* Cliquen beschreiben die a priori Wahrscheinlichkeit $P(\omega_i)$, die ein Signifikanzwert für eine Hypothese besitzt. Da dieses Clique-Potential weder von einer Nachbarschaftsinformation noch von der Bildinformation profitiert, wird hier eine Gleichverteilung dieser Wahrscheinlichkeit für das gesamte Feld angenommen. Mit dieser Annahme bleibt diese Wahrscheinlichkeit konstant und ist zudem für jeden Knoten identisch. Das Clique-Potential kann dementsprechend vernachlässigt werden.

In Abschnitt 6.2.1 ist in unterstützende und konkurrierende Kontextbezüge zwischen den Gruppierungshypothesen unterschieden worden. Diese Unterscheidung wird für die Definition der Clique-Potentiale für zweielementige Cliquen übernommen und deren Menge dementsprechend aufgeteilt:

$$\mathcal{C}_2 = \mathcal{C}_2^{supp} \cup \mathcal{C}_2^{comp}$$

Im folgenden wird das Potential für zweielementige Cliquen kurz mit unterstützendes V_c^{supp} bzw. konkurrierendes Potential V_c^{comp} bezeichnet. Ein unterstützendes Potential soll die Kompatibilität und Korrektheit der Interpretationen beurteilen, die von zwei Hypothesen g_i, g_k vorgenommen werden. Eine optimale Unterstützung besteht, wenn beide Hypothesen einen hohen Signifikanzwert besitzen. Der Potentialwert soll in diesem Fall die Energie absenken und wird negativ gewählt. Ist in einer Clique der Hypothese g_i eine niedrige Signifikanz und g_k eine hohe Signifikanz zugewiesen, so widerspricht dies einer unterstützenden Relation. Es entsteht eine inkonsistente Interpretation, in der eine nicht signifikante Hypothese und eine signifikante sich gegenseitig unterstützen sollen. Für solche Konstellation wird ein positives Potential definiert. Sind beide Signifikanzwerte niedrig, wird ein neutrales Potential gewählt, da keine Aussage über die Konsistenz und Korrektheit der Interpretation getroffen werden kann. Mit dieser Motivation entsteht eine Unterteilung des Definitionsbereichs von V_c^{supp} in vier Quadranten, die in Abbildung 6.3(a) qualitativ dargestellt sind. In drei dieser Quadranten wird die Potential-Funktion durch eine entsprechend parametrisierte Gauß-Funktion φ modelliert, während die Neutralität im vierten Quadranten durch den Wert 0 repräsentiert wird. Neben einer Normierung des Wertebereichs von φ auf ein Intervall $[0, 1]$, wird die

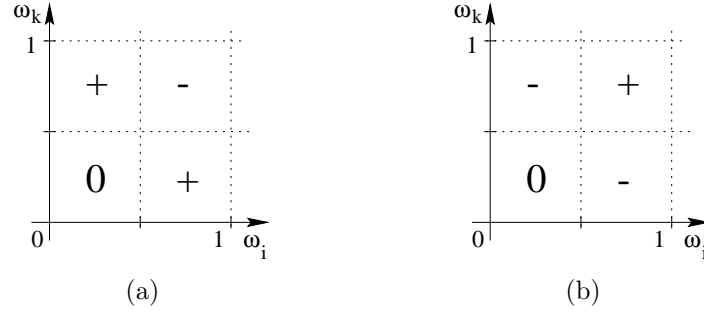


Abbildung 6.3.: Qualitative Modellierung des Clique-Potentials für zweielementige Cliques. Die Darstellung erfolgt in Abhängigkeit der Signifikanzwerte ω_i und ω_k für (a) unterstützende und (b) konkurrierende Relationen.

Standardabweichung zu $\sigma = 0.2$ gewählt, um nahezu glatte Übergänge zwischen den Quadranten zu gewährleisten:

$$\varphi(\omega_i, \omega_k) = \frac{1}{8\pi\sigma^2} e^{-\frac{1}{2\sigma^2} ((\omega_i - \mu_i)^2 + (\omega_k - \mu_k)^2)} \quad (6.13)$$

Die Erwartungswerte μ_i, μ_k werden in Abhängigkeit von den Signifikanzwerten der Hypothesen gesetzt und es ergibt sich folgende Funktion für den gesamten Definitionsbereich:

$$\varphi^{supp}(\omega_i, \omega_k) := \begin{cases} 0 & : 0 \leq \omega_i \leq 0.5 \quad \wedge \quad 0 \leq \omega_k \leq 0.5 \\ \varphi(\omega_i, \omega_k) & : 0 \leq \omega_i \leq 0.5 \quad \wedge \quad 0.5 < \omega_k \leq 1 \quad \wedge \\ & \mu_i = 0 \quad \wedge \quad \mu_k = 1 \\ -\varphi(\omega_i, \omega_k) & : 0.5 < \omega_i \leq 1 \quad \wedge \quad 0.5 < \omega_k \leq 1 \quad \wedge \\ & \mu_i = 1 \quad \wedge \quad \mu_k = 1 \\ \varphi(\omega_i, \omega_k) & : 0.5 < \omega_i \leq 1 \quad \wedge \quad 0 \leq \omega_k \leq 0.5 \quad \wedge \\ & \mu_i = 1 \quad \wedge \quad \mu_k = 0 \end{cases} \quad (6.14)$$

Der Funktionsgraph ist in Abbildung 6.4 für die einzelnen Quadranten dargestellt. Es ist deutlich ersichtlich, daß eine Entscheidung für Werte im mittleren Definitionsbereich zu neutralen Potentialwerten führt. Mit diesem neutralen Beitrag zur Gesamtenergie werden zunächst Auswertungen weiterer Kontextbezüge dieser Hypothesen abgewartet, bevor eine eindeutige Entscheidung getroffen wird.

Die Definition des unterstützenden Potentials durch die Funktion aus Gleichung (6.14) ist in bestimmten Situationen problematisch. In Abbildung 6.5 erhält die Gruppierung $g_1 \frown g_2 \frown g_3$ allein durch die höhere Anzahl ihrer konstituierenden Elemente eine stärkere Unterstützung als die Gruppierung $g_1 \frown g_4$. Dies entspricht nicht der gewünschten Modellierung der Kontextabhängigkeiten. Die Bewertung der Gruppierungen bezüglich ihrer Signifikanz soll in einem globaleren Kontext erfolgen und nicht von der Anzahl der

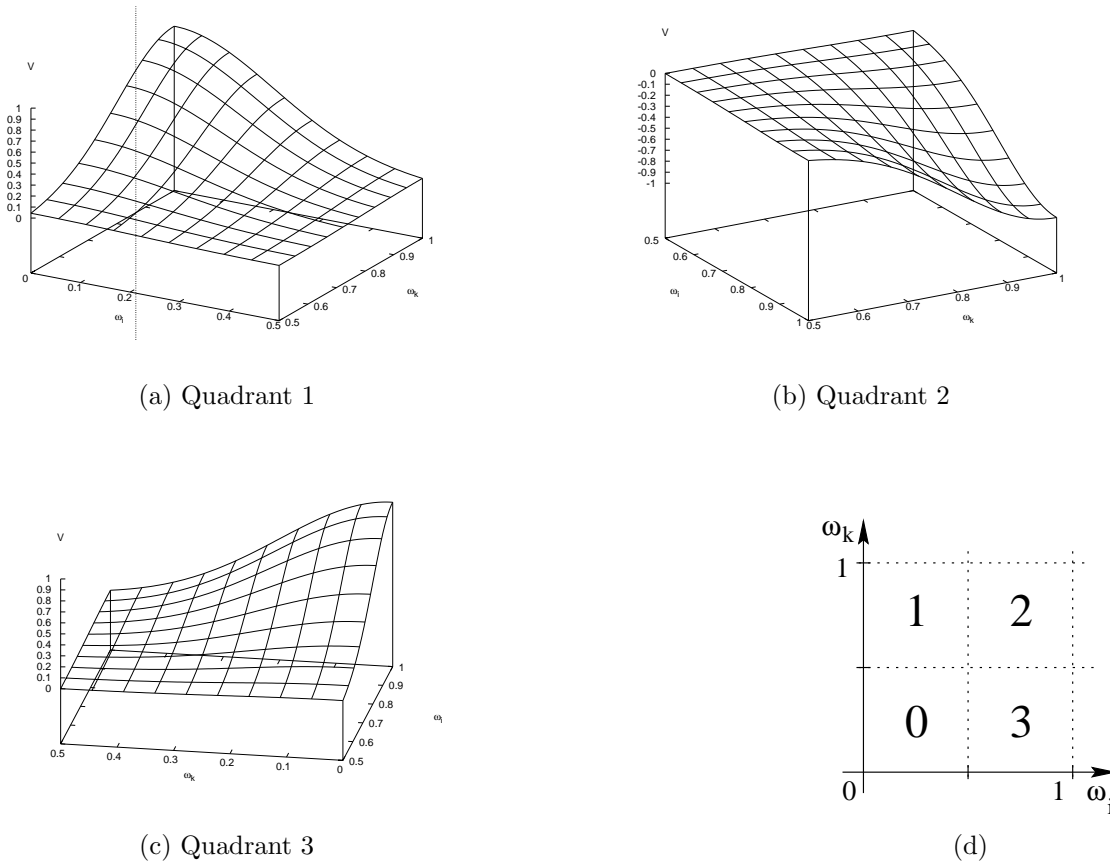


Abbildung 6.4.: Funktionsverlauf des unterstützenden Potential φ^{supp} in Abhängigkeit der Signifikanzwerte ω_i, ω_j . Der Verlauf ist in den jeweiligen Quadranten dargestellt und in (d) ist die Zuordnung der Quadranten für den gesamten Definitionsbereich durch Angabe der Quadrantennummer markiert.

konstituierenden Elemente einer Struktur dominiert werden. Aus diesem Grund wird φ^{supp} jeweils mit der Anzahl der konstituierenden Elemente einer Gruppierung normiert. Dies gilt für alle unterstützenden Relationen, die sich zwischen den Ebenen der Gruppierungshierarchie ergeben. Sei g_k konstituierendes Element der Gruppe g_i , so ergibt sich das Clique-Potential für die unterstützende Relation zwischen g_k und g_i zu:

$$V_{\mathbf{c}}^{supp}(\boldsymbol{\omega}) := \frac{\varphi^{supp}(\omega_i, \omega_k)}{|GElem(g_i)|}, \quad \text{mit } \mathbf{c} = \{g_i, g_k\} \in \mathcal{C}_2^{supp} \quad (6.15)$$

Das konkurrierende Potential für zweielementige Cliquen ist in Analogie zum unterstützenden Potential definiert. Allerdings soll mit diesem Potential die Inkompatibilität von zwei Hypothesen unterstrichen werden. Dementsprechend wird Hypothesen,

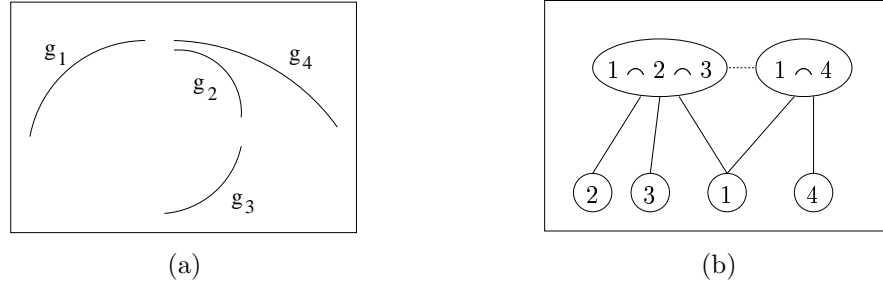


Abbildung 6.5.: Motivation zur Normierung des unterstützenden Potentials. (a) Ausschnitt einer initialen Kontursegmentierung. (b) Gruppierungsgraph mit den aus (a) entstehenden Knoten und Kanten.

die in einem konkurrierenden Kontext stehen und beide einen hohen Signifikanzwert beanspruchen, ein positives Potential zur Erhöhung der Gesamtenergie zugewiesen. Damit wird im Energieminimierungsprozeß eine Veränderung dieser Signifikanzwerte signalisiert. Im Gegensatz dazu wird ein negatives Potential gewählt, wenn eine Hypothese g_i einen hohen Signifikanzwert besitzt, während die im konkurrierenden Kontext auftretende Hypothese g_k einen niedrigen Signifikanzwert hat. Die Inkompatibilität konnte damit zu einer signifikanten und einer nicht signifikanten Interpretation aufgelöst werden und befindet sich in einem optimalen Zustand. Wie bei dem unterstützenden Potential, wird auch bei konkurrierenden Relationen ein neutrales Potential gewählt, falls keine der Hypothesen als signifikant interpretiert wird (siehe auch Abbildung 6.3(b)). Mit dieser Motivation ergibt sich das konkurrierende Potential durch eine Invertierung des unterstützenden Clique-Potentials:

$$V_{\mathbf{c}}^{comp}(\boldsymbol{\omega}) := -\varphi^{supp}(\omega_i, \omega_k), \quad \text{mit } \mathbf{c} = \{g_i, g_k\} \in \mathcal{C}_2^{comp} \quad (6.16)$$

Die in Abbildung 6.4 gezeigten Teilfunktionen werden hier nicht noch einmal explizit für ein konkurrierendes Potential abgebildet, da sie prinzipiell identisch sind. Es wird lediglich in der jeweiligen Teilfunktion die V -Achse negativ zum unterstützenden Potential aufgetragen.

Die Konkurrenzrelationen zwischen Geschlossenheiten bilden eine Ausnahme bezüglich der Potentialberechnung. Das konkurrierende Potential wird in diesem Fall zusätzlich mit einem Faktor gewichtet, der die Inkompatibilität bzw. die Inkonsistenz der Hypothesen berücksichtigt. Dieser Faktor δ nimmt in Abhängigkeit vom Überdeckungsgrad zwischen zwei Geschlossenheitshypothesen g_i, g_k zu:

$$\delta(g_i, g_k) := \frac{A(g_i) \cap A(g_k)}{\min(A(g_i), A(g_k))} \quad (6.17)$$

Die zusätzliche Berücksichtigung dieses Faktors im Clique-Potential bewirkt damit eine stärkere Konkurrenz von Geschlossenheiten, die zunehmend stärker überlappen. So

führt beispielsweise eine vollständige Überdeckung zu einer Verdoppelung des normalen Potentialwertes einer Konkurrenzrelation:

$$V_{\mathbf{c}}^{comp}(\boldsymbol{\omega}) := -(1 + \delta(g_i, g_k)) * \varphi^{supp}(\omega_i, \omega_k), \quad \text{mit } \mathbf{c} = \{g_i, g_k\} \in \mathcal{C}_2^{comp} \quad (6.18)$$

Diese Modellierung motiviert sich aus der Zielsetzung, flächenmäßig möglichst disjunkte Geschlossenheiten zu detektieren, die idealerweise Projektionen von Objektoberflächen beschreiben.

Nach der Konstruktionsvorschrift für den Gruppierungsgraphen entstehen auch Cliques höheren Grades im G-MRF. Diese Cliques werden hier nicht weiter berücksichtigt, bzw. wird für sie ein neutrales Clique-Potential gewählt. Dies begründet sich durch die Annahme, daß das Wissen über Kontextbezüge zwischen Gruppierungshypothesen in den Clique-Potentialen für zweielementige Cliques hinreichend abgebildet ist. Darüber hinaus lassen sich aufgrund der Vielfalt der möglichen internen Kontextbezüge von Gruppierungshypothesen in mehrelementigen Cliques a priori keine geeigneten Zielkonfigurationen für die Belegung der Label in solchen Cliques formulieren. Aus diesem Grund können auch keine entsprechenden Potentialfunktionen definiert werden. Damit ist die Beschreibung der Modellierung des a priori Wissen abgeschlossen und im folgenden wird die Modellierung der Datenabhängigkeit beschrieben.

6.2.3. Datenabhängigkeit

In dem Gruppierungs-MRF wird das Ziel verfolgt, eine Konfiguration $\boldsymbol{\omega}$ der Signifikanzwerte zu wählen, die der minimalen a posteriori Energie $U(\boldsymbol{\omega}|\mathbf{d})$ und damit einer MAP-Lösung der Bewertungsaufgabe entspricht. Für diese Bewertung der Gruppierungshypothesen ist neben dem beschriebenen a priori Wissen eine Modellierung einer Datenabhängigkeit der Signifikanzwerte einer Gruppierung erforderlich. Mit dieser Datenabhängigkeit kann dann die Formulierung eines a posteriori MRFs erfolgen.

In der Einleitung zu Abschnitt 6.2 wurde darauf eingegangen, daß bei der Generierung der Gruppierungshypothesen eine binäre Entscheidung über die Gruppierungswirkung von zwei Strukturen getroffen wird: Eine Gruppierung ist möglich, oder nicht. Nun wird diese binäre Entscheidung quantifiziert, um einen Hinweis zu erlangen, wie *gut* eine Hypothese allein aufgrund ihrer zugrundeliegenden Gruppierungsdaten die entsprechende Gruppierungswirkung erfüllt. Zu diesem Zweck wird eine *datengetriebene Signifikanz* $\underline{\omega}$ für die Gruppierungen eingeführt, die genau diese Quantifizierung leistet und in Analogie zu dem Signifikanzbegriff der Labelmenge eine Bewertung der Hypothese auf das Intervall $[0, 1]$ abbildet. Diese datengetriebene Signifikanz wird im folgenden für die unterschiedlichen Gruppierungstypen vorgestellt, wobei einheitlich das Symbol $\underline{\omega}$ gewählt wird, da sich die Berechnungsvorschrift implizit aus dem Gruppierungstyp ergibt. Die Wahl der einzelnen Beurteilungskriterien und deren Parametrisierung wird aus einer qualitativen Einschätzung der Gruppierungsleistung der einzelnen Generierungsschritte abgeleitet. Dies begründet sich aus der Tatsache, daß eine empirisch ermittelte Wahl

der Parameter, etwa durch eine Stichprobe, aufgrund der einflußnehmenden Faktoren sehr komplex ist: Es muß nicht nur eine binäre Entscheidung über die Gruppierungswirkung getroffen werden, sondern zusätzlich eine Quantifizierung derselben und im Zusammenhang damit, auch die Wechselwirkung mit anderen einflußnehmenden Gruppierungsphänomenen berücksichtigt werden. Eine mögliche Strategie und sinnvolle Erweiterung des Verfahrens für eine fundierte Parameterwahl könnte daher der Einsatz von geeigneten Lernverfahren sein. Dies wurde hier jedoch nicht weiter untersucht.

Für jeden Gruppierungstyp sind unterschiedliche Faktoren ausschlaggebend, die in eine Beurteilung der vorgenommenen Interpretation einer Hypothese eingehen:

Konturmodell: Die datengetriebene Signifikanz für Konturmodelle wird über den mittleren quadratischen Approximationsfehler $\bar{d}_\psi(P(g_i))$ (vergleiche Formel (2.1), Seite 17) der Kantenelemente P eines Konturmodells g_i errechnet. Für eine Normierung auf das Intervall $[0, 1]$ wird der maximal und minimal auftretende Approximationsfehler über alle Konturmodelle im Bild bestimmt. Die datengetriebene Signifikanz für jedes Konturmodell berechnet sich damit zu:

$$\underline{\omega}(g_i) := 1 - \frac{\bar{d}_\psi(P(g_i)) - \min_{g_k} \bar{d}_\psi(P(g_k))}{\max_{g_k} \bar{d}_\psi(P(g_k)) - \min_{g_k} \bar{d}_\psi(P(g_k))} \quad (6.19)$$

Hat ein Konturmodell den minimal auftretenden Approximationsfehler, so wird ihm eine datengetriebene Signifikanz von 1 zugeordnet.

Kollinearität/Kurvilinearität: Die datengetriebene Signifikanz für eine Kollinearität oder Kurvilinearität bezieht zwei Faktoren in ihre Beurteilung ein. Wie bereits in Abschnitt 4.2.2 erwähnt, werden Gruppierungshypothesen dieses Typs während der Generierung auf Basis ihrer konstituierenden Elemente erneut approximiert. Der dabei ermittelte Approximationsfehler liefert ein hinreichend quantifizierendes Maß, wie gut die zugrundeliegenden Daten durch eine Gruppierungshypothese dieses Typs interpretiert werden. Implizit wird dabei auch eine Beurteilung der Orientierungsdifferenz zwischen den beteiligten linearen Gruppen vorgenommen. In Analogie zu den Konturmodellen ergibt sich damit der erste Faktor zu:

$$\underline{\omega}_1^{kol}(g_i) := 1 - \frac{\bar{d}_\psi(P(g_i)) - \min_{g_k} \bar{d}_\psi(P(g_k))}{\max_{g_k} \bar{d}_\psi(P(g_k)) - \min_{g_k} \bar{d}_\psi(P(g_k))} \quad (6.20)$$

Auch hier erfolgt die Normierung über die Extremwerte aller Kollinearitäten bzw. Kurvilinearitäten im Bild.

Bei der Beurteilung durch $\underline{\omega}_1^{kol}(g_i)$ wird nur die Projektion der gruppierten Kantenelemente auf die entsprechend parametrisch beschriebene Gruppierung berücksichtigt. Zusätzlich soll nun der Abstand *zwischen* den einzelnen Konturmodellen, die die Gruppierung bilden, berücksichtigt werden. Hierzu wird die mittlere

Lückenlänge $\bar{g}(g_i)$ zwischen den Endpunkten, sowie die durchschnittliche analytische Länge $\bar{l}(g_i)$ der gruppierten Konturmodelle ermittelt. Diese Werte werden zur Berechnung des zweiten beeinflussenden Faktors eingesetzt:

$$\underline{\omega}_2^{kol}(g_i) := \max\left(\frac{\bar{l}(g_i) - \bar{g}(g_i)}{\bar{l}(g_i)}, 0\right) \quad (6.21)$$

Die Berechnung und Bezeichnung für kurvilineare Gruppen $\underline{\omega}_{1,2}^{kur}$ erfolgt entsprechend. Die datengetriebene Signifikanz für Kollinearitäten bzw. Kurvilinearitäten wird nun durch eine gewichtete Linearkombination dieser beiden Faktoren ermittelt:

$$\begin{aligned} \underline{\omega}(g_i) &:= \alpha_1 \underline{\omega}_1^{kol}(g_i) + \alpha_2 \underline{\omega}_2^{kol}(g_i), \quad \text{bzw.} \\ \underline{\omega}(g_i) &:= \alpha_1 \underline{\omega}_1^{kur}(g_i) + \alpha_2 \underline{\omega}_2^{kur}(g_i) \end{aligned} \quad (6.22)$$

Um den Wertebereich auf das Intervall $[0, 1]$ zu beschränken sind die Gewichte dabei so zu wählen, daß $\alpha_1 + \alpha_2 = 1$ gilt.

Proximität: Für die Berechnung der datengetriebenen Signifikanz einer Proximität wird alleine der Abstand zwischen den Endpunkten der beiden konstituierenden linearen Gruppen verwendet. Damit ergibt sich in Analogie zu Definition (6.21):

$$\underline{\omega}(g_i) := \max\left(\frac{\bar{l}(g_i) - g(g_i)}{\bar{l}(g_i)}, 0\right) \quad (6.23)$$

Hierbei bezeichnet $g(g_i)$ die Lückenlänge zwischen den Endpunkten der linearen Gruppen.

Parallelität: Die datengetriebene Signifikanz einer Parallelitätshypothese g_i setzt sich aus drei verschiedenen Bewertungskriterien zusammen. Dabei wird insbesondere die räumliche Relation der zwei Geradenstücke g_1 und g_2 , die die Gruppe bilden, untersucht (siehe Abbildung 6.6 für eine Motivation):

- $\angle(g_1, g_2)$, die Orientierungsdifferenz zwischen den beiden Geradenstücken.
- $\bar{d}_{norm}(g_1, g_2)$, der normierte mittlere Abstand zwischen den beiden Geradenstücken.
- $\tau(g_1, g_2)$, der Überlappungsbereich in dem sich beide Geradenstücke vollständig überlagern, wenn sie aufeinander projiziert werden.

Die Berücksichtigung der Orientierungsdifferenz erfolgt in ähnlicher Weise, wie bei der Generierung der Parallelitätshypothese (vergleiche Abschnitt 4.2.1). Das dabei zugelassene Intervall von $[0, \Delta_\phi(\parallel)]$ wird auch hier verwendet, um eine Beurteilung der Orientierungsdifferenz zwischen g_1 und g_2 auf das Intervall $[0, 1]$ abzubilden:

$$\underline{\omega}_1^{par}(g_i) := 1 - \frac{\angle(g_1, g_2)}{\Delta_\phi(\parallel)} \quad (6.24)$$

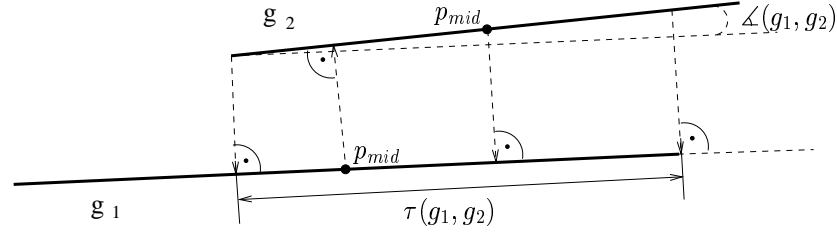


Abbildung 6.6.: Faktoren der datengetriebenen Signifikanz bei Parallelitäten.

Die Beurteilung des Abstandes zwischen den Geradenstücken wird ebenfalls auf Basis der Stichprobe zur Bestimmung der Aufmerksamkeitsbereiche vorgenommen. Aus der Stichprobe für Parallelitätshypothesen wird der normierte mittlere Abstand zu $\Delta_{\bar{d}}^{(II)} = 0.157$ bestimmt und ins Verhältnis zum normierten mittleren Abstand $\bar{d}_{norm}(g_2, g_1)$ in der aktuellen Hypothese gesetzt:

$$\underline{\omega}_2^{par}(g_i) := \begin{cases} \frac{\bar{d}_{norm}(g_2, g_1)}{\Delta_{\bar{d}}^{(II)}} & , \text{ wenn } \bar{d}_{norm}(g_2, g_1) < \Delta_{\bar{d}}^{(II)} \\ \frac{\Delta_{\bar{d}}^{(II)}}{\bar{d}_{norm}(g_2, g_1)} & , \text{ sonst} \end{cases} \quad (6.25)$$

Der mittlere Abstand errechnet sich aus dem Mittel der senkrechten Abstände der Mittelpunkte p_{mid} der beiden Geradenstücke zu dem jeweiligen komplementären Gegenstück. Die Normierung erfolgt mit der mittleren analytischen Länge der beiden Geradenstücke:

$$\bar{d}_{norm}(g_2, g_1) = \frac{d(p_{mid}(g_2), g_1) + d(p_{mid}(g_1), g_2)}{l(g_2) + l(g_1)}$$

Demnach liefert Definition (6.25) den höchsten anteiligen Signifikanzwert von 1, wenn gilt: $\bar{d}_{norm}(g_2, g_1) = \Delta_{\bar{d}}^{(II)}$. Für kleinere Abstände nimmt der Wert linear ab, während er bei größeren Abständen hyperbolisch abnimmt.

Als dritte Komponente berücksichtigt die datengetriebene Signifikanz für Parallelitäten den Überlappungsbereich $\tau(g_2, g_1)$, den beide Geradenstücke zueinander bilden. Zur Bestimmung dieses Wertes werden die Endpunkte der Geradenstücke wechselseitig auf das gegenüberliegende Geradenstück projiziert, und dann wird auf einem der beiden der Überlappungsbereich bestimmt (vergleiche Abbildung 6.6). Dieser Überlappungsbereich wird dann mit der mittleren analytischen Länge der beiden Geradenstücke normiert:

$$\underline{\omega}_3^{par}(g_i) := \frac{\tau(g_2, g_1)}{\bar{l}(g_i)} \quad (6.26)$$

Die ermittelten Werte werden dann durch eine gewichtete Linearkombination zur gesamten, datengetriebenen Signifikanz einer Parallelitätshypothese g_i kombiniert:

$$\underline{\omega}(g_i) := \beta_1 \underline{\omega}_1^{par}(g_i) + \beta_2 \underline{\omega}_2^{par}(g_i) + \beta_3 \underline{\omega}_3^{par}(g_i) \quad (6.27)$$

Auch hier sind die Gewichte so zu wählen, daß $\beta_1 + \beta_2 + \beta_3 = 1$ gilt.

Geschlossenheit: Die datengetriebene Signifikanz einer Geschlossenheitshypothese wird, wie schon bei den linearen Gruppen und Proximitäten, nur unter Berücksichtigung der Lückenlänge zwischen den Endpunkten der konstituierenden linearen Gruppen gebildet. Die mittlere Lückenlänge wird daher ebenfalls ins Verhältnis zur mittleren Länge der linearen Gruppen gesetzt:

$$\underline{\omega}(g_i) := \max\left(\frac{\bar{l}(g_i) - \bar{g}(g_i)}{\bar{l}(g_i)}, 0\right) \quad (6.28)$$

Mit der Definition der datengetriebenen Signifikanz für jeden Hypothesentyp kann nun – durch die Integration dieser rein datengetriebenen Bewertung – die a posteriori Wahrscheinlichkeit $P(\boldsymbol{\omega}|\mathbf{d})$ und damit die dazu äquivalente a posteriori Energie $U(\boldsymbol{\omega}|\mathbf{d})$ vollständig formuliert werden. Zu diesem Zweck wird zunächst zur Modellierung der Datenabhängigkeit die datengetriebene Signifikanz zur Definition der Potentialfunktion der einelementigen Cliques verwendet. Die Wahl dieser Potentialfunktion ist durch folgende Zusammenhänge motiviert. Unter der Annahme, daß die datengetriebene Signifikanz $\underline{\omega}(g_i)$ eine adäquate Bewertung der Gruppierungshypothese g_i aufgrund der Daten d_i leistet, so sollte der Wert für ω_i in einer Konfiguration $\boldsymbol{\omega}$ ohne Berücksichtigung des Kontextes gleich dem Signifikanzwert $\underline{\omega}(g_i)$ sein. Eine Abweichung von ω_i zu $\underline{\omega}(g_i)$, die sich dann aus dem Kontext ergeben muß, läßt die durch die Gruppierungshypothese vorgenommene Interpretation unwahrscheinlich werden. Diese schlechtere a posteriori Wahrscheinlichkeit ($P(d_i|\boldsymbol{\omega}) \ll 1$) wird durch eine Erhöhung des Potentials abgebildet. Sind die Signifikanzwerte jedoch ähnlich, liegt wahrscheinlich eine korrekte Interpretation vor, und die a posteriori Wahrscheinlichkeit für die Konfiguration wird vorwiegend durch das kodierte a priori Kontextwissen dominiert. Das entsprechende Potential verhält sich in diesem Fall neutral.

In Anlehnung an die Darstellungen in Abschnitt 6.1.1 erfolgt die Kodierung der a posteriori Wahrscheinlichkeit nun durch die Summe der Clique-Potentiale der einelementigen Cliques:

$$-T \sum_{i=1,\dots,n} \ln P(d_i|\boldsymbol{\omega}) = \sum_{i=1,\dots,n} V_{\mathbf{c}}^{data}(\boldsymbol{\omega}), \quad \text{mit } \mathbf{c} = \{g_i\} \in \mathcal{C}_1 \quad (6.29)$$

Das Clique-Potential für einelementige Cliques $V_{\mathbf{c}}^{data}(\boldsymbol{\omega})$ ergibt sich dabei entsprechend der obigen Motivation:

$$V_{g_i}^{data}(\boldsymbol{\omega}) := |\omega_i - \underline{\omega}(g_i)| \quad (6.30)$$

Das Potential erhöht also die a posteriori Energie, wenn eine Abweichung zwischen der datengetriebenen Signifikanz und der durch eine Konfiguration vorgegebenen Signifikanz vorliegt. Der Beitrag steigt dabei entsprechend dem Grad der Abweichung. Damit ist die Modellierung der Datenabhängigkeit vollständig beschrieben.

6.2.4. Energieminimierung mit HCF

Mit der Definition der Clique-Potentiale für die Datenabhängigkeit und die Kontextabhängigkeiten von Gruppierungshypothesen kann nun die a posteriori Energie für das Gruppierungs-MRF vollständig angegeben werden. Sie wird durch die Summierung der verschiedenen Potentiale gebildet:

$$U(\boldsymbol{\omega}|\mathbf{d}) = \sum_{\mathbf{c} \in \mathcal{C}_2^{supp}} V_{\mathbf{c}}^{supp}(\boldsymbol{\omega}) + \sum_{\mathbf{c} \in \mathcal{C}_2^{comp}} V_{\mathbf{c}}^{comp}(\boldsymbol{\omega}) + \sum_{\mathbf{c} \in \mathcal{C}_1} V_{\mathbf{c}}^{data}(\boldsymbol{\omega}) \quad (6.31)$$

Zur Minimierung dieser Energie wird hier das Verfahren des Highest Confidence First aus Abschnitt 6.1.2 (Seite 130) eingesetzt.

Die Definition des G-MRF mit der zugehörigen a posteriori Energie erfüllen insgesamt die Voraussetzungen zum Einsatz des HCF. Es kann eine lokale Energie entsprechend Definition (6.8) bestimmt werden und damit auch die Stabilität einer Signifikanzentscheidung ermittelt werden (vergleiche Definition (6.10)). Initial werden alle Gruppierungshypothesen ebenfalls mit einem Label $l_{uncommitted}$ zur Repräsentation eines *unentschiedenen* Zustandes belegt. Diese Zuweisung führt jedoch zu einer problematischen Situation in Bezug auf die initiale Stabilität jedes Knotens. Mit den Definitionen zur Berechnung der lokalen Energie $U_i(\omega_i|\mathbf{d}, \boldsymbol{\omega})$ und der Clique-Potentiale $V_{\mathbf{c}}'(\boldsymbol{\omega})$ wird die Stabilität nach Definition (6.10) für *alle* Knoten initial auf den Wert $S_i(\boldsymbol{\omega}) = -\Delta_{sig}$ gesetzt. Dies begründet sich dadurch, daß nach Definition (6.9) bereits mit *einem* unentschiedenen Knoten in einer Clique $V_{\mathbf{c}}'(\boldsymbol{\omega}) = 0$ gilt. Da zu Beginn alle Knoten unentschieden sind, reduziert sich die *initiale* lokale Energie (Definition (6.8)) für jeden Knoten s_i zu

$$U_i(\omega_i|\mathbf{d}, \boldsymbol{\omega}) = -T \ln P(d_i|\omega_i) = V_{g_i}^{data}(\boldsymbol{\omega}) = |\omega_i - \underline{\omega}(g_i)|.$$

Diese Energie wird für jeden Knoten minimal, wenn als Label in Definition (6.10) die entsprechend datengetriebene Signifikanz gewählt wird $\hat{\omega} = \underline{\omega}(g_i)$. Die minimale Energiedifferenz ergibt sich dann für jeden Knoten mit einem Label $\omega' = \underline{\omega} \pm \Delta_{sig}$. Als Resultat erhalten alle Knoten als initiale Stabilität den Wert $S_i(\boldsymbol{\omega}) = -\Delta_{sig}$, womit keine sinnvolle Reihenfolge für die Bearbeitung der einzelnen Knoten ermittelt werden kann.

Aus diesem Grund wird hier eine andere initiale Stabilität für jeden Knoten gewählt. Sie motiviert sich aus der Definition der datengetriebenen Signifikanz:

$$S_i^{init} := -10 \cdot \underline{\omega}(g_i) \quad (6.32)$$

Mit dieser initialen Stabilität beginnt der Minimierungsprozeß bei den Gruppierungshypothesen, die die höchste datengetriebene Signifikanz besitzen. Es werden also die Hypothesen gewählt, die alleine aufgrund ihrer Daten eine signifikante Interpretation des Bildinhaltes zu sein scheinen. Der Vorfaktor dient einer höheren Gewichtung gegenüber bereits entschiedenen Knoten, die bereits ihre Stabilität aufgrund von Definition (6.10) bestimmt haben.

Mit der Beschreibung des Energieminimierungsprozesses ist die Vorstellung des Markov Random Fields zur Bewertung von Gruppierungshypothesen abgeschlossen. Nach einer Energieminimierung befindet sich das G-MRF in einem Energieminimum, welches eine gute Approximation des optimalen Minimums darstellt. Die entsprechende Konfiguration stellt dann eine relativ gute MAP-Lösung des Bewertungsproblems dar. Gruppierungshypothesen können damit anhand ihres Signifikanzwertes für eine weitere Bildanalyse selektiert werden. Die Darstellung der Ergebnisse, sowie die Wahl von Parametern erfolgt in Abschnitt 7.2.

6.3. MRF zur Kontursegmentierung in Bildfolgen

In dem bisher vorgestellten Gruppierungsansatz erfolgt die Detektion von signifikanten konturbasierten Strukturen nur in einzelnen Bildern, ohne die Gruppierungsinformation aus dem vorherigen Bild zu berücksichtigen. Dies resultiert in einer zeitaufwendigen Vorgehensweise und in einer isolierten Strukturierung des Bildinhalts für Zeitpunkte t , die in keinem zeitlichen Kontext zu vorher detektierten Strukturen zu Zeitpunkten $t - 1$ stehen. In diesem Abschnitt wird die Konzeption für ein Verfahren vorgestellt, das einen grundlegenden Baustein für die Berücksichtigung eines zeitlichen Kontextes beschreibt.

Die Kernidee des Konzeptes ist es, eine Kontursegmentierung in Bildsequenzen zu realisieren, die eine Identifikation von korrespondierenden primitiven Konturmodellen (Geradenstücke, Kreis-, Ellipsenbögen) in zeitlich aufeinanderfolgenden Bildern beinhaltet (siehe Abbildung 6.7 und [Cor00]). Hierzu wird zunächst auf jedem Bild eine unabhängige Konturdetektion durchgeführt (vergleiche Abschnitt 2.1.1). Die daraus entstandenen Kantenelemente p_i^t werden anschließend durch den Einsatz eines MRFs in einen zeitlichen Kontext mit Konturmodellen m_j^{t-1} aus vorherigen Bildern gebracht. In dem Schritt der Approximation wird abschließend eine Identifizierung korrespondierender Konturen vorgenommen $m_j^{t-1} \leftrightarrow m_j^t$. Hierbei können jedoch zunächst nur statische, kaum bewegte Konturmodelle berücksichtigt und als identische Bildprimitiva in einer Bildsequenz identifiziert werden. Insgesamt wird jedoch mit dieser Konzeption eine wichtige Grundlage für ein temporal ausgelegtes perzeptives Gruppieren geschaffen.

Für den Gruppierungsprozeß kann die Korrespondenzinformation in starkem Maße genutzt werden. Bei der Generierung können beispielsweise die Gruppierungen des letzten Bildes, die aus korrespondierend markierten Konturmodellen erzeugt wurden, als Vorerwartung für die zu erzeugenden Gruppierungen dienen. Damit reduziert sich der Suchaufwand nach potentiellen Gruppierungshypothesen, und für einzelne Hypothesen

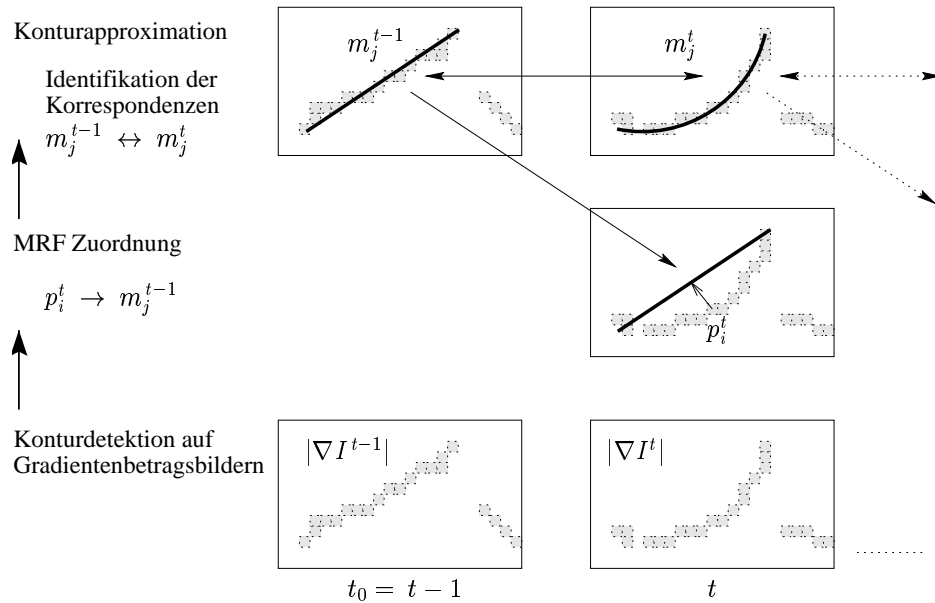


Abbildung 6.7.: Motivation zum Verfahren der Kontursegmentierung in Bildfolgen.

ergibt sich eine zusätzliche Evidenz für ihre Korrektheit aus dem zeitlichen Kontext. Diese Evidenz kann wiederum in dem Gruppierungs-MRF auf die Bewertung Einfluß nehmen. Im Rahmen dieser Arbeit wird auf diese Aspekte jedoch nicht eingegangen, da bisher keine algorithmische Umsetzung dieser Perspektiven erfolgt ist. Für die Kontursegmentierung resultiert aus der vorgeschlagenen Konzeption neben dem Zugewinn an Information auch eine deutlich kürzere Bearbeitungszeit der Einzelbilder.

Im folgenden wird zunächst auf die verschiedenen Bestandteile des MRFs eingegangen, das für eine Realisierung des Konzeptes eingesetzt wird. Dabei werden die Knotenmenge und die Labelmenge des Kontur-MRF (K-MRF), sowie das zugeordnete Nachbarschaftssystem definiert. Anschließend werden die Kontextabhängigkeiten, die Datenabhängigkeit und die eingesetzte Methodik zur Bestimmung der entsprechenden Parameter beschrieben. Die Energieminimierung wird ebenfalls mit dem *Highest Confidence First* Verfahren realisiert (vergleiche Abschnitt 6.1.2). Hierdurch wird dann implizit die Identifizierung korrespondierender Konturen respektive ihrer Kantenelemente vorgenommen. Abschließend erfolgt die Beschreibung, wie die Ergebnisse der MRF-Anwendung für eine Detektion zeitlich korrespondierender Konturmodelle genutzt wird. Zur klaren Unterscheidung zeitlicher Zuordnungen wird folgende Notation für Bildprimitiva \mathbf{b} und Bilder I eingeführt. Für ein Symbol wird der zugehörige Zeitpunkt durch einen hochgestellten Index markiert: \mathbf{b}^t markiert das Bildprimitivum \mathbf{b} , das aus dem Bild I^t vom Zeitpunkt t ermittelt worden ist.

6.3.1. Konturgraph und Labelmenge

Die Definition der Stellen und des Nachbarschaftssystems wird für diesen Ansatz ähnlich zu anderen MRF-Ansätzen aus dem low-level Bereich der Bildverarbeitung gewählt (vergleiche etwa [Li95]). Die Stellen sind durch die Kantenelemente eines Gradientenbetragsbildes definiert, die in einer unabhängigen Konturdetektion ermittelt werden. Beschreibe P^t die Menge der Kantenelemente in dem Bild I^t , so ist die Menge der Stellen zu $\mathcal{S}^t = \{p_i | p_i \in P^t\}$ definiert. Das eingesetzte Nachbarschaftssystem motiviert sich in naheliegender Weise über die zugrundeliegende Bildmatrix. In dem hier beschriebenen K-MRF wird die Nachbarschaft zwischen den Kantenelementen auf Basis einer 8-er Nachbarschaft definiert. In Analogie zu dem Gruppierungsgraph läßt sich damit ein ungerichteter Konturgraph aufstellen, in dem die Kantenelemente als Knoten und die 8-er Nachbarschaftsbeziehung zu anderen Kantenelementen durch Kanten repräsentiert sind.

Labelmenge

Mit der Definition der Stellen und des Nachbarschaftssystems stellt sich nun die Frage nach der Definition der Labelmenge. Es liegt nahe hier die Kantenelemente aus dem Bild I^{t-1} als Labelmenge für die Stellen des aktuellen Bildes zu verwenden. Damit wird jedoch eine sehr große Labelmenge für jedes Kantenelement erzeugt, die zu einer unüberschaubaren Anzahl von möglichen Konfigurationen $\omega \in \Omega$ führen würde. Aus diesem Grund wird hier eine andere Labelmenge definiert, die auch mehr dem Charakter einer Interpretation der Stellen entspricht.

In Abschnitt 2.1.3 wird das Verfahren zur Konturapproximation von Leonardis vorgestellt, das auch hier wieder zur Approximation der Kantenelemente zum Einsatz kommt. In dem Verfahren wird über eine *Objective Function* (Gleichung (2.5)) eine Menge \mathbf{m} von Konturmodellen definiert, die die im Bild enthaltenen Kantenelemente optimal approximieren. Diese Menge von Konturmodellen stellt damit eine wesentlich abstraktere Bildbeschreibung dar. Aus diesem Grund werden den Stellen \mathcal{S}^t des K-MRFs die Konturmodelle \mathbf{m}^{t-1} aus dem Bild I^{t-1} als Labelmenge zugewiesen:

$$\mathcal{L}^t := \{m_i^{t-1} | m_i^{t-1} \in \mathbf{m}^{t-1}\} \quad (6.33)$$

Mit der Repräsentation jedes Kantenelementes durch eine Stelle bleibt aber auch bei dieser Wahl von \mathcal{L} immer noch ein großer Konfigurationsraum Ω bestehen. Ohne die formalen Definitionen zu verletzen, läßt sich jedoch die Labelmenge in positionsabhängige Teilmengen aufteilen. Mit dieser Vorstellung wird daher für jedes Kantenelement p_i^t folgende Labelmenge gewählt:

$$\mathcal{L}_i^t = \{m_j^{t-1} | m_j^{t-1} \in \mathbf{m}^{t-1} \wedge p_i^t \in R_j\} \cup \{l_{unresolved}\} \quad (6.34)$$

Dabei beschreibt R_j einen Bereich um das Modell m_j^{t-1} im Bild I^t , der zur Einschränkung der Labelmenge für p_i^t eingesetzt wird. Alle Kantenelemente p_i^t , die innerhalb des Bereiches R_j lokalisiert sind, enthalten in ihrer Labelmenge \mathcal{L}_i^t das Konturmodell m_j^{t-1} . Die

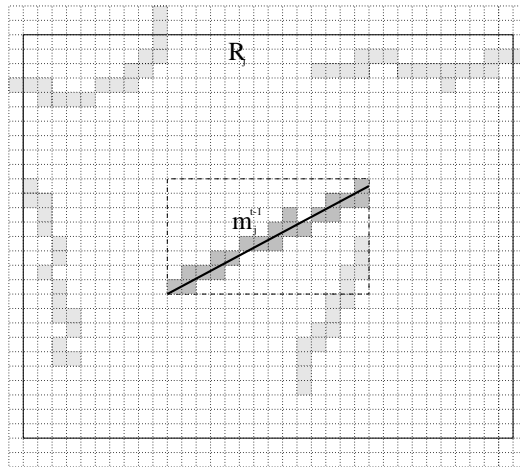


Abbildung 6.8.: Motivation zur Einschränkung der Labelmenge eines Kantenelementes. Der Einflußbereich R_j von m_j^{t-1} ist mit einer durchgezogenen Linie markiert. Für alle Kantenelemente p_i^t , die innerhalb dieses Bereiches lokalisiert sind, ist m_j^{t-1} Element der Labelmenge \mathcal{L}_i^t .

Ausdehnung dieses Einflußbereiches läßt sich beispielsweise in Experimenten auf Bildsequenzen mit statischen Szenen ermitteln. Vereinfacht kann dann der Einflußbereich zum Beispiel auf das umschließende Rechteck eines Konturmodells zuzüglich eines Offsets in jede Richtung festgelegt werden (siehe Abbildung 6.8). Er könnte aber auch einem Korridor um den Verlauf des Konturmodells m_j^{t-1} entsprechen. Des weiteren können auch Kantenelemente im Bild auftreten, die in keinem der Einflußbereiche der Konturmodelle \mathbf{m}^{t-1} enthalten sind. Um diesen Zustand ebenfalls abbilden zu können, wird ein Label $l_{unresolved}$ eingeführt, das genau dieser Interpretation entspricht. Mit dieser Definition der Labelmengen und der Stellen wird im folgenden Abschnitt die Modellierung der Kontextabhängigkeiten und der Datenabhängigkeit beschrieben.

6.3.2. Modellierung des a priori und a posteriori Wissens

Für die Zuordnung von Konturmodellen des Zeitpunktes $t - 1$ zu Kantenelementen des Zeitpunktes t kann a priori Wissen über Abhängigkeiten einer solchen Zuordnung formuliert werden. Wie bei dem Gruppierungs-MRF erfolgt dies auch hier über die Clique-Potentiale V_c , wobei ebenfalls auf die explizite Modellierung des a priori Wissens in den *einelementigen* Cliquen verzichtet wird (vergleiche Abschnitt 6.2.2). Die Modellierung des a posteriori Wissens erfolgt analog durch zusätzliche Summanden für die einelementigen Cliquen. Sie beurteilen im Wesentlichen, ob ein Kantenelement p_i^t mit seinen Daten zu einem Konturmodell m_j^{t-1} gehören könnte.

Kontextabhängigkeiten

Das *zweielementige* Clique-Potential kodiert das Wissen über den Kontext, in dem die Zuweisung eines Konturmodells zu einem Kantenelement beurteilt werden muß. Ein Konturmodell m^{t-1} ist aus mehreren benachbarten Kantenelementen p^{t-1} generiert worden. Es liegt daher nahe, daß eine Zuweisung eines Modells m_j^{t-1} zu einem Kantenelement p_i^t umso wahrscheinlicher ist, wenn in der Nachbarschaft N_i ebenfalls das gleiche Konturmodell m_j^{t-1} zugewiesen ist. Der Potentialwert für eine Clique mit dieser Interpretation sollte also klein gewählt werden. Sind jedoch zwei benachbarten Kantenelementen unterschiedliche Konturmodelle zugewiesen, so entspricht dies einer inkonsistenten Interpretation und der Potentialwert dieser Clique wird entsprechend größer gewählt. Auch die Zuweisung des Labels $l_{unresolved}$ zu einem der Kantenelemente in einer Clique führt zu einem erhöhten Potentialwert, da die Entscheidung für ein Label aus \mathcal{L}^t unterstützt werden soll. Aus dieser Motivation werden die Potentiale für zweielementige Cliquen folgendermaßen definiert:

$$V_{\mathbf{c}}(\boldsymbol{\omega}) := \begin{cases} v_{cor} & , \text{ wenn } \omega_i = \omega_j \wedge \omega_i, \omega_j \neq l_{unresolved} \\ v_{err} & , \text{ wenn } \omega_i \neq \omega_j \wedge \omega_i, \omega_j \neq l_{unresolved} \\ v_{res} & , \text{ sonst} \end{cases} \quad (6.35)$$

Eine Methode zur Wahl der Parameter v wird in Abschnitt 6.3.3 näher erläutert, da zur Modellierung der Datenabhängigkeit ähnliche Parameter eingesetzt werden, und die Bestimmung dieser Parameter mit der gleichen Methode erfolgen kann. Aus obiger Motivation läßt sich jedoch jetzt schon folgende Relation der einzelnen Parameter angeben:

$$v_{cor} < v_{res} < v_{err}$$

Dies bringt zum Ausdruck, daß eine inkonsistente Interpretation mit einem hohen Potentialwert bestraft wird, wohingegen eine Interpretation als „nicht zuzuordnendes“ Kantenelement weniger starke Auswirkungen hat. Eine korrekte Zuweisung eines Konturmodells zu einem Kantenelement, soll die wahrscheinlichste Konfiguration sein und entspricht daher dem geringsten Potentialwert.

Die Potentialwerte *höherwertiger* Cliquen werden in dieser Konzeption neutral gewählt. Hierfür wird, wie bei den Bewertung von Gruppierungshypothesen, von der Annahme ausgegangen, daß mit der Definition der Kontextabhängigkeit über die zweielementigen Clique-Potentiale eine ausreichende Modellierung des a priori Wissens gewählt wird.

Datenabhängigkeit

Die Modellierung des a posteriori Wissen für die Zuordnungsaufgabe motiviert sich in naheliegender Weise aus den Verfahren zur Konturapproximation. Eine Zuordnung eines Konturmodells m_j^{t-1} als abstraktere Beschreibung des Kantenelementes p_i^t ist umso

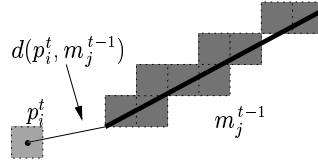


Abbildung 6.9.: Motivation des Abstandes $d(p_i^t, m_j^{t-1})$ zur datengetriebenen Beurteilung einer Zuweisung eines Konturmodells zu einem Kantenelement.

wahrscheinlicher, je besser die Einschränkungen aus den Verfahren zur Konturapproximation erfüllt werden. Allerdings muß aufgrund des zeitlichen Versatzes eine stärkere Abweichung von der idealen Interpretation zugelassen werden. Diese höheren Abweichungen begründen sich bei statischen Szenen in erster Linie durch Rauscheffekte der eingesetzten Hardware. Aus dieser Motivation ergibt sich *ein* Anteil der datengetriebenen Beurteilung einer Zuweisung zu:

$$\underline{\omega}_1(p_i^t) := \frac{d(p_i^t, m_j^{t-1}) - \min_k d(p_k^t, m_j^{t-1})}{\max_k d(p_k^t, m_j^{t-1}) - \min_k d(p_k^t, m_j^{t-1})}, \text{ mit } p_k^t \in R_j \quad (6.36)$$

In dieser Definition wird der kürzeste Abstand eines Kantenelementes zu einem Konturmodell mit $d(p_i^t, m_j^{t-1})$ bezeichnet (siehe Abbildung 6.9). Dieser Anteil der datengetriebenen Beurteilung nimmt damit kleine Werte für Kantenelemente p_i^t an, die sehr nah an dem Konturmodell m_j^{t-1} lokalisiert sind und große Werte für weiter entfernte Kantenelemente. Die Normierung des Wertes auf ein Intervall $[0, 1]$ bezieht sich nur auf den Einflußbereich R_j des Konturmodells m_j^{t-1} und erzeugt dadurch einen lokalen Charakter der Bewertung.

Als weitere Komponente zur datengetriebenen Beurteilung einer Zuweisung wird die Abweichung des Gradientenbetrages $|\nabla I_i^t|$ eines Kantenelementes p_i^t zu einem lokalen Mittelwert auf dem Konturmodell m_j^{t-1} berücksichtigt. Zu diesem Zweck wird das Kantenelement p_i^t über die kürzeste Distanz auf das Konturmodell projiziert und in einem Bereich D_q um den Projektionspunkt q der mittlere Gradientenbetrag $|\overline{\nabla I}_q^{t-1}|$ ermittelt. Die Ausdehnung des Bereiches D_q kann beispielsweise auf einer Stichprobe ermittelt werden. Sei $P_{j_q}^{t-1}$ die Menge der Kantenelemente, die von dem Konturmodell m_j^{t-1} im Bereich D_q approximiert werden, so ergibt sich der lokale mittlere Gradientenbetrag zu:

$$|\overline{\nabla I}_q^{t-1}| = \frac{1}{|P_{j_q}^{t-1}|} \sum_{p_k \in P_{j_q}^{t-1}} |\nabla I_k^{t-1}| \quad (6.37)$$

Mit diesem lokalen mittleren Gradientenbetrag aus dem vorherigen Bild wird der zweite

Anteil der datengetriebenen Beurteilung einer Zuweisung ermittelt:

$$\underline{\omega}_2(p_i^t) := \frac{d_\nabla(p_i^t, m_j^{t-1}) - \min_k d_\nabla(p_k^t, m_j^{t-1})}{\max_k d_\nabla(p_k^t, m_j^{t-1}) - \min_k d_\nabla(p_k^t, m_j^{t-1})}, \quad \text{mit } p_k^t \in R_j \text{ und} \quad (6.38)$$

$$d_\nabla(p_i^t, m_j^{t-1}) := \left| |\nabla I_i^t| - |\overline{\nabla I_q^{t-1}}| \right|$$

Dieser Anteil wird ebenfalls auf ein Intervall $[0, 1]$ bezüglich des Bereiches R_j normiert. Mit diesem Anteil wird damit anschaulich die Konstanz der Gradientenbeträge über die Zeit in einem lokalen Umfeld der Konturmodelle beurteilt.

Die beiden datengetriebenen Beurteilungen werden schließlich zur Formulierung des Potentials für die a posteriori Energie kombiniert:

$$V_{p_i^t}^{data}(\omega) := \begin{cases} v_{data} & , \text{ wenn } \omega_i = l_{unresolved} \\ \frac{\omega_1(p_i^t) + \omega_2(p_i^t)}{2} & , \text{ sonst} \end{cases} \quad (6.39)$$

Auch die Zuordnung des Labels $l_{unresolved}$ muß zur Modellierung der Datenabhängigkeit berücksichtigt werden, da idealerweise natürlich alle Kantenelemente ein Konturmodell zugewiesen bekommen sollten. Dies erfolgt hier durch einen konstanten Wert v_{data} , für dessen Bestimmung in Abschnitt 6.3.3 eine Methode vorgeschlagen wird.

Mit der Formulierung der Datenabhängigkeit ist die Kodierung des a priori und a posteriori Wissens vollständig beschrieben. Ein Vorschlag zur Parametrisierung der Clique-Potentiale wird ebenfalls im nächsten Abschnitt beschrieben. Unabhängig davon läßt sich nun, in Analogie zum MRF für die Bewertung von Gruppierungshypothesen, die a posteriori Energie folgendermaßen formulieren:

$$U(\omega|\mathbf{d}) = \sum_{c \in \mathcal{C}_2} V_c(\omega) + \sum_{c \in \mathcal{C}_1} V_c^{data}(\omega), \quad \text{mit } \mathcal{C}_1 = P^t \quad (6.40)$$

6.3.3. Parameterbestimmung

Im Forschungsumfeld von MRFs wird der Parametrisierung von Potentialfunktionen viel Aufmerksamkeit geschenkt. Wie bereits bei der formalen Einführung zu MRFs herausgestellt wurde, wird in den Potentialfunktionen das gesamte a priori Wissen über eine Problemstellung kodiert. Die Wahl dieser Potentiale ist daher von entscheidender Bedeutung für eine erfolgreiche Lösung eines Problems. In dieser Konzeption wird vorgeschlagen, die Parameter der Clique-Potentiale aus Untersuchungen einer Stichprobe abzuleiten. Im Gegensatz zur Modellierung der Clique-Potentiale im G-MRF, für das die Ableitung der Potentialwerte aus einer Stichprobe sehr schwer zu realisieren ist, da a priori keine Vorgaben aufgrund der komplexen Wechselwirkungen gemacht werden können, ist dies in dem K-MRF möglich.

In der Untersuchung einer Stichprobe wird zwischen der Bestimmung des Parameters v_{data} für das datenabhängige Potential und der Parameter v_{cor} , v_{err} und v_{res} für

die zweielementigen Clique-Potentiale unterschieden. Die grundlegende Idee für die Untersuchung ist jedoch identisch. Auf einer Anzahl Testbildern wird zunächst eine Kontursegmentierung durchgeführt. Anschließend wird für jedes Testbild eine Bildfolge der Länge zwei erzeugt, die aus zwei identischen Bildern $I^0 = I^1$ besteht. Damit kann eine eindeutige Abbildung von Kantenelemente zwischen zwei Bildern gewährleistet werden $p_i^0 = p_i^1$, die nicht durch Rauschen gestört ist. Die Parameter der Potentiale sollen nun so festgelegt werden, daß in dem K-MRF die korrekte Konfiguration mit minimaler a posteriori Energie gewählt wird, die jedem Kantenelement p_i^1 genau das Konturmodell m_j^0 als Label zuweist, das auch p_i^0 approximiert. Mit dieser Vorgehensweise kann anschließend eine qualitative Beurteilung der Labelzuordnungen vorgenommen werden. Eine *richtige* Zuordnung liegt vor, wenn das Kantenelement p_i^1 das Label m_j^0 erhält, das auch schon p_i^0 approximiert hat. Hingegen liegt eine *falsche* Zuordnung vor, wenn p_i^1 als Label m_k^0 oder $l_{unresolved}$ erhält, obwohl $p_i^0 \in m_j^0$ und $j \neq k$ gilt. Jedoch werden bei einer Konturapproximation nicht alle Kantenelemente durch ein Konturmodell beschrieben. Für diese nicht approximierten Kantenelemente wird die Zuordnung von $l_{unresolved}$ als *richtig* und jede andere Belegung als *falsch* interpretiert. Eine Auswertung der Anzahl *falscher* und *richtiger* Zuordnungen über alle verwendeten Bildfolgen gibt Hinweise auf die Veränderung der Parameter für die Potentialfunktionen.

Potential bei fehlender Datenabhängigkeit

Das Potential für die einelementigen Cliques des K-MRF modelliert die Datenabhängigkeit einer Labelzuweisung und ist in Definition (6.39) angegeben. Demnach ist das Potential entweder von den zugrundeliegenden Daten abhängig, oder wird auf den konstanten Wert v_{data} gesetzt, falls die Datenabhängigkeit fehlt. Der Wert von v_{data} unterliegt dabei zwei qualitativen Einschränkungen:

1. Die Wahl von $l_{unresolved}$ muß unwahrscheinlicher sein, als eine *richtige* Zuordnung eines Konturmodells zu einem Kantenelement. Daher muß der Wert von v_{data} größer sein als der Potentialwert, der sich aus den zugrundeliegenden Daten ergibt.

$$v_{data} > \frac{\omega_1 + \omega_2}{2} \quad , \text{ falls } p_i^1 \rightarrow m_j^0 \wedge p_i^0 \in m_j^0$$

2. Die Wahl von $l_{unresolved}$ soll hingegen wahrscheinlicher sein, als eine *falsche* Zuordnung. In diesem Fall soll der Wert von v_{data} kleiner sein als der Potentialwert, der sich aus den zugrundeliegenden Daten ergibt.

$$v_{data} < \frac{\omega_1 + \omega_2}{2} \quad , \text{ falls } p_i^1 \rightarrow m_k^0 \wedge p_i^0 \in m_j^0 \wedge j \neq k$$

Hierbei zählt auch die Zuweisung eines Konturmodells zu einem in I^0 nicht approximierten Kantenelement als *falsche* Interpretation.

Zur Bestimmung von v_{data} werden nun dessen Werte im Intervall $[0, 1]$ bei festem zweielementigen Cliquepotential variiert, und eine Auswertung der *richtigen* und *falschen* Zuordnungen von Labeln unter obigen Randbedingungen für alle verwendeten Testbilder durchgeführt. Aus dieser Statistik wird der Wert für v_{data} dann so gewählt, daß die Anzahl *richtiger* Zuordnungen maximiert und gleichzeitig die Anzahl *falscher* Labelzuordnungen im Sinne obiger Interpretation minimiert wird.

Potentiale für Kontextabhängigkeiten

Das Clique-Potential zur Modellierung von Kontextbezügen zwischen den Kantenelementen ist in Definition (6.35) angegeben. Dort wird auch eine Motivation für die qualitative Modellierung der Potentialwerte gegeben: Eine kompatible Zuordnung soll eine hohe a priori Wahrscheinlichkeit haben, eine konträre hingegen eine deutlich geringere.

Die Bestimmung der Potentialwerte für diese Konstellationen erfolgt ebenfalls auf einer Stichprobe von Testbildern. Allerdings ist die Ermittlung der Werte v deutlich komplexer und nicht so direkt möglich, wie bei v_{data} . Es müssen Wechselwirkungen der Cliques untereinander und mit den datenabhängigen Potentialen berücksichtigt werden. Aus der Motivation der qualitativen Modellierung lassen sich zunächst folgende Zusammenhänge spezifizieren. Der Wert von v_{cor} interpretiert eine kompatible Zuweisung eines Konturmodells zu beiden Kantenelementen. Dies entspricht der Standardsituation, daß zwei benachbarte Kantenelemente in der Regel durch das *gleiche* Konturmodell approximiert werden. Es soll deutlich wahrscheinlicher sein, als eine Zuweisung des Labels $l_{unresolved}$ zu einem oder zu beiden Kantenelementen ($\Rightarrow v_{cor} < v_{res}$). In solchen Situationen ist noch keine vollständig korrekte Interpretation erfolgt. Die unwahrscheinlichste Konfiguration liegt jedoch vor, wenn zwei benachbarte Kantenelemente durch unterschiedlich Konturmodelle interpretiert werden ($\Rightarrow v_{res} < v_{err}$). Diese Zuordnung sollte auf jedenfall im Minimierungsprozeß eine Veränderung auslösen.

Aus diesen Zusammenhängen kann folgende Strategie zur Bestimmung der Parameter aus der Stichprobe vorgeschlagen werden. Die einzelnen Potentialwerte werden in bestimmten Intervallen variiert, die die obigen Einschränkungen bezüglich der beiden anderen Potentialwerte erfüllen. Dabei werden diese anderen Potentialwerte jeweils konstant gewählt. Eine Auswertung der erzielten Ergebnisse führt schließlich zur Wahl der Potentialwerte, die die geringste Anzahl *falscher* Labelzuordnungen auslöst. Mit diesem empirischen Vorgehen kann dann ein entsprechender Parametersatz für die Potentialwerte bestimmt werden. Damit ist der Vorschlag für eine Parametrisierung der ein- und zweielementigen Clique-Potentiale beschrieben.

6.3.4. Detektion korrespondierender Konturmodelle

Das Ziel dieses Ansatzes ist die Identifizierung korrespondierender Konturmodelle in Bildfolgen. Zu diesem Zweck wird nun das vorgestellte K-MRF eingesetzt.

In Abschnitt 2.1 ist ausführlich das Verfahren der Kontursegmentierung auf einzelnen Bildern vorgestellt worden. Für die Segmentierung in Bildfolgen wird der Schritt der Konturdetektion weiterhin auf jedem Bild unabhängig von einem zeitlichen Kontext durchgeführt. Dies motiviert sich durch die pixelorientierten Verfahren, die bei diesen Teilschritten zum Einsatz kommen und nur in geringem Maße von einer Integration bereits vorhandener Information profitieren würden. Daher wird hier die Integration in dem Schritt der Approximation von Kantenelementen in Konturmodelle mit dem Verfahren von Leonardis (vergleiche Abschnitt 2.1.3) vorgeschlagen. Damit stellt sich der Prozeß der Kontursegmentierung in Bildfolgen folgendermaßen dar (siehe auch Abbildung 6.7).

Auf einer Bildfolge $\mathcal{I} = \{I^0, \dots, I^N\}$ wird initial auf dem Bild I^0 eine Kontursegmentierung, wie in Abschnitt 2.1 beschrieben, durchgeführt. Für jedes folgende Bild I^t wird nur noch die Kantendetektion in ein Gradientenbetragsbild $|\nabla I^t|$ durchgeführt. Die Kantenelemente dieses Bildes bilden die Menge der Stellen \mathcal{S}^t eines K-MRFs. Die Labelmenge ist durch die Konturmodelle \mathbf{m}^{t-1} des letzten Bildes I^{t-1} spezifiziert. Auf dem so definierten K-MRF wird eine Energieminimierung mit dem HCF-Verfahren durchgeführt. Dabei wird das HCF-Verfahren, im Gegensatz zum Gruppierungs-MRF, genau so angewendet, wie es in der Originalarbeit von Chou und Brown in [Cho90] und in Abschnitt 6.1.2 beschrieben wird. Ist mit dem HCF-Verfahren eine Konfiguration ermittelt worden, die einer MAP-Lösung des K-MRF entspricht, so sind den Kantenelementen $p_i^t \in |\nabla I^t|$ die Konturmodelle \mathbf{m}^{t-1} aus dem Vorgängerbild zugewiesen. In dieser Phase ist damit die Zuordnung von Informationen aus den zeitlich versetzten Bildern etabliert. Sie kann nun genutzt werden, um die Approximation der Kantenelemente im aktuellen Bild zu unterstützen und eine Identifizierung identischer Konturmodelle zu realisieren.

Für die Approximation der Kantenelemente p_i^t mit dem Verfahren nach Leonardis wird hier der Schritt der *Model Recovery* angepaßt (vergleiche Algorithmus 2.1, Seite 19, und Algorithmus 6.1, Seite 159). Im Originalverfahren werden die initialen Konturmodelle aus den Kantenelementen der fest vorgegebenen Saatfenster ermittelt und sukzessive um kompatible Kantenelemente erweitert. Dieser Schritt wird nun zunächst umgangen. In der initialen Phase werden alle Kantenelemente aufgesammelt, denen das gleiche Konturmodell m_j^{t-1} zugeordnet worden ist. Damit wird die Bestimmung der initialen Saaten P für eine Konturapproximation im Bild I^t durch die Zuordnung der Konturmodelle m_j^{t-1} aus dem Bild I^{t-1} bestimmt. Diese initialen Saaten werden dann unter Verwendung des Approximationstyps (Geradenstück, Kreis- oder Ellipsenbogen) des Modells m_j^{t-1} neu approximiert und so ein neues initiales Modell m_j^t ermittelt. Anschließend werden die einschränkenden Bedingungen des Originalverfahrens ausgewertet und gegebenenfalls der Modelltyp alterniert, um eine adäquate Approximation zu erzielen. Kann für ein initiales Modell m_j^t kein Parametersatz ermittelt werden, der die einschränkenden Bedingungen erfüllt, werden die entsprechenden Kantenelemente zunächst mit dem Label $l_{unresolved}$ belegt (siehe Algorithmus 6.1).

Nachdem diese Phase einer integrierenden Approximation für alle Modelle \mathbf{m}^{t-1} abgeschlossen ist, startet der Prozeß der *Model Recovery* für alle Kantenelemente, die das Label $l_{unresolved}$ aufweisen, in seiner ursprünglichen Version. Ebenfalls analog zu dem

Model Recovery with Correspondence

```

 $M^t = \emptyset$  {Menge der Konturmodelle}
for all  $m_j^{t-1} \in \mathbf{m}^{t-1}$  do
     $M^t = M^t \cup$  Initial Model ( $m_j^{t-1}$ ) {Zuordnung Konturmodelle ermitteln}
end for

for all Saatfenster do
     $M^t = M^t \cup$  Initial Unresolved Models (Saatfenster)
end for

for all  $m \in M^t$  do
    repeat
         $\vdots$  {siehe Algorithmus 2.1, Seite 19}
    until Expansion  $m$  beendet
end for

```

Initial Model (m_j^{t-1})

```

 $P^t =$  alle Kantenelemente mit Label  $\omega_i = m_j^{t-1}$ 
 $m^t =$  Approximation von  $P^t$  mit Modelltyp von  $m_j^{t-1}$ 
if Gleichung (2.1) für  $m^t$  erfüllt then
    return  $\{m^t\}$ 
else
    for Modelltyp  $\in$  {Geradenstück, Kreis- Ellipsenbogen} do
         $m^t =$  Approximation von  $P^t$  mit Modelltyp
        if Gleichung (2.1) für  $m^t$  erfüllt then
            return  $\{m^t\}$ 
        end if
    end for
    for all  $p_i \in P^t$  do
         $\omega_i = l_{unresolved}$ 
    end for
end if
return  $\{\}$ 

```

Initial Unresolved Models (*Saatfenster*)

```

 $m =$  Geradenapproximation mit Kantenelementen  $P$  aus
    Saatfenster und Label  $\omega = l_{unresolved}$ 
 $\vdots$  {siehe Initial Modells, Algorithmus 2.1}

```

Algorithmus 6.1: Model Recovery zur Bestimmung korrespondierender Konturmodelle.

Originalverfahren erfolgt die Auswahl der korrekten Modelle. Hier wird die Phase der *Model Selection* des Originalverfahrens verwendet (siehe Algorithmus 2.2, Seite 21).

Mit dieser Vorgehensweise ist es nunmehr möglich korrespondierende initiale Konturmodelle in einer Bildsequenz zu ermitteln. In der dargestellten Form ist das Verfahren in der Lage nur relativ statische Konturen als korrespondierend zu identifizieren. Jedoch könnte hier die Ausweitung der Einflußbereiche für die Bestimmung der Labelmengen eine mögliche Erweiterung sein, um auch bewegte Konturen verfolgen zu können. Ein anderer Ansatz zur Verfolgung bewegter Konturen bestünde in der Möglichkeit, ein Verfahren aus dem Bereich der Regionensegmentierung einzusetzen, das korrespondierende Regionen über die Zeit identifizieren und verfolgen kann (siehe beispielsweise [Bom97]). Mit dem Verfahren zur Zuordnung von Konturen und Regionen aus Kapitel 5 wäre es dann möglich, die Labelmengen der Kantenelemente entsprechend der Konturzuordnungen und den aktuellen Positionen der zugeordneten Regionen zu bestimmen.

Eine initiale Implementation dieses Konzeptes zur Kontursegmentierung in Bildfolgen wurde im Rahmen dieser Arbeit realisiert. Die damit erzielten Ergebnisse sind derzeit jedoch noch präliminar, so daß hier keine Evaluierung derselben stattfindet. Sie lassen allerdings bereits im aktuellen Zustand auf die Tragfähigkeit dieser Konzeption schließen, und versprechen neben der Identifikation korrespondierender Konturmodelle, eine Reduzierung der Berechnungszeiten für die Approximation initialer Konturmodelle in einer Größenordnung von ca. 75% (vergleiche [Cor00]). Im folgenden Kapitel werden nun die Ergebnisse der in den vorangegangenen Kapiteln vorgestellten Gruppierungsprozesse an Beispielbildern diskutiert.

7. Ergebnisse und Diskussion

In diesem Kapitel erfolgt die Präsentation und Diskussion von Ergebnissen, die sich bei der Anwendung der vorgestellten Techniken zum hierarchischen konturbasierten perzeptiven Gruppieren und der Kombination von Konturen und Regionen ergeben. Die Evaluation beider Verfahren findet gemeinsam statt, da beide Techniken in ihrer Anwendung eng miteinander verknüpft sind. Für die Generierung der Ergebnisse werden dieselben zehn Bilder herangezogen, die bereits bei der Diskussion der Ergebnisse der Kontursegmentierung und Regionensegmentierung verwendet wurden (vergleiche Anhang A und Abschnitt 2.1.4, Seite 22). Die Bilder 11 und 12 dienen der zusätzlichen Evaluation der Gruppierungen linear angeordneter geschlossener Figuren und fungieren darüber hinaus als Beispiel für einen rekursiv angelegten Gruppierungsprozeß.

Bereits zu Beginn des Kapitels 4 wurde dargestellt, daß mit dem konturbasierten perzeptiven Gruppieren das Ziel verfolgt wird, die Fragmentierung der initialen Konturmodelle aufzulösen und so möglichst eine vollständige Beschreibung der Objektstrukturen und auch darüber hinaus der bildinhärenten Strukturen zu erhalten. Dabei gilt es zum einen, die lokalen Fragmentierungen aufgrund lokaler Bildstörungen aufzulösen und mehr noch, die Projektionen von Umrissen verdeckter Objekte ähnlich einem menschlichen Beobachter zu rekonstruieren. Mit dieser Zielsetzung ergibt sich auch die generelle Problematik einer quantitativen Evaluation der erzielten Ergebnisse zum perzeptiven Gruppieren. Selbst ein unbedarfter menschlicher Beobachter, der die Objekte in der abgebildeten Szene nicht kennt, bringt ein gewisses Erfahrungswissen für die Strukturierung der Bildinhalte mit. Aufgrund der hohen Variabilität der initialen Kontursegmente kann es hingegen durch das System zu einer hohen Anzahl von Gruppierungshypothesen kommen, die unter einer lokalen Betrachtung zwar größtenteils eine korrekte Interpretation vornehmen, aber in einem globaleren Kontext zu Mehrdeutigkeiten oder Fehlinterpretationen führen. Die automatische Bewertung dieser Hypothesen durch das System strebt eine Reduzierung der Mehrdeutigkeiten und Fehlinterpretationen an, und verfolgt letztendlich eine Fokussierung auf die durch die Daten unterstützten und unter einer globaleren Wechselwirkung signifikanten Gruppierungshypothesen. Diese Aspekte in einer quantitativen Evaluierung zu berücksichtigen erfordert eine umfangreiche handmarkierte Stichprobe, in der die verschiedenen Variabilitäten (beispielsweise verrauschte Lokalisierung) der Eingangsdaten in repräsentativer Anzahl vorhanden sind, und entsprechend von einem menschlichen Beobachter berücksichtigt und akzeptiert werden. Neben dem sehr hohen Aufwand, der mit der Aufnahme einer solchen Stichprobe verbunden ist, ist es extrem schwierig den Versuchspersonen eine genaue Zielsetzung für

diese Handmarkierung vorzugeben. Dabei stellt sich die Frage, was ist überhaupt eine korrekte Struktur, wenn kein Domänenwissen berücksichtigt wird? Aus diesem Grund erfolgt die Beurteilung der erzielten Ergebnisse eher in einer qualitativen Form, wobei einzelne Phänomene der Gruppierungsleistung diskutiert und individuell anhand der abgebildeten Objekte beurteilt werden. Darüber hinaus werden quantitative Ergebnisse bezüglich der Reduzierung der Mehrdeutigkeiten der konturbasierten Gruppierungen und des Laufzeitverhaltens gegeben. Eine ausführliche Diskussion über die verschiedenen Schwierigkeiten einer Evaluierung von Gruppierungsansätzen, die dann auch einen Vergleich unterschiedlicher Verfahren erlauben, wird in [Bor97] geführt. Für die Kombination von Konturgruppen und Regionen erfolgt eine quantitative Auswertung auf Basis der handmarkierten Stichprobe (vergleiche Abschnitt 5.2.3, Seite 104).

Im nächsten Abschnitt werden die generierten Gruppierungen dargestellt und diskutiert. Daraufhin erfolgt die Diskussion der Bewertung der konturbasierten Gruppierungen. Abschließend werden Ergebnisse für die integrierte Objekterkennung gezeigt.

7.1. Generierung von Gruppierungshypothesen

Die Generierung der konturbasierten Gruppierungshypothesen für die Beispielbilder erfolgt unter Einsatz der Parameter, die bereits in den jeweiligen Abschnitten vorgestellt wurden (vergleiche Tabelle 4.1, Seite 63 und Tabelle 4.2, Seite 67). Nach der Diskussion der Gruppierungen mit linearem Charakter (**1D**-Ebene, **2×1D**-Ebene) werden die Ergebnisse einer Kombination der linearen Gruppen mit Regionen vorgestellt. Diese Ergebnisse beeinflussen wiederum die Generierung von geschlossenen Figuren (**2D**-Ebene). Abschließend wird die Organisation von geschlossenen Figuren, sowohl konturbasierter Natur als auch regionenbasierter Form, in lineare Strukturen zusammen mit den dafür verwendeten Parametern vorgestellt (**N×2D**-Ebene). In Tabelle 7.1 ist die Anzahl der jeweils generierten Hypothesen für jedes der Bilder angegeben.

Lineare Gruppen

Der Tabelle 7.1 ist zu entnehmen, daß trotz der zum Teil hohen Anzahl initialer Geradenstücke und Ellipsenbögen¹, zumeist eine vergleichsweise niedrige Anzahl von kollinearen und kurvilinearen Gruppierungen generiert wird. Dies unterstreicht, daß bereits die initiale Konturapproximation recht gute Ergebnisse bereitstellt, in denen die sichtbaren Konturen von Objekten relativ vollständig approximiert werden (vergleiche Anhang A). Dennoch kann es durch lokale Bildstörungen zu Fragmentierung von Objektumrissen kommen, wie sie beispielsweise in Abbildung A.5 an der Oberkante der linken Leiste und auf der rechten Leiste zu sehen sind. Auch in Abbildung A.9 treten viele fragmentierte Geradenstücke im Bereich der Garageneinfahrt auf. Durch die Gruppierung in Kollinearitäten können diese sehr lokalen Fragmentierungen überwunden werden und

¹ Kreisbögen werden hier und im folgenden unter Ellipsenbögen subsumiert.

Gruppierungshypothesen											
Bilder:	1	2	3	4	5	6	7	8	9	10	
Approximation											
Geradenstücke	35	62	143	89	114	120	101	106	130	123	
Ellipsenbögen	46	65	127	84	146	89	99	56	48	34	
Gruppierung 1D, 2×1D											
Kollinearitäten	11	51	10	71	6	16	38	48	63	74	
Kurvilinearitäten	1	25	27	11	31	19	45	3	2	-	
Proximitäten	85	244	281	368	314	208	321	265	323	333	
Parallelitäten	85	196	73	199	42	78	71	91	224	244	
Kombination											
Umriß	75	64	179	113	175	118	125	127	-	-	
Struktur	4	46	98	74	89	105	44	33	-	-	
Gruppierung 2D											
Geschlossenheiten	53	39	44	95	53	33	52	35	82	103	
regionenbasierte Proximitäten	21	37	73	47	60	67	67	66	-	-	
regionenbasierte Geschlossenheiten	9	97	27	111	14	22	22	55	-	-	
Gruppierung N×2D											
kollineare Geschlossenheiten	21	-	2	7	6	9	2	9	-	-	
kurvilineare Geschlossenheiten	5	3	5	18	2	1	4	2	-	-	
Summe:	372	819	812	1100	788	662	822	736	876	911	
Zu Bewertende:	240	535	451	660	406	377	428	394	549	578	

Tabelle 7.1.: Anzahl generierter Gruppierungshypothesen. In der Summe sind die Gruppierungen aus der Kombination von Konturen und Regionen nicht enthalten. Mit regionbasierten Proximitäten werden Relationen zwischen linearen Gruppen bezeichnet, die auf Basis einer Regionengrenze generiert wurden (vergleiche Abschnitt 5.3.3). Regionenbasierte Geschlossenheiten sind durch die Nutzung dieser Proximitäten entstanden. In der letzten Zeile ist die Anzahl der Hypothesen aufgeführt, die einer Bewertung mit dem Verfahren aus Abschnitt 6.2 unterzogen werden (vergleiche mit Tabelle 7.6). Hierzu gehören: Geradenstücke, Ellipsenbögen, Kol- und Kurvilinearitäten, Parallelitäten, Geschlossenheiten und regionbasierte Geschlossenheiten.

führen so zu einer Rekonstruktion der realen Kontur, wie in Abbildung B.5(a) und Abbildung B.9(a) zu sehen ist. Allerdings kann es hier zu vielen Mehrdeutigkeiten kommen, da viele parallel verlaufende Teilstücke vorliegen (Abbildung B.9(a)). Diese Mehrdeutigkeiten sollen jedoch durch die nachgelagerte Bewertung in einem globalen Kontext aufgelöst werden. Weiterhin Anlaß für eine Fragmentierung ergibt sich aus fehlendem Kontrast der Objektfläche zum Hintergrund. Dies ist sehr gut in dem Bild 6 im Bereich der zwei verschraubten Leisten zu erkennen. In Abbildung 7.1 ist der Grauwertauszug des Bildes gezeigt, der den fehlenden Kontrast veranschaulicht. Daher können

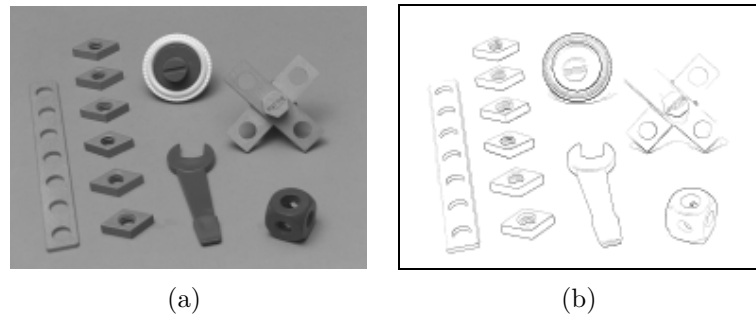


Abbildung 7.1.: (a) Grauwertauszug des Bildes 6. (b) Verdünntes Gradientenbetragsbild, das zur besseren Visualisierung logarithmisch skaliert und invertiert dargestellt ist. So entsprechen dunkle Stellen höheren Gradientenbeträgen.

hier nur kollineare Gruppen generiert werden, soweit Fragmente den Umriss der einzelnen Leisten beschreiben (siehe Abbildung B.6(a)). Für die Oberkante der Leiste mit dem Würfel in Bild 2 ergibt sich eine Situation, die neben dem geringen Kontrast im unteren Bereich der Leiste, auch durch ihre Kreuzung mit einem anderen Objekt zu einer Fragmentierung der Kante führt. Diese kann ebenfalls durch eine kollineare Gruppierung der Teilstücke aufgelöst werden (Abbildung B.2(a)). In den meisten Fällen kommt es zur Unterbrechung der initialen Konturmodelle, wenn eine lokale Nachbarschaft von verschiedenen Objektumrissen im Bild, oder kompletten Verdeckungen einzelner Objekte durch andere vorliegt. In Bild 8 sind Geradenstücke unterbrochen, die die Umrisse der Leisten für die vorderen Tragflächen bilden. Diese Unterbrechung ist durch die Nachbarschaft von diversen Schraubköpfen bedingt und kann durch kollineare Gruppierungen überwunden werden. Auch die Fragmentierung von Objektumrissen, die sich aus einer Verdeckung anderer Objekte ergibt, kann in den meisten Fällen aufgelöst werden: In den Abbildungen B.4(a), B.5(a), B.7(a) ist zu sehen, daß entsprechend unterbrochene Kanten der überdeckten Leisten durch Kollinearitäten gut rekonstruiert werden können. In Bild 10 kommt es ebenfalls an der Garageneinfahrt zu einer Verdeckung der unteren Kante durch den Modell-Lkw, die zudem durch lokale Bildstörungen stark fragmentiert ist. Auch hier wird eine große Anzahl von kollinearen Gruppierungen generiert, da im Bereich des Fensters zusätzlich mehrere kleine Geradenstücke approximiert werden. Daneben erfolgt auch eine Gruppierung von Geradenstücken unterschiedlicher Objekte und trägt zu einer Strukturierung des Bildes bei. In Bild 1 werden so zum Beispiel verbindende kollineare Gruppen zwischen Geradenstücken generiert, die die Ober- und Unterkanten der in einer Flucht liegenden Leisten beschreiben (Abbildung B.1(a)). Auch in Abbildung B.4(a) sind diese Effekte für die in Bild 4 linear angeordneten Rauten zu beobachten.

Für die Gruppierung von Ellipsenbögen in Kurvilinearitäten lassen sich ähnliche Phänomene beobachten. So werden im Bild 2 kurvilineare Gruppen für Ellipsenbögen generiert, die die Kontur der Löcher auf der langen Leiste beschreiben und aufgrund

lokaler Bildstörungen, beziehungsweise fehlendem Kontrast zum Untergrund, fragmentiert sind (Abbildung B.2(a)). Auch für den Reifen und den äußeren Umriß des blauen Würfels werden in diesem Bild kurvilineare Gruppen gebildet. Doch bereits bei dem Reifen wird eine entscheidende Problematik für diesen Gruppierungstyp offensichtlich: Der relativ große Toleranzbereich in der Orientierungsdifferenz für die Tangenten, die in den Endpunkten bestimmt und zur Gruppierung der Ellipsenbögen verwendet werden (vergleiche Abschnitt 4.2.2, Seite 61), führt zu Gruppen von strukturbeschreibenden Ellipsenbögen auf dem Profil des Reifens. Es entsteht eine große Anzahl vernachlässigbarer Gruppierungen, die häufig nicht die realen Konturen widerspiegeln. Dies verstärkt sich zudem bei einer Frontalansicht des Reifens, wie in Bild 6 und 7. Der Großteil der generierten Kurvilinearitäten in den Abbildungen B.6(a) und B.7(a) beruht auf diesem Phänomen. Hier könnte möglicherweise die Berücksichtigung des Approximationsfehlers der zugrundeliegenden Kantenelemente als eine weitere einschränkende Bedingung in den Gruppierungsprozeß eingebracht werden. Bisher ist dies jedoch nicht der Fall. Ungeachtet dessen werden aber auch Gruppierungen generiert, die eine entsprechende Rekonstruktion der wahren Kontur vornehmen. In Bild 3 wird so beispielsweise die Oberfläche des grünen Würfels durch entsprechende kurvilineare Gruppen beschrieben (Abbildung B.2(a)). Auch die Umrisse einiger Schraubköpfe in den Bildern 5 und 8, und des Würfels in Bild 5 werden durch kurvilineare Gruppen angenähert (vergleiche Abbildungen B.5(a), B.8(a)). Allerdings ist die kurvilineare Gruppierung von Ellipsenbögen insgesamt weniger erfolgreich, als die Gruppierung kollinearere Geradenstücke.

Eine anschauliche Diskussion der generierten Proximitätsgruppen verschließt sich aufgrund der großen Anzahl der generierten Hypothesen (vergleiche Tabelle 7.1). Da die wesentliche Rolle dieser Gruppierungen in dem strukturbildenden Charakter für die Generierung geschlossener, konturbasierter Figuren liegt, wird in Zusammenhang mit deren Diskussion nochmal auf einige Aspekte dieses Gruppierungstyps eingegangen.

Parallele Gruppierungen werden in den Beispielen hauptsächlich für die Kanten der Leisten und Rauten generiert (Abbildungen B.6(b), B.8(b)). Dies bedingt sich wesentlich durch die gewählte Domäne, in der diese Objekte die häufigsten parallelen Strukturen enthalten. Einige Parallelitäten werden zudem für parallele Strukturen auf den Schraubköpfen und den Schraubgewinden detektiert. Aber auch bei den Bildern 9 und 10 ist die Detektion von parallelen Gruppen zu beobachten (Abbildungen B.9(b), B.10(b)). Bei der Generierung von Parallelitäten kommt es prozessbedingt zu mehrdeutigen Hypothesen, die sowohl zwei Kollinearitäten gruppieren, als auch deren konstituierenden Teilstücke als parallel hypothetisieren. Diese Vielfalt soll durch den Bewertungsprozeß reduziert werden, und die wesentlichen parallelen Strukturen sollen als signifikant markiert werden.

Kombination linearer Gruppen und Regionen

Auf Basis der linearen Gruppen und der Regionen, deren Ergebnisse bereits in Abschnitt 2.2.3 dargelegt wurden, werden im folgenden die Ergebnisse der Kombination

Kombination Kontur - Region			
Polynomgrad	4	5	6
Zuordnung			
Umriß	97.65 %	97.91 %	96.74 %
Struktur	80.08 %	83.61 %	85.24 %
Gesamt	88.71 %	90.63 %	90.89 %

Tabelle 7.2.: Klassifikationsergebnis für die Kombination von Konturen und Regionen mit einem *leave-one-out* Test auf der Trainingsstichprobe in Prozent. Die Stichprobe umfaßt insgesamt 1559 Zuordnungen, davon 766 umrissbeschreibende und 793 strukturbeschreibende Zuordnungen. Die Prozente sind in Abhängigkeit vom Polynomgrad angegeben.

von Konturgruppen und Regionen diskutiert. Neben der exemplarischen Diskussion der Ergebnisse für einige Beispielbilder, kann hier zunächst eine quantitative Evaluierung auf Basis der Trainingsstichprobe für den entsprechenden Polynomklassifikator gegeben werden (vergleiche Abschnitt 5.2.3). So wurde mit einer Auslassungsmethode (*leave-one-out*-Verfahren) der Klassifikationsfehler in Abhängigkeit des Polynomgrades bestimmt. In Tabelle 7.2 sind die ermittelten Prozente einer korrekten Klassifikation angegeben. Hierbei steht *Umriß* für eine umrissbeschreibende Zuordnung einer Konturgruppe zu einer Region, während *Struktur* die Zuordnung einer Konturgruppe bezeichnet, die potentiell eine Struktur innerhalb einer Region approximiert. Mit dieser Auswertung wurde dann für den eingesetzten Polynomklassifikator ein Polynom sechsten Grades gewählt.

Die Anzahl der Zuordnungen, die auf den jeweiligen Beispielbildern ermittelt werden, ist in Tabelle 7.1 aufgeführt. In den Abbildungen B.1 – B.8 (c),(d) (Seite 202ff) sind die Zuordnungen für die ersten acht Beispielbilder gezeigt. Insgesamt ist zu erkennen, daß in fast allen Fällen die Zuordnungen korrekt erfolgen. In Abbildung B.1(c) werden beispielsweise alle umrissbeschreibenden Konturen für den Leistenkörper und die Lochregionen korrekt den entsprechenden Regionen zugewiesen. Ähnliches läßt sich auch für die Leistenkörper in den anderen Bildern beobachten. In Abbildung B.4(d) ist zu erkennen, daß die kollinearen Gruppierungen für die vorderen Oberkanten der gekreuzten Leisten in korrekter Weise als Struktur der verschmolzenen Holzregion zugeordnet wird. Das Fehlen der hinteren kollinearen Gruppierungen als Strukturbeschreibung für die Oberkanten der Leisten ist auf eine einschränkende Bedingung zurückzuführen. So müssen Konturgruppen für eine Strukturbeschreibung zu mindestens 50% ihrer Länge innerhalb der Regionengrenze verlaufen (vergleiche Abschnitt 5.2.3, Seite 107), was für die oberen kollinearen Gruppen nicht der Fall ist. Damit begründet sich auch, warum in dem Bild 8 in Abbildung B.8(d) teilweise die langen kollinearen Gruppen für die Leistenkanten der Tragflächen und für den Rumpf des Flugzeuges nicht zugeordnet werden. Für die Rauten in Abbildung B.6 kann hingegen eine gute Diskriminierung in umriss- und strukturbeschreibende Konturen realisiert werden. Gleiches gilt für die Struktur der Würfel und der Schraubgewinde in den Abbildungen B.4(d) und B.5(d). In den Bildern 7

und 8 mit komplexeren Baugruppen wird, bis auf die fehlenden langen Kollinearitäten als Strukturen, ebenfalls eine gute Unterscheidung in Struktur- und Umrißapproximation vorgenommen (siehe Abbildungen B.7 und B.8). Insgesamt kann also resümiert werden, daß mit dem gewählten Ansatz eines Klassifikators das Ziel, Konturgruppen und Regionen einander zuzuordnen, wenn sie das selbe Ereignis im Bild beschreiben, in fast allen Fällen erreicht wird.

Geschlossene Figuren

Die Gruppierung von Konturen in geschlossene Figuren stellt gemäß der Gruppierungshierarchie ein hohes Abstraktionsniveau der Bildstrukturen dar. Zumeist werden mit diesem Gruppierungstyp Oberflächenprojektionen von den Objekten in der Szene beschrieben. Im folgenden werden zunächst nur Geschlossenheiten betrachtet, die ohne die Verwendung von regionenbasierten Proximitäten generiert werden. Diese sind jeweils in den Abbildungen B.1(e)–B.8(e) und in den Abbildungen B.9(c) und B.10(c) für die Bilder 9 und 10 dargestellt.

In diesen Beispielbildern können die Projektionen für Leistenkörper, Rauten und für die Löcher in den Leistenkörpern überwiegend durch entsprechende Geschlossenheitshypothesen gut rekonstruiert werden. Wie der Tabelle 7.1 zu entnehmen ist, kommt es jedoch teilweise zu sehr vielen Hypothesen im Vergleich zu der Anzahl der abgebildeten Objekte. Dies ist beispielsweise für Bild 1 in Abbildung B.1(e) der Fall. Hier werden für die beiden in der Mitte liegenden Leisten aus den verschiedenen Kombinationen der einzelnen Geradenstücke und Kollinearitäten insgesamt 16 Geschlossenheiten gebildet. Davon beschreiben jeweils vier die untere und obere Leistenprojektion und nochmal acht Hypothesen die beiden Leistenprojektionen zusammengefaßt als *eine* geschlossene Figur. Die jeweils vier Hypothesen ergeben sich dabei aus der Kombination der parallel verlaufenden Kanten der Leisten, wohingegen die acht weiteren Hypothesen aufgrund von Kombinationen mit den verbindenden Kollinearitäten für diese Kanten entstehen (vergleiche Abbildung B.1(a)). Ähnliche Phänomene sind in den übrigen Bildern zu beobachten. In Abbildung B.4(e) werden so 46 unterschiedliche Geschlossenheitshypothesen für die linear angeordneten Rauten gebildet, die alle zwei oder mehr Rautenoberflächen in einer geschlossenen Figur zusammenfassen. Zusätzlich werden jeweils die einzelnen Rautenoberflächen noch als Geschlossenheiten hypothetisiert. Auch im Bild 9 werden ca. 50 Geschlossenheitshypothesen für die Vereinigung der Garagenfront mit dem Seitengiebel erzeugt (siehe Abbildung B.9(c)). Dies begründet sich aus der hohen Anzahl von über 20 parallel verlaufenden Kollinearitätshypothesen am unteren Rand der Garageneinfahrt und weiteren 2 Kollinearitäten an der vorderen Kante des Daches. Aber auch die einzelnen Strukturen aus Forderfront, Garageneinfahrt und Seitengiebel sowie das Dach werden detektiert. Darüber hinaus werden die Aufbauten der Lkw's sowohl in ihren einzelnen Flächen, als auch in ihrer Gesamtheit detektiert. In Bild 7 kann beispielsweise auch sehr gut der Leistenkörper für das Trittbrett des Rollers rekonstruiert werden. Und beim Flugzeug, in Bild 8, werden ebenfalls Hypothesen über die verwen-

deten Leistenkörper generiert, wenn auch die vordere Tragfläche zunächst fehlt. Zudem werden Hypothesen generiert, die die Leiste der vorderen Tragfläche mit dem Mittelstück verbinden und das Mittelstück selber beschreiben. Die Reduzierung dieser Mehrfachinterpretationen der selben Strukturen im Bild, die praktisch in jedem Bild mehr oder weniger häufig auftreten, soll durch den Bewertungsprozeß erreicht werden. Jedoch ist aufgrund der zum Teil hohen Anzahl der konkurrierenden Hypothesen nicht immer eine eindeutige Auflösung zu erwarten.

Die Projektionen der Würfel können kaum durch entsprechende geschlossene Figuren rekonstruiert werden. Lediglich in Bild 5 in Abbildung [B.5\(e\)](#) wird der Würfel 4 mal in seinem Umriß als geschlossen hypothetisiert. Dabei werden Kurvilinearitäten und einzelne Geradenstücke am unteren Rand der Projektion verwendet. In Bild 4 wird der Würfel hingegen mehrfach in Verbindung mit der angeschraubten Leiste und dem teilweise sichtbaren Schraubkopf als eine zusammengefaßte geschlossene Struktur hypothetisiert (Abbildung [B.4\(e\)](#)). Die innere Oberflächenstruktur der Würfel kann in Bild 3 durch fast geschlossene Kurvilinearitäten gut beschrieben werden (Abbildung [B.3\(e\)](#)). Das Fehlen der übrigen Würfelstrukturen ist in erster Linie durch überlappende Approximationen von Ellipsenbögen begründet, die identische Kantenelemente auf dem Würfelumriß beschreiben (siehe beispielsweise Abbildung [A.6](#)). Daher können in ihren Endpunkten keine Proximitätsbeziehungen generiert werden. In Bild 3 kann aus diesem Grund auch für die linke verschraubte Raute keine Geschlossenheit hypothetisiert werden. Zur Reduzierung dieser Fälle müßte die Generierung von initialen Konturmodellen in ihrer Akzeptanz überlappender Modelle stärker eingeschränkt werden, was wiederum zu stärker fragmentierten initialen Beschreibungen führt. Dem gegenüber werden für die kreisrunden Objekte wie Reifen, Felgen und Distanzringe in den überwiegenden Fällen fast geschlossene Ellipsen zur Beschreibung einer Geschlossenheit für diese Objekte verwendet (vergleiche Abbildungen [B.2\(e\)](#), [B.4\(e\)](#), [B.6\(e\)](#) und [B.7\(e\)](#)). Häufig sind aber generell zu große Lücken in den umrissbeschreibenden Konturen vorhanden, so daß in dem Proximitätsgraph keine Nähebeziehungen zwischen diesen Konturen enthalten sind, und damit keine Zyklen gefunden werden können. Dies trifft auch sehr oft bei den Rauten zu. So kann der komplette Umriß der Rauten mit dem rein konturbasierten Ansatz in keinem der Bilder gefunden werden.

Zur Verbesserung der Detektion von geschlossenen Figuren in genau diesen Fällen wird nun die Berücksichtigung von regionenbasierten Proximitätsbeziehungen diskutiert. Die Geschlossenheitshypothesen, die eine regionenbasierte Proximität verwenden, sind in den Abbildungen [B.1\(f\)](#)–[B.8\(f\)](#) gezeigt. Für die Bilder 9 und 10 wird hier keine Berücksichtigung von Regioneninformation evaluiert. In Abbildung [A.6](#) für Bild 6 sind prototypische Situationen fehlender Proximitäten sehr gut ersichtlich. So entstehen an den Begrenzungen, der zu einem Kreuz verschraubten Leisten, sehr große Lücken zu benachbarten Umrisskonturen. Diese sind auf den mangelnden Kontrast im Grauwertkanal dieses Bildes zurückzuführen (vergleiche Abbildung [7.1](#)). Sie können von dem Gruppierungsprozeß aufgrund der eingeschränkten Größe des Suchbereiches für Proximitätsbeziehungen nicht überbrückt werden. Werden regionenbasierte Proximitäten gemäß dem

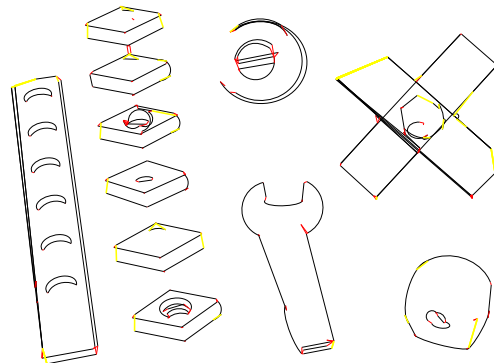


Abbildung 7.2.: Proximitätsgraph für das Bild 6. Regionenbasierte Proximitäten sind gelb markiert, während die Proximitäten aus den Bereichen perzeptiver Aufmerksamkeit rot dargestellt sind. Die linearen Gruppen sind schwarz gezeichnet. Lineare Gruppen, die in einem Endpunkt (Knoten) keine Verbindung zu anderen linearen Gruppen besitzen, sind nicht eingezeichnet (vergleiche auch Abschnitt 4.3, Seite 66ff).

Verfahren aus Abschnitt 5.3.3 zusätzlich in den Proximitätsgraphen eingefügt, so können fehlende Geschlossenheiten in diesem Graphen detektiert werden. In Abbildung 7.2 ist der erweiterte Proximitätsgraph für das Bild 6 dargestellt. Es ist deutlich ersichtlich, daß wichtige, zuvor fehlende Verbindungen eingefügt werden. Mit dieser Erweiterung lassen sich dann zusätzlich Geschlossenheitshypothesen für die Leiste des Kreuzes bilden. Auch der Leistenkörper auf der linken Seite wird detektiert und durch eine geschlossene Figur beschrieben (siehe Abbildung B.6(f) versus B.6(e)). Zudem werden auch für vier der sechs Rauten Geschlossenheitshypothesen aufgestellt, die den gesamten Umriß ihrer Projektion beschreiben. In den übrigen Beispielbildern sind ähnliche Verbesserungen des Generierungsprozesses zu beobachten. So kann beispielsweise auch die vordere Leiste der Tragflächen in Bild 8 hypothetisiert werden, die ebenfalls aufgrund fehlenden Kontrastes im unteren Bereich bisher nicht detektiert wurde (Abbildung B.8(f)). In den Bildern 2 und 3 werden ebenfalls, neben zusätzlichen Hypothesen für die Leistenkörper, Teilstrukturen der Würfel detektiert (Abbildungen B.2(f), B.3(f)). Die Anzahl der zusätzlich generierten Hypothesen für alle Bilder ist in Tabelle 7.1 aufgeführt. Daran ist auch zu erkennen, daß erwartungsgemäß die Mehrdeutigkeiten von Geschlossenheitsinterpretationen zunehmen. Es werden wesentlich mehr Hypothesen generiert, die verschiedene Objektstrukturen verschmelzen. Die Hoffnung ist, daß diese Mehrdeutigkeiten ebenfalls durch den Bewertungsprozeß korrekt aufgelöst werden.

Trotz einiger fehlender bzw. nicht adäquat beschriebener Objektstrukturen (insbesondere sind hier noch die Schrauben zu erwähnen) zeigt sich insgesamt, daß die wesentlichen konturbasierten geschlossenen Figuren erzeugt werden. Mit Ausnahme einer Leiste

Klassifikation Ähnlichkeit - Nähe			
Polynomgrad	2	3	4
Ähnlichkeit	90.6 %	91.5 %	91.5 %
Nähe			
N1	86.9 %	89.6 %	46.5 %
N2	87.3 %	90.4 %	44.4 %
N3	87.7 %	91.1 %	43.9 %
N4	87.7 %	91.7 %	46.3 %

Tabelle 7.3.: Korrekte Klassifikationen für die Ähnlichkeitswirkung und Nähewirkung von Polygonpaaren mit einem *leave-one-out* Test auf der Trainingsstichprobe in Prozent. Die Prozente sind in Abhängigkeit vom Polynomgrad und bei der Nähewirkung für jede Normierungsdefinition angegeben (vergleiche Abschnitt 4.4.1, Seite 73).

in Bild 3 werden so für die 26 teilweise verdeckten Leistenkörper in den Bildern mindestens eine korrekte Geschlossenheit für deren Oberflächenprojektion generiert. Ähnliches gilt für die Reifen und Rauten, die bis auf wenige Ausnahmen alle korrekt beschrieben werden können. Hierbei ist insbesondere hervorzuheben, daß in dem Ansatz nur ein überaus geringes Maß an Domänenabhängigkeit durch die Bestimmung der Suchbereichsmasken eingebracht wird. Dies unterstreichen auch die Ergebnisse auf den Bildern 9 und 10, die nicht Bestandteil der *baufix*[®] Domäne sind.

Lineare Gruppen von geschlossenen Figuren

Bevor die Ergebnisse zur Generierung von Gruppen linear angeordneter Geschlossenheiten diskutiert werden können, muß zunächst auf die Wahl einiger Parameter eingegangen werden. In Abschnitt 4.4 wurde im Rahmen der Vorstellung des Verfahrens angeführt, daß bei der Nähebestimmung die Normierungsdefinition N_4 und jeweils ein Polynom dritten Grades für die Klassifikatoren für Nähe- und Ähnlichkeitswirkung von Polygonpaaren eingesetzt wird. Diese Wahl begründet sich aus einer Analyse der Klassifikationsergebnisse mittels eines *leave-one-out*-Tests auf der Trainingsstichprobe. In dieser Stichprobe sind 4445 Polygonpaare bezüglich ihrer Nähewirkung und 2992 Polygonpaare in bezug auf ihre Ähnlichkeitswirkung durch eine Versuchsperson markiert worden (vergleiche Abschnitt 4.4.3). In Tabelle 7.3 sind die Gesamtergebnisse dieses Tests für eine korrekte Klassifikation in eine der jeweiligen Klassen in Abhängigkeit vom Polynomgrad und der Normierungsdefinition angegeben. Die Auswertung dieser Werte führte zur Wahl der genannten Klassifikationsparameter, da sie das beste Klassifikationsergebnis erzielen (in Tabelle 7.3 fett gedruckt).

Neben diesen Faktoren, die die einzelnen modellierten Gruppierungsprinzipien beeinflussen, müssen die Parameter zur Bestimmung einer guten Fortsetzung spezifiziert werden. In Tabelle 7.4 ist der gesamte Parametersatz angegeben. Zunächst erfolgt mit

	Parameter	Wert
Gewichtung Gruppierungsstärke		
	κ_{sn}	0.5
Gruppierungsgraph		
	Δ_{edge}	0.25
Approximation		
	$\Delta_{w_{sn}}$	0.25
	$\Delta_{continuity}(\ominus)$	0.25
	$\Delta_{\phi}(\ominus)$	60°
	$\Delta_{accept}(\ominus)$	0.05
	$\Delta_{compatibility}(\ominus)$	0.1
	$\kappa_1(\ominus)$	1.0
	$\kappa_2(\ominus)$	0.5
	$\kappa_3(\ominus)$	3.5
	$\kappa_4(\ominus)$	12

Tabelle 7.4.: Der gewählte Parametersatz zur Ermittlung von Konturmodellen für linear angeordnete geschlossene Figuren (zur Erläuterung siehe Text).

der Wahl von $\kappa_{sn} = 0.5$ eine Gleichgewichtung der Gruppierungswirkung aufgrund von Ähnlichkeit und Nähe in einen gemeinsamen Wert, da a priori nicht offensichtlich ist, welches der beiden Gruppierungsprinzipien stärkeren Einfluß für die Gruppierung von Polygonpaaren haben soll (vergleiche Definition (4.28), Seite 89). Der Wert für Δ_{edge} zur Generierung von Kanten im Gruppierungsgraphen wird so gewählt, daß auch noch Kanten zwischen Polygonen möglich sind, die sich eher unähnlich sind ($w_s < 0.5$), aber doch nahe beieinander liegen ($w_n > 0.5$, vergleiche Definition (4.29)). Die Bestimmung der initialen Saaten aus drei im Gruppierungsgraphen benachbarten Polygonen wird durch die Werte $\Delta_{w_{sn}}$, $\Delta_{continuity}(\ominus)$ und $\Delta_{\phi}(\ominus)$ eingeschränkt (siehe Seite 92). Es sollen die Gruppierungsstärken der beteiligten Polygone stark genug, das Verhältnis der Abstände ähnlich und alle Polygone einer initialen Saat möglichst geradlinig angeordnet sein. Diese Forderungen werden mit den gewählten Werten adäquat eingeschränkt. Die Expansion dieser Saaten wird schließlich durch die weiteren Parameter $\Delta_{accept}(\ominus)$ und $\Delta_{compatibility}(\ominus)$ reglementiert (vergleiche Seite 92). Im Gegensatz zur initialen Kontursegmentierung sind diese Werte hier dimensionslos und schränken die Abweichung eines möglicherweise kompatiblen Polygons im Verhältnis zu dem mittleren internen Abstand der aktuellen Gruppe ein. Dies gilt auch für den Wert von $\Delta_{continuity}(\ominus)$. Die Wahl der Gewichte für die Selektion von Konturmodellen durch die Optimierungsfunktion kann aus der initialen Kontursegmentierung übernommen werden. Lediglich die zusätzliche Berücksichtigung der Gruppierungsstärke ist noch anzugeben, die hier sehr starken Einfluß hat, um möglichst die Konturmodelle zu wählen, die auch eine hohe Gruppierungswirkung haben. Alle Werte wurden in umfangreichen Tests auf Bildern der **baufix**[®] Domäne bestimmt.

In den Abbildungen **B.11–B.18** sind die Ergebnisse für die Approximation linear angeordneter Geschlossenheiten durch Geradenstücke und Ellipsenbögen dargestellt. Die Anzahl der jeweils erzeugten kollinearen und kurvilinearen Strukturen von geschlossenen Figuren ist in Tabelle **7.1** angegeben. Die Löcher in den Leisten können in fast allen Fällen in lineare Gruppen organisiert werden, sofern sie in ausreichender Zahl und mit ähnlicher Form vorhanden sind (siehe beispielsweise Abbildungen **B.11(a)**, **B.12(a)**, **B.15(a)**, **B.16(a)**). In Bild 1 werden so auch für die vielen mehrdeutigen Geschlossenheiten für die Leistenkörper Geradenstücke generiert, die deren lineare, parallele Anordnung beschreiben (siehe Abbildung **B.11(a)**). In Bild 2 (Abbildung **B.12(a)**) ist sehr gut zu sehen, daß eine kurvilineare Anordnung für die Löcher detektiert wird, die durchaus von einem menschlichen Beobachter nachvollzogen werden kann, wenn er nur die Löcher betrachtet. Auch in den Abbildungen **B.12(b)** und **B.16(a)** sind ähnliche Gruppierungen zu erkennen. Die Projektionen der Rautenoberflächen und deren lineare Anordnung können ebenfalls jeweils gut durch Geradenstücke und Ellipsenbögen rekonstruiert werden (vergleiche Abbildungen **B.13(a)**, **B.14(b)**, **B.14(b)**). Es sei hier nochmal explizit darauf hingewiesen, daß möglichst äquidistante und gleichförmige Figuren gruppiert werden sollen. Dies ist in Bild 3 und 4 an einigen Stellen nicht gelungen (vergleiche Abbildungen **B.12(b)** und **B.13(a)**). In beiden Bildern ist eine krasse Fehlklassifikation in der Ähnlichkeit der Objekte für deren Gruppierung verantwortlich. Ein weiterer Fall ist in Abbildung **B.14(b)** zu erkennen.

In Bild 11 wurde extra eine regelmäßige Anordnung der Objekte in der Szene hergestellt und ein konturbasiertes Gruppieren durchgeführt (wird hier nicht weiter diskutiert). In Abbildung **B.17(a)** sind die Ergebnisse für die Gruppierung der darunterliegenden Geschlossenheiten gezeigt. Die parallelen Anordnungen gleichförmiger Objekte können gut approximiert werden. Einige vertikale Gruppen beschreiben die Anordnung von Lochregionen in den Objekten. Der Kreis ist eher als eine Fehlinterpretation zu beurteilen. In den Abbildungen **B.17(b)** und **B.17(c)** wurde das Gewicht für den Einfluß von Ähnlichkeit und Nähe variiert. Bei einer Ähnlichkeit bevorzugenden Gruppierung (Abbildung **B.17(b)**, $\kappa_{sn} = 0.1$) werden sowohl die horizontal ähnlichen Objekte gruppiert, als auch die Projektionen der Rauten in vertikaler Richtung. Die verschiedenen Kreisbögen veranschaulichen, daß eine höhere Flexibilität im Prozeß vorhanden ist und nicht offensichtliche (wenn zum Teil auch nachvollziehbare) Gruppierungen gebildet werden. In Abbildung **B.17(c)** hingegen wurde die Nähewirkung stark gewichtet ($\kappa_{sn} = 0.9$). Sehr gut ist zu erkennen, daß nun nur noch vertikale Gruppen zwischen den näher beieinander liegenden Objektprojektion gebildet werden. Dieses Verhalten weist auf Parallelen zu der menschlichen Wahrnehmung hin. Denn auch der Mensch kann seine Wahrnehmung von Strukturen bewußt in die eine oder andere Richtung (Nähe, Ähnlichkeit, horizontal, vertikal) beeinflussen. Jedoch bleibt die Frage offen, welches die vorzugsweise Richtung ist, und welches die einflußnehmenden Faktoren beim Menschen sind.

In Abbildung **B.18(a)** wird das Verfahren auf die Ergebnisse einer Regionensegmentierung mit dem Island Growing angewendet. In der Szene in Bild 12 ist ein Gullydeckel abgebildet, dessen regelmäßige Strukturen es zu rekonstruieren gilt. Leider ist es jedoch

schwierig, bereits eine gute initiale Segmentierung zu erreichen, und so sind einige strukturbildende Elemente nicht adäquat beschrieben (vergleiche Abschnitt 2.2.3). Trotzdem kann mit den *gleichen* Parametereinstellungen ein sehr gutes Ergebnis erzielt werden. So werden die Löcher in mehrere aneinandergrenzende Kreisbögen gruppiert, und auch auf dem äußeren Ring erfolgt teilweise eine gute Rekonstruktion der kurvilinear angeordneten Steine. Die Begründung für Fragmentierungen ist zum einen in einer eingeschränkten Ähnlichkeit der nächsten benachbarten Objektprojektionen zu suchen, und zum anderen durch die im folgenden diskutierte Problematik bei der Selektion von Modellen gegeben.

In den Abbildungen B.11(b), B.13(b), B.15(b), B.16(b) und B.18(b) sind die initialen Ergebnisse eines rekursiven Gruppierens entsprechend dem Verfahren aus Abschnitt 4.4.5 abgebildet. In den hier verwendeten Bildern werden zunächst die Fragmentierungen der Konturmodelle aus der $\mathbf{N} \times \mathbf{2D}$ -Ebene in der $\mathbf{1D}$ -Ebene aufgelöst. Diese Fragmentierung bedingt sich durch die Selektion gruppierender Konturmodelle, in der Modelle, die einzelne Geschlossenheitsstrukturen mehrfach approximieren, zum Teil zurückgewiesen werden. In der $\mathbf{1D}$ -Ebene werden jedoch Gruppen erzeugt, die die zugrundeliegenden Strukturen vollständig rekonstruieren. Für einige Bilder werden zudem Parallelitätshypothesen aufgestellt (nicht abgebildet). Allerdings werden keine höheren Gruppierungen generiert, da keine entsprechenden Strukturen vorhanden sind. Für das Bild 11 kann jedoch bei der Wahl von $\kappa_{sn} = 0.1$ eine Geschlossenheit für die gesamte Anordnung der Objekte detektiert werden (siehe Abbildung B.17(d)). Zwei weitere Geschlossenheiten entstehen aus günstigen Anordnungen der Konturmodelle. Sie beschreiben allerdings keine wesentliche Struktur im Bild.

Insgesamt kann aufgrund der erzielten Ergebnisse geschlossen werden, daß mit dem vorgestellten Verfahren ein wichtiger Schritt in die Richtung einer automatischen Gruppierung von ähnlichen, linear angeordneten, geschlossenen Figuren realisiert wird. Dies wird auch durch die Anwendung in dem Bild 12 unterstützt.

7.2. Bewertung von Gruppierungshypothesen

In diesem Abschnitt wird nun die Bewertung der generierten Gruppierungshypothesen für die Beispielfiguren vorgestellt und diskutiert. Dabei werden jedoch die Proximitätsgruppierungen, deren Charakter mehr in einer Strukturbildung liegt, und die Gruppierungen der $\mathbf{N} \times \mathbf{2D}$ -Ebene für eine Bewertung nicht berücksichtigt. Letztere sind im Rahmen der vorliegenden Arbeit noch nicht im Bewertungssystem integriert worden.

Zur Diskussion der Ergebnisse von Bewertungen sind zunächst einige Parameter aus Abschnitt 6.2 zu spezifizieren. Beim Aufbau des Gruppierungsgraphen wird zur Detektion von konkurrierenden Beziehungen zwischen initialen Konturmodellen ein Mindestprozentsatz $\Delta_{overlap}(\cdot)$ für eine gegenseitige Überdeckung auf Basis der approximierten Kantenelemente gefordert (vergleiche Definition (6.12), Seite 137). Darüber hinaus müssen für die Bestimmung der datengetriebenen Signifikanz für Kol- und Kurvilinearitäten und für Parallelitäten Gewichte $(\alpha_{1,2}, \beta_{1,2,3})$ für die Linearkombination der Einzelfaktoren

	Parameter	Wert
Segmentüberdeckung		
	$\Delta_{overlap}(\cdot)$	0.6
Kol-/Kurvilinearitäten		
	$\alpha_{1,2}$	0.5
Parallelitäten		
	$\beta_{1,3}$	0.2
	β_2	0.6

Tabelle 7.5.: Parametersatz zur Bewertung von Gruppierungshypothesen (siehe Text).

angegeben werden (vergleiche Definitionen (6.22) und (6.27), Seite 145ff). Eine empirische Ableitung dieser Werte, etwa aus der Stichprobe für die perzeptiven Aufmerksamkeitsbereiche, ist schwer zu realisieren. Deshalb wurden die Werte in Experimenten durch Variationen und subjektive Einschätzung der erzielten Ergebnisse ermittelt. Die resultierenden Werte sind in Tabelle 7.5 angegeben.

Mit diesen Parameterwerten werden die generierten Gruppierungshypothesen der Bilder 1–10 als signifikant oder nicht signifikant für die Strukturierung des Bildinhaltes markiert. Zur Auswahl der signifikanten Gruppierungshypothesen wird hier ein Wert $\omega_{sig} = 0.8$ angenommen. Damit gelten alle Hypothesen, deren Labelwert $\omega_i \geq \omega_{sig}$ ist, als signifikante Gruppierungshypothesen, während die Übrigen als nicht signifikant aufgefaßt werden. Allerdings ist es schwer eine harte Entscheidung zu treffen, da ein Signifikanzwert von $\omega_i = 0.5$ eine eher unsichere Bewertung signalisiert. Solche Hypothesen tragen möglicherweise auch zu Strukturen im Bild bei, jedoch kann im Energieminimierungsprozeß des G-MRF's trotz verschiedener Kontextabhängigkeiten keine sichere Entscheidung getroffen werden. Die Anzahl der signifikanten Gruppierungshypothesen ($\omega_i \geq 0.8$) sind in Tabelle 7.6 aufgeführt. Die Ergebnisse der Bewertung für die einzelnen Bilder sind in Anhang C ab Seite 215 dargestellt. Eine anschauliche Diskussion der gesamten Ergebnisse ist aufgrund der hohen Anzahl signifikant bewerteter Hypothesen sehr schwer. Aus diesem Grund wird im folgenden auf verschiedene Phänomene in den einzelnen Bildern eingegangen.

In den Abbildungen C.1(a)–C.10(a) sind die als signifikant markierten linearen Gruppen dargestellt. Sie enthalten neben den signifikanten Kol- und Kurvilinearitäten auch die signifikanten initialen Konturmodelle, die zur Generierung von höheren Gruppierungen verwendet werden. Initiale Konturmodelle, die zu *keiner* Gruppierung beitragen, sind nicht dargestellt, da sie im Bewertungsprozeß nicht berücksichtigt werden. Es fällt beispielweise in der Abbildung C.1(a) auf, daß die Geradenstücke für die Schmalseiten der Leistenkörper zum Teil fehlen, obwohl sie Elemente von signifikanten Geschlossenheitshypothesen sind. Dieser Effekt veranschaulicht die komplexen Wechselwirkungen innerhalb des G-MRF's. Die oberen beiden Leisten werden 16 mal zusammen als eine geschlossene Oberflächenprojektion (Geschlossenheit) hypothetisiert. Alle Hypothesen werden jedoch als nicht signifikant markiert (vergleiche Abbildung C.1(c)). Im Ener-

Signifikante Gruppierungshypothesen										
Bilder:	1	2	3	4	5	6	7	8	9	10
Approximation										
Geradenstücke	24	39	102	37	77	97	69	66	61	81
Ellipsenbögen	26	35	70	38	85	58	47	28	10	10
Gruppierung 1D, 2×1D										
Kollinearitäten	6	22	7	27	2	11	19	26	29	34
Kurvilinearitäten	1	12	10	6	19	11	20	1	1	-
Parallelitäten	44	82	49	96	29	56	44	52	105	116
Gruppierung 2D, regionenbasiert										
Geschlossenheiten	34	22	25	50	32	22	29	21	33	38
regionenbasierte Geschlossenheiten	5	31	14	31	4	11	10	21	-	-
Summe:	140	243	277	285	248	266	238	215	239	279

Tabelle 7.6.: Anzahl signifikanter Gruppierungshypothesen (vergleiche mit Tabelle 7.1).

gieminimierungsprozeß bewirkt dies auch eine Veränderung der Signifikanzwerte für die konstituierenden Elemente. Für die jeweils *einzelnen* Leisten werden hingegen insgesamt 9 Hypothesen als signifikant bewertet, die ebenfalls diese Geradenstücke beinhalten. Der Einfluß dieser signifikanten Hypothesen auf die Geradenstücke für die Schmalseiten reicht jedoch nicht aus, um den negativen Einfluß der 16 nicht signifikanten Hypothesen aufzuheben. Umgekehrt reicht aber der Einfluß der weniger signifikanten Geradenstücke nicht aus, um die Signifikanz der Geschlossenheiten für die einzelnen Leisten zu schwächen. In den übrigen Bildern sind ähnliche Effekte zu beobachten (vergleiche beispielsweise Abbildungen C.2(a), C.4(a)). In den überwiegenden Fällen besitzen die einzelnen konstituierenden Hypothesen aber dennoch Signifikanzwerte im Bereich von $\omega_i = 0.5 \dots 0.8$, so daß hier eher eine unsichere Entscheidung aufgrund der vielen Bestandteilsbeziehungen zu den signifikanten und nicht signifikanten Geschlossenheiten signalisiert wird. Neben einigen dieser unsicheren Bewertungen kommt es jedoch zu wesentlichen Reduzierungen fehlerhafter Interpretationen. So wird beispielsweise für die oben in der Mitte des Bildes 5 positionierte Schraube eine Kurvilinearität generiert, die den Schraubkopf mit einer Gewindestruktur gruppiert (vergleiche Abbildung B.5(a)). Diese wird durch den Bewertungsprozeß als nicht signifikant markiert (siehe Abbildung C.5(a)). Eine ähnlich fehlerhafte Gruppierung des linken Würfels in Bild 3 mit einem Schraubkopf wird ebenfalls als nicht signifikant markiert. Die Eliminierung der texturbedingten Kurvilinearitäten im Bereich der Reifen ist nur eingeschränkt gelungen. Allerdings liegen hier auch sehr viele initiale Ellipsenbögen vor, die eine Annahme korrekter Strukturen im System unterstützen (siehe Abbildungen C.6(a) und C.7(a)). Insgesamt werden viele der wesentlichen linearen Strukturen in dem Bewertungsprozeß als signifikant markiert. Des weiteren kann eine deutliche Reduzierung der Anzahl der generierten linearen Gruppen erreicht werden (vergleiche Tabelle 7.1 und 7.6). Die als weniger signifikant markierten linearen Gruppen können zudem über die als signifikant markierten höheren Gruppen selektiert werden, sofern sie als konstituierende Elemente auftreten.

Die Bewertung der Parallelitätshypothesen führt in starkem Maße zu einer Reduzierung der Anzahl der generierten Hypothesen. So werden durchschnittlich fast die Hälfte aller Hypothesen als nicht signifikant markiert. Eine vollständige Auflösung der Mehrdeutigkeiten wird zwar nicht erreicht, aber alle wichtigen und strukturbildenden Hypothesen bleiben erhalten.

Mit den Geschlossenheitshypothesen wird der aktuell höchste Strukturierungsgrad beschrieben, der im G-MRF bewertet wird. Primäres Ziel ist es hier, die Geschlossenheiten als signifikant zu markieren, die die Projektionen von Objektoberflächen im Bild beschreiben. Dies kann in allen betrachteten Bildern zumeist sehr gut erreicht werden. Als weiteres Ziel wird die Reduzierung von mehrdeutigen Interpretationen angestrebt. So werden in Bild 1 alle 20 Geschlossenheitshypothesen, die die Oberflächenprojektionen der Leistenkörper in der Mitte und der Leisten darüber beschreiben, als nicht signifikant markiert. Dem gegenüber werden alle Geschlossenheitshypothesen für die einzelnen Leisten als signifikant eingestuft. In Bild 2 wird zum Beispiel eine deutliche Reduzierung der Mehrdeutigkeiten erreicht. Von den insgesamt 87 Hypothesen für die Projektion der beiden gekreuzten Leisten, werden nur 34 als signifikant markiert. Wiederum beschreiben von diesen 34 Hypothesen korrekterweise 12 die Projektion der kurzen Leiste und 10 die der langen Leiste (siehe Abbildung C.11(a)). Die übrigen 12 Hypothesen stellen Strukturen dar, die den rechten Teil der langen Leiste mit der kurzen Leiste verbinden (Abbildung C.11(b)). Damit konnte zwar keine vollständige Auflösung der Mehrdeutigkeiten erzielt werden, aber zumindest wurden diese in hohem Maße reduziert. Ähnliche Effekte sind auch in den anderen Bildern zu verzeichnen. In Bild 3 wird neben der Projektion des Leistenkörpers auch dessen Verschmelzung mit dem unteren Teil der gekreuzten Leiste als signifikant markiert (Abbildung C.3(c)). Auch in Bild 4 (Abbildung C.4(c)) werden Mehrdeutigkeiten für die gekreuzten Leisten nicht vollständig aufgelöst. Jedoch wird jede Hypothese für die einzelnen Leisten als signifikant eingestuft. Die gute Bewertung für die verbundenen Rauten ist hier zwar prinzipiell nicht gewollt, kann aber unter dem Gesichtspunkt einer Strukturierung des Bildinhalts durchaus als korrekt empfunden werden. Es könnte hier auch eine stark fragmentierte Projektion einer Leiste vorliegen. Zudem werden die Projektionen der Rautenoberflächen durchgehend als signifikant bewertet. In der Mitte dieses Bildes wird die Struktur aus Würfel und verschraubter Leiste nur als Ganzes signifikant eingestuft, obwohl auch deren Einzelprojektionen als Geschlossenheiten hypothetisiert werden. Hier wäre eine bessere Auflösung wünschenswert gewesen. In Bild 5 (Abbildung C.5(c)) werden die drei Geschlossenheiten für die Projektionen der linken Leiste nur mit einem Signifikanzwert von $\omega_i = 0.6$ markiert. Dies begründet sich durch eine ausgewogene Bewertung in der jeweiligen datengetriebenen Signifikanz, die auch im Kontext konkurrierender und unterstützender Hypothesen keine eindeutige Entscheidung unterstützt. In Bild 2 wird die generierte Teilstruktur für den Würfel hinter dem Reifen mit einer Signifikanz von $\omega_i = 0.4$ bewertet (vergleiche Abbildungen B.8(f) versus C.8(c)). Diese Struktur wird von einer regionenbasierten Geschlossenheit beschrieben, die zum Teil sehr große Lücken zwischen den konstituierenden linearen Gruppen enthält. Mit der Berücksichtigung der Lückenlänge

in der datengetriebenen Signifikanz für Geschlossenheiten erfolgt daher eine eher nicht signifikante Einstufung. In Bild 6 ist dieser Effekt unter anderem ebenfalls dafür verantwortlich, daß die regionenbasierte Geschlossenheit für die vordere der gekreuzten Leisten als nicht signifikant bewertet wird (siehe Abbildung C.6(c)). Für das Spielzeugflugzeug in Bild 8 werden die einzelnen Leisten für die Tragflächen und das Leitwerk als signifikant markiert (siehe Abbildung C.8(c)). Das Mittelstück wird zudem mit dem Propeller als eine ganze Struktur verschmolzen und als signifikant bewertet. Die Geschlossenheit für die gesamte Mittelleiste (siehe Abbildung B.8(e)) wird aufgrund der Konkurrenz zu den mehrfach generierten Strukturen inklusive Propeller als nicht signifikant eingestuft. In den Bildern 9 und 10 werden ebenfalls alle Projektionen der Objektoberflächen als einzelne signifikante Hypothesen markiert. Jedoch kommt es auch hier zu Verschmelzungen benachbarter Oberflächen, die ebenfalls als signifikant eingestuft werden. Lediglich für den Aufbau der Lkw's werden in allen Fällen dekomponierende Geschlossenheitshypothesen für die einzelnen Oberflächen als signifikant markiert.

Die globale Bewertung kann also einen Großteil der signifikanten Strukturen im Bild als solche identifizieren. Durch die Definition von konkurrierenden und unterstützenden Relationen werden viele Mehrdeutigkeiten aufgelöst, wenn auch in einigen Fällen, aufgrund der hohen Anzahl der generierten Hypothesen, keine eindeutige Auflösung erzielt werden kann. So kommt es beispielsweise für das Bild 2 zu ca. 9500 Cliques im G-MRF, von denen alleine ca. 5500 die Konkurrenz zwischen Geschlossenheiten abbilden. Eine eindeutige Auswahl ist angesichts der damit verbundenen komplexen Wechselwirkungen nicht zu erwarten. Zudem ist hier a priori auch nicht immer festzulegen, welches die korrekten Strukturen sind. Insgesamt kann geschlossen werden, daß alle wesentlichen Strukturen auch durch den Bewertungsprozeß als signifikant eingestuft werden, wobei gleichzeitig eine deutliche Reduzierung der Anzahl zu betrachtender Hypothesen erreicht wird. Auch die Geschlossenheiten, die verschiedene Projektionen von Objektoberflächen verschmelzen, beschreiben ohne ein explizit modelliertes Domänenwissen durchaus korrekte Strukturen in den diskutierten Bilddaten.

Laufzeiten

Abschließend erfolgt nun eine kurze Darstellung der benötigten Rechenzeiten der einzelnen Module. Alle Gruppierungshypothesen wurden auf einer DEC Alpha Workstation XP 1000 / 667MHz berechnet. Die Laufzeiten sind in Tabelle 7.7 für jedes der zehn Testbilder in Sekunden angegeben.

Der Tabelle ist zu entnehmen, daß die Rechenzeiten speziell für das G-MRF entscheidend von der Anzahl der Hypothesen abhängt, die in konkurrierenden Relationen zu anderen Hypothesen stehen. So sind die Rechenzeiten für die Bewertung der Bilder 2, 4, 8, 9 und 10 aufgrund einer hohen Anzahl mehrdeutiger Geschlossenheitshypothesen deutlich länger als bei den übrigen Bildern. Die Rechenzeiten für das Gruppieren linear angeordneter Geschlossenheiten liegt bei allen Bildern über dem Durchschnitt der anderen Module. Dies begründet sich zum einen aus einer aktuell ineffizienten Implementierung

Laufzeiten der Gruppierungsmodule											
Bilder:	1	2	3	4	5	6	7	8	9	10	
Kontursegmentierung											
Detektion	0.42	0.38	0.46	0.46	0.46	0.48	0.46	0.46	0.46	0.46	
Approximation	2.71	4.08	3.36	4.18	3.10	6.18	3.35	1.85	2.25	1.86	
Gruppierung 1D, 2×1D											
	0.06	0.35	0.35	0.50	0.31	0.28	0.62	0.32	0.53	0.43	
Kombination											
	0.15	0.33	0.32	0.33	0.28	0.26	0.41	0.45	-	-	
Gruppierung 2D											
	0.05	0.21	0.15	0.37	0.15	0.13	0.24	3.08	0.20	0.18	
Gruppierung N×2D											
Nachbarschaft	7.40	1.00	4.76	365.0	5.08	5.76	10.90	11.01	-	-	
Approximation	0.25	0.00	0.16	15.6	0.00	0.20	0.07	0.03	-	-	
Bewertung mit G-MRF											
Gruppierungsgraph	0.90	6.75	1.00	8.56	0.43	0.58	0.63	4.90	5.36	4.75	
Minimierung	0.45	16.98	0.38	9.18	0.15	0.17	0.73	3.15	4.55	8.45	
Summe:	12.39	30.08	10.94	404.2	9.96	14.04	17.41	25.25	13.35	16.13	

Tabelle 7.7.: Berechnungszeiten der einzelnen Gruppierungsmodule für die zehn Beispielbilder in Sekunden.

und zum anderen durch teilweise aufwendige Berechnungen der verwendeten Merkmale zur Klassifikation von Nähe- und Ähnlichkeitswirkung von Polygonpaaren. Alle Zeiten beziehen sich auf die reine Bearbeitungszeit, ohne Ein- und Ausgaben von Zwischenergebnissen.

7.3. Integrierte Objekterkennung

In Abschnitt 5.4 wurde ein Konzept vorgestellt, das eine integrierte bzw. rückgekoppelte Objekterkennung durch die Kombination von Konturen und Regionen vorschlägt. Anhand der Beispielbilder 6 und 8 werden hier initiale Ergebnisse dieser Anwendung vorgestellt.

In Abbildung D.1 sind die erzielten Ergebnisse dargestellt. Zunächst werden die Bilder, wie in Abschnitt 5.4.1 beschrieben, durch einen regionenbasierten Ansatz segmentiert und ein hybrider Objekterkenner stellt Objekthypothesen für diese Regionen auf (siehe Abbildungen D.1(a) und (b)). Kann der Objekterkenner für eine Region keine Hypothese aufstellen, so klassifiziert er diese Region als *unbekannt*. Dies ist in Abbildung D.1(a) für die gekreuzten Leisten der Fall und auch für den Reifen. In Abbildung D.1(b) wird der gesamte Rumpf inklusive Tragflächen von der Regionensegmentierung in eine Region segmentiert. Für diese Region kann der auf Einzelobjekte ausgelegte Objekterkenner erwartungsgemäß ebenfalls keine Hypothese aufstellen. Diese unbekannt Objekte werden nun dem konturbasierten Gruppierungsmodul zugeführt, welches nur in diesen

Bildbereichen eine konturbasierte Interpretation durchführt. In den Abbildungen [D.1\(c\)](#) und [D.1\(d\)](#) sind die Ergebnisse dieser fokussierten Gruppierung für die Bilder 6 und 8 dargestellt. Es ist deutlich zu erkennen, daß für die verschiedenen Objektprojektionen in diesen unbekannt klassifizierten Bereichen entsprechende Geschlossenheiten generiert werden. Diese werden nun, wie in Abschnitt [5.4.2](#) beschrieben, dem Objekterkennungssystem als alternative Regionen zugeführt und von diesem entsprechend interpretiert. Die so erzielten zusätzlichen Objekthypothesen sind in den Abbildungen [D.1\(e\)](#) und [D.1\(f\)](#) dargestellt. Mit dieser Integration können nun die beiden gekreuzten Leisten in Bild 6 identifiziert werden und auch in Bild 8 werden für die Leistenkörper der Tragflächen und des Rumpfes korrekte Hypothesen aufgestellt. Prozessbedingt kommt es auch zu zusätzlichen Hypothesen, die wiederum als unbekannt klassifiziert bzw. falsch interpretiert werden (vergleiche Abbildungen [D.1\(e\)](#) und [D.1\(f\)](#)). Jedoch liegt hier eine generelle Problematik des Objekterkennungssystems vor. So werden neben zumeist korrekten Interpretationen bereits durch den rein regionenbasierten Ansatz Fehlklassifikationen vorgenommen. In Abbildung [D.1\(a\)](#) werden für die oberen beiden Rautenmutter Schrauben und für die Felge des Reifens eine Rautenmutter hypothetisiert. Auch in Abbildung [D.1\(b\)](#) tritt dieser Effekt am Propeller auf, der zu einer roten Schraube klassifiziert wird. Diese Fehlklassifikationen könnten durch eine in Abschnitt [5.4](#) vorgeschlagene Erweiterung des Objekterkennungssystems, mit einer Berücksichtigung aller ermittelten konturbasierten Gruppierungen, vermutlich deutlich verringert werden.

Da der Fokus der vorliegenden Arbeit mehr in der Modellierung von automatischen Gruppierungsprozessen liegt, wurden im Rahmen dieser Arbeit keine umfangreichen Tests auf entsprechenden Stichproben durchgeführt, die eine aussagekräftige Evaluation über die Steigerung der Erkennungsrate des Objekterkennungssystems mit dem Integrationsansatz erlauben würden. Vielmehr sollte die Konturinformation insgesamt zunächst in das Objekterkennungssystem integriert werden, und dann eine entsprechende Evaluierung erfolgen. Die präsentierten Ergebnissen lassen jedoch eine signifikante Steigerung erwarten.

7.4. Resümee und Ausblick

Mit den erzielten Ergebnissen wird insgesamt gezeigt, daß ein Großteil der angestrebten Gruppierungshypothesen mit den entwickelten Methoden erzeugt werden können.

Im Verfahren zur Hypothesierung von linearen Gruppen werden zumeist die offensichtlichen Fragmentierungen der initialen Konturmodelle aufgelöst. Zudem wird in den meisten Fällen eine Rekonstruktion von Objektkonturen erzielt, die bei einer Projektion aufgrund von Verdeckungen nicht abgebildet wurden. Im Fall der Kurvilinearitäten könnte die Berücksichtigung eines Approximationsfehlers zur Vermeidung von offensichtlich nicht korrekten Hypothesen ein robusteres Gruppierungsverhalten realisieren. Auch bei der Skalierung der Masken bietet sich eine Erweiterung an, um die Generierung vieler mehrdeutiger Hypothesen zu vermeiden, oder fehlende Hypothesen aufgrund zu großer Lücken erzeugen zu können. So könnte die Berücksichtigung der Dichte, bezogen

auf die Anzahl der Konturmodelle innerhalb des aktuellen Suchbereich eines gruppierenden Modells, als ein zusätzlicher Faktor eine dynamische Skalierung des Suchbereichs unterstützen (vergleiche beispielsweise [Low85], Seite 40ff). Damit würde auch das rekursive Gruppierungsmodell für lineare Gruppen von geschlossenen Figuren dynamisch auf die dabei untersuchten globaleren Strukturen angepaßt.

Für die Kombination von linearen Gruppen und Regionen können bereits sehr gute Ergebnisse erzielt werden. Eine Erweiterung des Merkmalsvektors zur Klassifikation einer Zuordnung wurde bereits mit verschiedenen zusätzlichen Merkmalen untersucht. Hierbei ergaben sich jedoch nur marginale Veränderungen in der Gruppierungsleistung des Verfahrens. Aus diesem Grund bleibt hier als eine mögliche Erweiterung die Stichprobe zu vergrößern (auch um domänenfremde Segmentierungen), und damit eine generalisierte Basis für das Verfahren zu entwickeln. Dies könnte dann möglicherweise auch die zusätzliche einschränkende Bedingung für eine strukturbeschreibende Zuordnung eliminieren.

Im Verfahren zur Hypothesisierung von geschlossenen Figuren konnte bereits durch die Integration von regionenbasierten Proximitäten eine deutlich Steigerung der Generierung korrekter Gruppierungen erreicht werden. Darüber hinaus verspricht eine relaxierte Behandlung von Proximitätsrelation zwischen zwei sich teilweise überdeckenden linearen Gruppen (Überdeckung im Sinne einer gemeinsamen Approximation identischer Kantenelemente) eine Steigerung der Anzahl korrekter Geschlossenheitshypothesen. Alternativ dazu, wie bereits in der Ergebnisdiskussion angesprochen, könnte eine Anpassung der initialen Konturapproximation durchgeführt werden. Mit einer dynamisch angepaßten Skalierung der Proximitätsmasken, wie bei den linearen Gruppen angesprochen, können darüber hinaus zusätzliche korrekte Hypothesen erwartet werden, da damit größere Lücken überbrückt werden könnten.

Für die Gruppierung von linear angeordneten geschlossenen Figuren, liegt die Berücksichtigung weiterer Merkmale nahe. So läßt die Verwendung des Gestaltgesetzes der Symmetrie eine Verbesserung der Ergebnisse erwarten. In der aktuellen Konzeption berücksichtigt das Verfahren diesen Faktor nicht. Beispielsweise werden daher Leistenkörper, die in der Form eines doppelten Ts' $| - |$ angeordnet sind, aufgrund ihrer ähnlichen Form und der auf einer Geraden liegenden Schwerpunkte durch ein Geradenstück gruppiert. Zusätzlich bietet sich bei der Gruppierung linear angeordneter Regionen die Berücksichtigung von Farbe an, die ein weiteres starkes Ähnlichkeitsmerkmal darstellt (vergleiche Abschnitt 3.1.2).

Das Design des Markov Random Fields zur Bewertung der konturbasierten Gruppierungshypothesen hat sich insgesamt bewährt, wie auch die erzielten Ergebnisse zeigen. Eine zunächst anstehende Erweiterung ist die Integration der Gruppierungen aus linear angeordneten geschlossenen Figuren. Des weiteren kann die Berücksichtigung der Kombination aus Konturen und Regionen zusätzliche Evidenz in die Bestimmung der jeweiligen datengetriebenen Signifikanz einbringen. Hier ist jedoch zunächst nicht offensichtlich, wie diese Information quantitativ einbezogen werden kann, bzw. ob nur eine binäre Aussage zusätzlich berücksichtigt wird.

8. Zusammenfassung

Die Motivation für die vorliegende Arbeit ergibt sich aus den perzeptiven Fähigkeiten des menschlichen Sehsystems, die Umwelt in figuralen Einheiten, sogenannten Gestalten, wahrzunehmen. Dabei beruhen diese Fähigkeiten in hohem Maße auf der Identifikation von Regularitäten in einer Szene, welche eine Gruppierung von primitiven Bildelementen in komplexere Einheiten bewirken. Diese komplexeren Einheiten unterstützen dann eine symbolische Beschreibung des wahrgenommenen Bildinhaltes, beispielsweise eines Objektes, und letztendlich eine vollständige Interpretation der Szene.

Bereits Anfang des letzten Jahrhunderts beschäftigten sich Gestaltpsychologen mit diesen Wahrnehmungsprozessen des menschlichen Sehsystems. Aus ihren Untersuchungen formulierten sie Gestaltgesetze, die die grundlegenden Prinzipien für die Identifikation von Regularitäten in einer betrachteten Szene beschreiben. In den letzten 20 Jahren wurden diese Gesetze zunehmend auch als Motivation herangezogen, ähnliche Gruppierungsprozesse in automatische Bildanalyzesysteme zu integrieren und damit eine Szeneninterpretation zu unterstützen. Zur Begründung einer strukturierten Wahrnehmung wurde hierzu das Prinzip des „*non-accidentalness*“ vorgeschlagen, welches besagt, daß die Beobachtung von Regelmäßigkeiten mit großer Sicherheit auf reale Ereignisse in der Umwelt zurückzuführen sind und vermutlich eine gemeinsame kausale Ursache besitzen, beispielsweise das Vorhandensein eines Objektes.

Aus diesen Überlegungen heraus ergibt sich auch die Zielsetzung der vorliegenden Arbeit, Gruppierungsprozesse für eine automatische Bildanalyse algorithmisch zu realisieren, und so eine stärkere Strukturierung und Abstraktion einer symbolischen Bildbeschreibung für Interpretationsaufgaben zur Verfügung zu stellen. Hierzu werden Methoden zum perzeptiven Gruppieren von initialen Konturbeschreibungen entwickelt. Das gesamte System ist dabei hierarchisch angelegt, wobei von einer Ebene zur nächsten eine zunehmend abstraktere Beschreibung der Strukturen im Bild modelliert wird. Darüber hinaus wird eine Kombination der konturbasierten Gruppierungen und der dazu dualen Bildbeschreibung durch Regionen vorgestellt. Diese wird genutzt, um zum einen den konturbasierten Gruppierungsprozeß zu unterstützen und zum anderen die Erkennungsleistung einer regionenbasierten Objekterkennung zu verbessern. Dabei ist hervorzuheben, daß die entwickelten Gruppierungstechniken weitestgehend problemunabhängige Methoden verfolgen.

Ausgangspunkt für die entwickelten Gruppierungstechniken ist eine Kontursegmentierung, die im Intensitätskanal statischer Bilder durchgeführt wird. Diese soll eine möglichst vollständige Beschreibung der sichtbaren Konturen im Bild erzielen. Ein So-

beloperator realisiert zunächst die Identifikation von Diskontinuitäten im Bild. Die ermittelten Helligkeitsübergänge werden anschließend mit einem *non-maximum-suppression* sowie einem Verfahren mit Hystereseschwellwert auf Bildpunkte mit maximalen Gradientenbetrag verdünnt. Zur Beschreibung dieser Kantenelemente durch geometrische Modelle wie Geradenstücke, Kreis- und Ellipsenbögen kommt das Verfahren nach Leonardis [Leo93] zur Konturapproximation zum Einsatz. In einer *Model Recovery* Phase wird hierzu sukzessive eine Menge von Konturmodellen ermittelt, die jeweils mit unterschiedlicher Parametrisierung und Komplexität die Menge der Kantenelemente adäquat beschreiben. Bedingt durch die Strategie der *Model Recovery* Phase kommt es jedoch zu redundanten Beschreibungen von Kantenelementen durch Konturmodelle. In einer *Model Selection* Phase wird daher über eine Zielfunktion die Menge der Konturmodelle selektiert, die die Kantenelemente möglichst vollständig, mit kleinem Approximationsfehler und wenigen Parametern beschreiben. Auf Basis von geometrischen Modellen erfolgt so initial eine adäquate symbolische Beschreibung des Bildinhaltes.

In einem parallelen Verarbeitungspfad realisiert ein Regionensegmentierungsverfahren eine Identifikation von Bereichen homogener Farbe in Farbbildern. Hierbei kommt ein domänenabhängiges Verfahren zum Einsatz, das zunächst über einen Polynomklassifikator für jeden Bildpunkt eine Reduzierung des Farbraumes auf die in der *baufix*[®] Domäne auftretenden Farben durchführt (vergleiche [Kum97]). Anschließend werden durch eine Glättung und ein *Merging* Verfahren Bereiche homogener Farbe ermittelt. Eine *Connected-Component* Analyse wird dann zur Regionenmarkierung eingesetzt. Die regionenbasierte Beschreibung des Bildinhaltes erfolgt abschließend durch polygonale Approximationen der Regionengrenzen. Alternativ zu dieser Methode kann auch das domänenunabhängige *Island Growing* Verfahren nach [Pri93] eingesetzt werden.

Für die Modellierung der konturbasierten Gruppierungsprozesse wird eine Gruppierungshierarchie vorgeschlagen, die verschiedene Typen von Gruppierungen in unterschiedlichen Ebenen definiert. Die verwendeten Gruppierungstypen sind durch die Gestaltgesetze für Nähe, Ähnlichkeit, gute Fortsetzung, Symmetrie und Geschlossenheit motiviert. In der **1D**-Ebene werden initiale Konturmodelle in Kollinearitäten, Kurvilinearitäten und, bezüglich einer Endpunktnähe, in Proximitäten gruppiert. Aufbauend auf diesen linearen Gruppen realisiert die nächste **2×1D**-Ebene Gruppierungen aufgrund von Parallelität und Symmetrie. Zur Quantifizierung der hierfür eingesetzten Gestaltgesetze wurden aus einer Stichprobe Suchbereiche perzeptiver Aufmerksamkeit individuell für jedes Gruppierungsprinzip ermittelt. Auch einschränkende Bedingungen wurden aus dieser Stichprobe abgeleitet und damit die Hypothesisierung auf eine empirische Basis gestellt. Die **2D**-Ebene beschreibt Strukturen, die das Gestaltgesetz von Geschlossenheit abbilden. Für ihre Hypothesisierung kommen initiale Konturmodelle und Gruppierungen aus der **1D**- und **2×1D**-Ebene als strukturbildende Elemente zum Einsatz. Den Proximitäten kommt dabei eine besondere Bedeutung zu, da sie Nachbarschaftsrelationen zwischen den linearen Gruppen etablieren. Auf Basis einer Graphrepräsentation werden einfache zyklische Pfade aus einer alternierenden Abfolge von linearen Gruppen und Proximitäten als geschlossene Figuren hypothesiert. Die oberste **N×2D**-Ebene or-

ganisiert geschlossene Figuren, motiviert durch die Gestaltgesetze der Nähe, Ähnlichkeit und guten Fortsetzung, wiederum in lineare Gruppen. Unter Verwendung verschiedener Merkmale wird hierzu zunächst die Gruppierungswirkung aufgrund von Ähnlichkeit und Nähe zwischen Polygonpaaren mit jeweils einem Polynomklassifikator bestimmt und beurteilt. Räumlich disjunkte geschlossene Figuren werden anschließend in Abhängigkeit dieser Beurteilung in Beziehung zueinander gesetzt. Eine Repräsentation dieser Beziehungen durch gewichtete Kanten in einem Graphen dient schließlich als Basis, um Pfade linear angeordneter geschlossener Figuren mittels des Approximationsverfahrens nach Leonardis durch Geradenstücke und Ellipsenbögen zu beschreiben. Damit liegt eine Beschreibung der abstrakten linearen Gruppen vor, die eine erneute Organisation in den Ebenen $1D$ bis $N \times 2D$ ermöglicht. Es wird so ein rekursives Modell der Gruppierungshierarchie entwickelt, das Strukturen mit immer abstrakterem und komplexerem Erscheinungsbild beschreiben kann. Zudem ist dieses Gruppierungsverfahren auch auf regionenbasierte Bildbeschreibungen anwendbar.

Neben diesen Methoden zur Generierung konturbasierter Gruppierungshypothesen wird ein Verfahren zur Kombination dualer Bildbeschreibungen entwickelt. Für den Menschen ist eine Szeneninterpretation alleine aus einer Beschreibungsform des Bildes häufig schwieriger, als die Ausnutzung mehrerer komplementärer Beschreibungen. Aus dieser Beobachtung heraus erfolgt der Vorschlag eines Verfahrens, das eine Kombination von Konturgruppen und Regionen etabliert, wenn beide Bildbeschreibungen mit hoher Wahrscheinlichkeit aufgrund der selben kausalen Ursache im Bild entstanden sind. Diese Ursache kann dabei beispielsweise die Projektion einer Objektoberfläche sein. So werden in diesem Verfahren Konturgruppen den Regionen zugeordnet, wenn sie den Umriß der Region approximieren oder eine Struktur innerhalb der Region beschreiben. Hierzu erfolgt eine Berechnung verschiedener Merkmale zwischen Konturgruppen und Regionen, die sich in einer räumlichen Nachbarschaft zueinander befinden. Ein Polynomklassifikator klassifiziert anschließend eine Konturgruppe als umrissbeschreibend oder strukturbeschreibend für eine Region. Eine solche Zuordnung bietet verschiedene Anwendungsfelder. In der vorliegenden Arbeit werden umrissbeschreibende Zuordnungen unter anderem genutzt, um auf Basis von Regionengrenzen zusätzliche Proximitätsbeziehungen zwischen linearen Konturgruppen zu hypothesieren. Der konturbasierte Gruppierungsprozeß wird somit um zusätzliche Evidenz für Gruppierungshypothesen angereichert. Aber auch die Integration in ein rückgekoppeltes hybrides Objekterkennungssystem wird als Anwendung vorgestellt. Dabei werden konturbasierte geschlossene Figuren als alternative Regionen aufgefaßt und unterstützen den Erkennungsprozeß bei komplexeren Objektkonstellationen.

Die oben beschriebenen Methoden zur Generierung konturbasierter Gruppierungshypothesen verwenden bisher nur lokale Bildinformation bei der Generierung einer Hypothese. Es erfolgt keine Berücksichtigung eines globalen Kontextes zu anderen Hypothesen. Dadurch kann es zu Mehrdeutigkeiten oder Inkonsistenzen kommen, deren Auflösung hier mit einer globalen Bewertung aller Hypothesen auf Basis von Markov Random Fields angestrebt wird. Letztendlich verfolgt diese Bewertung das Ziel, eine

Selektion von signifikanten Gruppierungshypothesen zu ermöglichen. Hierzu bilden die Gruppierungen die Knoten in einem Graphen, dessen Kanten die Abhängigkeiten zwischen Hypothesen repräsentieren. Diese Abhängigkeiten sind über Bestandteilsbeziehungen und konkurrierende Interpretationen von Gruppierungen definiert. Jeder Hypothese wird zudem eine Signifikanz zugeordnet, die die Korrektheit der vorgenommenen Interpretation sowohl aufgrund der zugrundeliegenden Daten, als auch im Kontext zu anderen Hypothesen beurteilt. Eine entsprechende Modellierung von Clique-Potentialen gewährleistet dabei die Berücksichtigung der jeweiligen Datenabhängigkeit, und die Repräsentation des a priori Wissens über die verschiedenen Kontextabhängigkeiten. So wird ein a posteriori Markov Random Field definiert, dessen Energie mit dem Verfahren des Highest Confidence First minimiert wird. Es zeigt sich, daß mit dieser Modellierung eine erfolgreiche Bewertung durchgeführt wird und *die* Gruppierungen bei der Energieminimierung als signifikant markiert werden, die wesentliche Bildstrukturen beschreiben. Gleichzeitig erfolgt eine starke Reduzierung von inkompatiblen und mehrdeutigen Gruppierungshypothesen.

Der Formalismus von Markov Random Fields wird ebenfalls verwendet, um eine Konzeption für ein Verfahren zur Kontursegmentierung in Bildfolgen zu entwickeln. Diese Konzeption verfolgt dabei das Ziel, eine Basis für ein temporales perzeptives Gruppieren zu schaffen. Zur Realisierung repräsentieren die initialen Kantenelemente eines Bildes die Stellen in einem Markov Random Field, dessen Nachbarschaftssystem über eine räumliche lokale Nachbarschaft der Kantenelemente definiert ist. Als Ziel sollen den Kantenelementen die Konturmodelle aus dem zuletzt verarbeiteten Bild zugeordnet werden, die auch die aktuellen Kantenelemente adäquat beschreiben. Hierzu wird in entsprechenden Clique-Potentialen a priori Wissen über Kontextabhängigkeiten zwischen den Kantenelementen und eine Datenabhängigkeit der einzelnen Kantenelemente modelliert. Abschließend realisiert eine Energieminimierung dann eine Zuordnung der zuvor generierten Konturmodelle zu den aktuellen Kantenelementen. Mit dieser Zuordnung wird dabei gleichzeitig eine Identifikation korrespondierenden Konturmodellen über die Zeit vorgenommen. Damit wird ein Konzept vorgeschlagen, daß es erlaubt in Verbindung mit den übrigen Gruppierungstechniken ein räumlich-zeitliches Gruppieren zu realisieren.

Insgesamt wird mit der vorliegenden Arbeit ein in sich geschlossener perzeptiver Gruppierungsansatz präsentiert, der eine zunehmend abstraktere Beschreibung bildinhärenter Strukturen realisiert. Darüber hinaus wird mit der Kombination dualer Bildbeschreibungen auf den Aspekt eingegangen, daß perzeptive Organisation ein ganzheitliches Konzept darstellen sollte. Die Konzentration von Gruppierungsprozessen auf nur eine Modalität erscheint auf Dauer nicht erfolgversprechend und wird zudem durch biologische Vorbilder nicht unterstützt. Den Rahmen der vorliegenden Arbeit bildet der visuelle Systemteil eines Situierten Künstlichen Kommunikators, dessen Leistungsfähigkeit für eine Szeneninterpretation mit den entwickelten Methoden signifikant gesteigert werden kann. Die allgemeine Tragfähigkeit und Problemunabhängigkeit der vorgeschlagenen Methoden wird zudem durch Ergebnisse in der Anwendung auf anderen Domänen unterstrichen.

Literaturverzeichnis

- [Ack97] F. Ackermann, A. Maßmann, S. Posch, G. Sagerer, D. Schlüter: *Perceptual Grouping of Contour Segments Using Markov Random Fields*, *International Journal of Pattern Recognition and Image Analysis*, Bd. 7, Nr. 1, 1997, S. 11–17.
- [Alq98] L. Alquier, P. Montesinos: *Representation of linear structures using perceptual grouping*, URL: http://marathon.csee.usf.edu/~sarkar/pocv_program.html, Juni 1998, IEEE Workshop on Perceptual Organization in Computer Vision.
- [Alt94] H. Alt, O. Aichholzer, G. Rote: *Matching Shapes with a reference point*, in *Proc. ACM Symp. on Computational Geometry*, 1994, S. 85–92.
- [Alt95] H. Alt, M. Godau: *Computing the Fréchet Distance between Polygonal Curves*, *International Journal of Computational Geometry and Applications*, Bd. 5, Nr. 1&2, 1995, S. 75–91.
- [Alt99] H. Alt, L. Guibas: *Discrete Geometric Shapes: Matching, Interpolation, and Approximation*, in J.-R. Sack, J. Urrutia (Hrsg.): *Handbook of Computational Geometry*, Elsevier Science Publishers B.V. North-Holland, Amsterdam, 1999, S. 121 – 153.
- [Ami98] A. Amir, M. Lindenbaum: *A Generic Grouping Algorithm and Its Quantitative Analysis*, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Bd. 20, Nr. 2, 1998, S. 168–185.
- [Ami99] A. Amir, M. Lindenbaum: *Ground from Figure Discrimination*, *Computer Vision and Image Understanding*, Bd. 76, Nr. 1, Okt. 1999, S. 7–18.
- [Ark91] E. M. Arkin, L. P. Chew, D. P. Huttenlocher, K. Kedem, J. S. B. Mitchell: *An Efficiently Computable Metric for Comparing Polygonal Shapes*, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Bd. PAMI-13, Nr. 3, März 1991, S. 209–216.
- [Bal82] D. Ballard, C. Brown: *Computer Vision*, Prentice Hall, London, 1982.
- [Bäs93] H. Bäsmann, P. W. Besslich: *Bildverarbeitung AdOculos*, Springer, Berlin, 2. Ausg., 1993.

- [Ben92] J. Benois, D. Barba: *Image segmentation by region-contour cooperation for image coding*, in *Proc. International Conference on Pattern Recognition*, Bd. C, 1992, S. 331–334.
- [Ben95] M. Ben-Av, D. Sagi: *Perceptual Grouping by Similarity and Proximity: Experimental Results can be Predicted by Intensity Autocorrelations*, *Vision Research*, Bd. 35, Nr. 6, März 1995, S. 853–866.
- [Bes74] J. Besag: *Spatial Interaction and Statistical Analysis of Lattice Systems*, *Journal of the Royal Statistical Society, Series B*, Bd. 36, Juli 1974, S. 192–236.
- [Bie87a] I. Biederman: *Matching Image Edges to Object Memory*, in *Proc. 1. Int. Conf. on Computer Vision*, London, 1987, S. 384–392.
- [Bie87b] I. Biederman: *Recognition-by-Components: A Theory of Human Image Understanding*, *Psychological Review*, Bd. 94, Nr. 2, 1987, S. 115–147.
- [Bom97] T. Bomberg, S. Posch: *Regionensegmentierung von Farbbildfolgen*, in E. Paulus, F.M.Wahl (Hrsg.): *Mustererkennung 1997. Proceedings 19. DAGM-Symposium*, Informatik Aktuell, Springer, 1997, S. 63–70.
- [Bor97] S. Borra, S. Sarkar: *A Framework for Performance Characterization of Intermediate Level Grouping Modules*, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Bd. 19, Nr. 11, nov 1997, S. 1306–1312.
- [Bor99] C. Borges: *On the Estimation of Markov Random Field Parameters*, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Bd. 21, Nr. 3, März 1999, S. 216–224.
- [Boy98] Y. Boykov, O. Veksler, R. Zabih: *Efficient Restoration of Multicolor Images with Independent Noise*, Technical Report TR98-1712, Cornell University, Computer Science, Okt. 1998.
- [Boy99a] K. L. Boyer, S. Sarkar: *Perceptual Organization in Computer Vision: Status, Challenges, and Potential*, *Computer Vision and Image Understanding*, Bd. 76, Nr. 1, Okt. 1999, S. 1–5.
- [Boy99b] Y. Boykov, D. Huttenlocher: *A New Bayesian Framework for Object Recognition*, in *Proc. Computer Vision and Pattern Recognition*, IEEE, 1999.
- [Boy00] K. L. Boyer, S. Sarkar (Hrsg.): *Perceptual Organization for Artificial Vision Systems*, Bd. 546 von *The Kluwer International Series in Engineering and Computer Science*, Kluwer Academic Publishers, Boston, März 2000.
- [Can83] J. Canny: *Finding Edges and Lines in Images*, Technical Report 720, MIT Artificial Intelligence Laboratory, 1983.

-
- [Cha97] Y.-L. Chang, J. K. Aggarwal: *Line Correspondences from Cooperating Spatial and Temporal Grouping Processes for a Sequence of Images*, *Computer Vision and Image Understanding*, Bd. 67, Nr. 2, Aug. 1997, S. 186–201.
- [Cha99] A. Chakraborty, J. S. Duncan: *Game-Theoretic Integration for Image Segmentation*, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Bd. 21, Nr. 1, Jan. 1999.
- [Che93a] R. Chellappa, A. Jain: *Markov Random Fields: Theory and Applications*, Academic Press, 1993.
- [Che93b] P. P. Chew, M. T. Goodrich, D. P. Huttenlocher, K. Kedem, J. M. Kleinberg, D. Kravets: *Geometric pattern matching under Euclidean motion*, in *Proc. 5th Canadian Conference on Computational Geometry*, 1993, S. 151–156.
- [Cho90] P. B. Chou, C. M. Brown: *The Theory and Practice of Bayesian Image Labeling*, *International Journal of Computer Vision*, Bd. 4, 1990, S. 185–210.
- [Cho93] P. Chou, P. Cooper, M. Swain, C. Brown, L. Wixson: *Probabilistic Network Inference for Cooperative High and Low Level Vision*, in R. Chellappa, A. Jain (Hrsg.): *Markov Random Fields, Theory and Application*, Academic Press, 1993, S. 211–243.
- [Cor00] B. Cordes: *Konturzuordnung mittels Markov Random Fields*, Diplomarbeit, Universität Bielefeld, Technische Fakultät, AG Angewandte Informatik, 2000.
- [Cre99] D. Crevier: *A Probabilistic Method for Extracting Chains of Collinear Segments*, *Computer Vision and Image Understanding*, Bd. 76, Nr. 1, Okt. 1999, S. 36–53.
- [Cre00] D. Crevier: *Bayesian Extraction of Collinear Segment Chains from Digital Images*, in K. L. Boyer, S. Sarkar (Hrsg.): *Perceptual Organization for Artificial Vision Systems*, Bd. 546 von *The Kluwer International Series in Engineering and Computer Science*, Kluwer Academic Publishers, Boston, März 2000, S. 311–323.
- [dB96] M. de Berg, O. Devillers, M. van Kreveld, O. Schwarzkopf, M. Teillaud: *Computing the maximum overlap of two convex polygons under translations*, Dept. of Comp. Science, University of Utrecht, 1996.
- [Den96] J. Denzler, H. Niemann: *A New Energy Term Combining Kalman-Filter and Active Contour Models for Object Tracking*, *Machine Graphics & Vision*, Bd. 5, Nr. 1/2, 1996, S. 157–165.
- [Den97] J. Denzler: *Aktives Sehen zur Echtzeitobjektverfolgung*, Bd. 163 von *Dissertationen zur künstlichen Intelligenz*, Infix, St. Augustin, 1997.

- [Des99] X. Descombes, R. Morris, J. Zerubia, M. Berthod: *Estimation of Markov Random Field Prior Parameters Using Markov Chain Monte Carlo Maximum Likelihood*, *IEEE Trans. on Image Processing*, Bd. 8, Nr. 7, Juli 1999, S. 954.
- [Ete92] A. Etemadi: *Robust Segmentation of Edge Data*, in *International Conference on Image Processing and its Applications*, Maastricht, April 1992, S. 311–314.
- [Fri97] N. S. Friedland, A. Rosenfeld: *An Integrated Approach to 2D Object Recognition*, *Pattern Recognition*, Bd. 30, Nr. 3, 1997, S. 525–535.
- [Gei91] D. Geiger, F. Girosi: *Parallel and Deterministic Algorithms from MRF's: Surface Reconstruction*, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Bd. 13, Nr. 4, 1991, S. 401–412.
- [Gem84] S. Geman, D. Geman: *Stochastic Relaxation, Gibbs Distribution, and the Bayesian Restoration of Images*, *IEEE Transactions On Pattern Analysis And Machine Intelligence*, Bd. 6, Nr. 6, 1984, S. 721–741.
- [Gev00] T. Gevers, H. Stokman: *Classifying Color Transitions into Shadowgeometry, Illumination, Highlight, or Material Edges*, in *Proc. International Conference on Image Processing*, IEEE, Sep. 2000.
- [Gon91] R. Gonzalez, P. Wintz: *Digital image processing*, Addison-Wesley, 2. Ausg., 1991.
- [HA93] J. B. Helmut Alt, Bernd Behrends: *Approximate Matching of Polygonal Shapes*, B 93–10, Freie Universität Berlin, Fachbereich Mathematik und Informatik, Institut für Informatik, 1993.
- [Har92] R. Haralick, L. Shapiro: *Computer and Robot Vision*, Addison-Wesley, 1992.
- [Hea97] M. D. Heath, S. Sarkar, T. Sanocki, K. W. Bowyer: *A robust visual method for assessing the relative performance of Edge-detection algorithms*, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Bd. 19, Nr. 12, dec 1997, S. 1338–1359.
- [Hei93] F. Heitz, P. Bouthemy: *Multimodal estimation of discontinuous optical flow using Markov random fields*, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Bd. 15, Nr. 12, 1993, S. 1217–32.
- [Hei98] G. Heidemann: *Ein flexibel einsetzbares Objekterkennungssystem auf der Basis neuronaler Netze*, Dissertation, Universität Bielefeld, Technische Fakultät, 1998, Infix, DISKI 190.

- [Hei99] G. Heidemann, N. Jungclaus, F. Kummert, G. Sagerer, H. Ritter: *Ein hybrides Bildanalyse-System für einen künstlichen Kommunikator*, *Kognitions-wissenschaft*, Bd. 8, Nr. 3, 1999, S. 101–107.
- [Hoc76] J. E. Hochberg, A. Silverstein: *A Quantitative Index of Stimulus Similarity: Proximity vs. Difference in Brightness*, in G. M. Murch (Hrsg.): *Studies in perception*, The Bobbs-Merrill Company, Inc., Indianapolis, 1976, S. 123–125.
- [Hua79] T. Huang, G. Yang, G. Tang: *A fast two-dimensional median filtering algorithm*, *IEEE Transactions on Acoustics, Speech, and Signal Processing*, Bd. 27, Nr. 1, Feb. 1979, S. 13–18.
- [Hue73] M. H. Hueckel: *A Local Visual Operator Which Recognizes Edges and Lines*, *Journal of the ACM*, Bd. 20, Nr. 4, Okt. 1973, S. 634–647.
- [JD97] A. Jacot-Descombes, T. Pun: *Asynchronous Perceptual Grouping: From Contours to Relevant 2-D Structure*, *Computer Vision and Image Understanding*, Bd. 66, Nr. 1, 1997, S. 1–24.
- [Kam00] S. Kamijo, Y. Matsushita, K. Ikeuchi, M. Sakauchi: *Occlusion Robust Tracking Utilizing Spatio-Temporal Markov Random Field Model*, in *Proc. International Conference on Pattern Recognition*, Bd. 1, Barcelona, Sep. 2000, S. 140–144.
- [Kan79] G. Kanizsa: *Organization in vision*, Praeger, 1979.
- [Kim93] I. Kim, H. Yang: *Efficient Image Labeling Based on Markov Random Field and Error Backpropagation Network*, *Pattern Recognition*, Bd. 26, 1993, S. 1695–1707.
- [Kim94] I. Kim, H. Yang: *A Systematic Way for Region-Based Image Segmentation Based on Markov Random-Field Model*, *Pattern Recognition Letters*, Bd. 15, Nr. 10, Okt. 1994, S. 969–976.
- [Kin80] R. Kindermann, J. L. Snell: *Markov Random Fields and their Applications*, American Mathematical Society, Providence, 1980.
- [Koh95] P. Kohlhepp: *Combining region and edge based segmentation for robust 3D localization in range images*, in *Proc. Asian Conference on Computer Vision*, Bd. 2, IEEE, Singapore, 1995, S. 254–258.
- [Kor94] M. Korff: *Regionensegmentierung auf der Basis von Farben*, Diplomarbeit, Universität Bielefeld, Technische Fakultät, AG Angewandte Informatik, 1994.

- [Kri95] S. Krishnamachari, R. Chellappa: *Multiresolution GMRF models for texture segmentation*, in *Proc. International Conference on Image Processing*, Washington D.C., USA, Okt. 1995, S. 2407–2410.
- [Kri96] S. Krishnamachari, R. Chellappa: *Delineating buildings by grouping lines with MRF's*, *IEEE Trans. on Image Processing*, Bd. 5, 1996, S. 164–168.
- [Kub00] M. Kubovy, S. Gepshtein: *Gestalt: From phenomena to laws*, in K. L. Boyer, S. Sarkar (Hrsg.): *Perceptual Organization for Artificial Vision Systems*, Bd. 546 von *The Kluwer International Series in Engineering and Computer Science*, Kluwer Academic Publishers, Boston, März 2000, S. 41–71.
- [Kum97] F. Kummert, G. Fink, G. Sagerer: *Schritthaltende hybride Segmentierung*, in E. Paulus, F.M.Wahl (Hrsg.): *Mustererkennung 1997. Proceedings 19. DAGM-Symposium*, Informatik Aktuell, Springer, 1997, S. 137–144.
- [Kum98a] F. Kummert: *Interpretation von Bild- und Sprachsignalen – Ein hybrider Ansatz*, Shaker Verlag, 1998.
- [Kum98b] F. Kummert, G. Fink, G. Sagerer, E. Braun: *Hybrid Object Recognition in Image Sequences*, in *14th International Conference on Pattern Recognition*, Bd. II, Brisbane, 1998, S. 1165–1170.
- [Lan00] S. Lang: *Perzeptives Gruppieren von Regionen*, Diplomarbeit, Universität Bielefeld, Technische Fakultät, AG Angewandte Informatik, 2000.
- [Lei97] J. M. N. Leitao, M. A. T. Figueiredo: *Unsupervised Image Restoration and Edge Location Using Compound Gauss-Markov Random Fields and the MDL Principle*, *IEEE Trans. on Image Processing*, Bd. 6, Nr. 8, 1997, S. 1089–.
- [Leo93] A. Leonardis: *Image Analysis Using Parametric Models*, Dissertation, University of Ljubljana, 1993.
- [Leu98] T. Leung, J. Malik: *Contour continuity in region based image segmentation*, in *Proc. European Conference on Computer Vision*, Bd. 1406 von *Lecture Notes in Computer Science*, 1998, S. 544–??
- [Li94] S. Z. Li: *A Markov Random Field Model for Object Matching under Contextual Constraints*, in *Proc. Computer Vision and Pattern Recognition*, IEEE Computer Society Press, Los Alamitos, CA, USA, Juni 1994, S. 866–869.
- [Li95] S. Li: *Markov Random Field Modeling in Computer Vision*, Springer, 1995.
- [Lin98] C. Lin, R. Nevatia: *Building Detection and Description from a Single Intensity Image*, *Computer Vision and Image Understanding*, Bd. 72, Nr. 2, Nov. 1998, S. 101–121.

- [Lin00a] M. Lindenbaum, A. Berengolts: *An Observation on Saliency*, in K. L. Boyer, S. Sarkar (Hrsg.): *Perceptual Organization for Artificial Vision Systems*, Bd. 546 von *The Kluwer International Series in Engineering and Computer Science*, Kluwer Academic Publishers, Boston, März 2000, S. 239–248.
- [Lin00b] M. Lindenbaum, A. Berengolts: *A Probabilistic Interpretation of the Saliency Network*, in *Proc. European Conference on Computer Vision*, 2000.
- [Low85] D. G. Lowe: *Perceptual Organization and Visual Recognition*, Kluwer Academic Publishers, 1985.
- [Lu00] Q. Lu, T. Jiang: *A New Bayesian Approach to Image Denoising with a Combination of MRFs and Pixon Method*, in *Proc. International Conference on Pattern Recognition*, Bd. 15/3, Barcelona, Sep. 2000, S. 726–729.
- [Lut94] F. Luthon, G. V. Popescu, A. Caplier: *An MRF based motion detection algorithm implemented on analog resistive network*, in *Proc. European Conference on Computer Vision*, Bd. 800 von *Lecture Notes in Computer Science*, 1994, S. 167–174.
- [Mal99] J. Malik, S. Belongie, J. Shi, T. Leung: *Textons, Contours and Regions: Cue Integration in Image Segmentation*, in *Proc. International Conference on Computer Vision*, Bd. 2, Corfu, Greece, 1999, S. 918–925.
- [Mal00] J. Malik, S. Belongie, J. Shi, T. Leung: *Contour and Texture Analysis for Image Segmentation*, in K. L. Boyer, S. Sarkar (Hrsg.): *Perceptual Organization for Artificial Vision Systems*, Bd. 546 von *The Kluwer International Series in Engineering and Computer Science*, Kluwer Academic Publishers, Boston, März 2000, S. 139–172.
- [Mar80] D. Marr, E. C. Hildreth: *Theory of edge detection*, in *Proc. of Royal Society London*, Bd. 207(B), 1980, S. 187–217.
- [Mar89] J. Marroquin: *A Markovian Random Field of Piecewise Straight Lines*, *Biological Cybernetics*, Bd. 61, 1989, S. 457–465.
- [Maß95a] A. Maßmann, S. Posch: *Bereiche perzeptiver Aufmerksamkeit für konturbasierte Gruppierung*, in G. Sagerer, S. Posch, F. Kummert (Hrsg.): *Mustererkennung 1995. Proceedings 17. DAGM-Symposium*, Informatik Aktuell, Springer, 1995, S. 602–609.
- [Maß95b] A. Maßmann, S. Posch: *Mask-Oriented Grouping Operations in a Contour-Based Approach*, in *Proc. Asian Conference on Computer Vision*, Bd. 3, IEEE, Singapore, 1995, S. 58–61.

- [Maß96] A. Maßmann, S. Posch, D. Schlüter: *Perzeptives Gruppieren von Ko- und Kurvilinearitäten mittels Markov Random Fields*, in B. Jähne, P. Geißler, H. Haußecker, F. Hering (Hrsg.): *Mustererkennung 1996. Proceedings 18. DAGM-Symposium*, Informatik Aktuell, Springer, 1996, S. 235–242.
- [Maß97] A. Maßmann, S. Posch, G. Sagerer, D. Schlüter: *Using Markov Random Fields for Contour-Based Grouping*, in *Proc. International Conference on Image Processing*, Bd. II, IEEE, 1997, S. 207–210.
- [Mil92] D. Milun, D. Sher: *Improving Edge Detectors on Compressed Images - A Trainable Markov Random Field Approach*, TR 92-22, SUNY at Buffalo, Sep. 1992.
- [Mod93] J. W. Modestino, J. Zhang: *A Markov Random Field Model-Based Approach to Image Interpretation*, in R. Chellappa, A. Jain (Hrsg.): *Markov Random Fields, Theory and Application*, Academic Press, 1993, S. 369–408.
- [Moh89] R. Mohan: *Perceptual Organization for Computer Vision*, Dissertation, University of Southern California, 1989.
- [Moh92] R. Mohan, R. Nevatia: *Perceptual Organization for Scene Segmentation and Description*, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Bd. 14, Nr. 6, 1992, S. 616–635.
- [Mok97] M. Mokhtari, R. Bergevin: *Multiscale Segmentation and Approximation for Significant Description of 2D Contours*, in *Proc. International Conference on Image Processing*, Bd. I, 1997, S. 212–215.
- [Mur93] V. Murino, G. Foresti, C. Regazzoni: *A Markovian Model for Perceptual Grouping of Different Shape Primitives*, in V. Roberto (Hrsg.): *Intelligent Perceptual Systems*, Springer, 1993, S. 161–180.
- [Nad96] S. Nadabar, A. Jain: *Parameter estimation in Markov random field contextual models using geometric models of objects*, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Bd. 18, Nr. 3, 1996, S. 326–329.
- [Nie83] H. Niemann: *Klassifikation von Mustern*, Springer-Verlag, Berlin, 1983.
- [Os98] W. Osberger, A. J. Maeder: *Automatic Identification of Perceptually Important Regions in an Image using a Model of the Human Visual System*, in B. C. L. Anil K. Jain, Svetha Venkatesh (Hrsg.): *Proc. International Conference on Pattern Recognition*, IEEE, Brisbane, Australia, Aug. 1998, S. 701–704.

-
- [Pag00] R. Paget, I. Longstaff: *Nonparametric Markov Random Field Model Analysis of the MeasTex Test Suite*, in *Proc. International Conference on Pattern Recognition*, Bd. 3, Barcelona, Sep. 2000, S. 939–942.
- [Pal94] S. Palmer, I. Rock: *Rethinking perceptual organization: The role of uniform connectedness*, *Psychonomic Bulletin & Review*, Bd. 1, Nr. 1, 1994, S. 29–55.
- [Par97] J. S. Park, J. H. Han: *Estimating optical flow by tracking contours.*, *Pattern Recognition Letters*, Bd. 18, Nr. 7, Juli 1997, S. 641–648.
- [Par98] J. S. Park, J. H. Han: *Contour motion estimation from images sequences using curvature information*, *Pattern Recognition*, Bd. 31, Nr. 1, 1998, S. 31–39.
- [Pau96] E. Pauwels, P. Fiddelaers, L. Van Gool: *Autonomous grouping of contour-segments using an adaptive region-growing algorithm*, in *Proc. International Conference on Pattern Recognition*, Bd. B, 1996, S. 586–590.
- [Pav74] T. Pavlidis, S. Horowitz: *Segmentation of plane curves*, *IEEE Transactions on Computers*, Bd. 23, 1974, S. 860–870.
- [Pav90] T. Pavlidis, Y.-T. Liow: *Integrating Region Growing and Edge Detection*, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Bd. 12, Nr. 3, März 1990, S. 225–233.
- [Pea88] J. Pearl: *Probabilistic Reasoning in Intelligent Systems*, Morgan Kaufmann, San Francisco, California, 1988.
- [Pie97] L. Piegel, W. Tiller: *The NURBS Book*, Monographs in Visual Communication, Springer, Berlin, 2. Ausg., 1997.
- [Pop94] A. Pope, D. Lowe: *Vista: A software environment for computer vision research*, in *Proc. Computer Vision and Pattern Recognition*, IEEE, 1994, S. 768–772.
- [Pos97] S. Posch: *Perzeptives Gruppieren und Bildanalyse*, Habilitationsschrift, Universität Bielefeld, 1997.
- [Pos98] S. Posch, D. Schlüter: *Perceptual Grouping using Markov Random Fields and Cue Integration of Contour and Region Information*, Technical Report 98/10, SFB 360, Universität Bielefeld, 1998.
- [Pre88] W. H. Press, et al: *Numerical recipes in C : the art of scientific computing*, Cambridge University Press, New York, 1988.

- [Pri93] L. Priebe, V. Rehrmann: *A Fast Hybrid Color Segmentation Method*, in S. Pöppel, H. Handels (Hrsg.): *Mustererkennung 1993. Proceedings 15. DAGM-Symposium*, Springer, 1993, S. 297–304.
- [Qui98] P. T. Quinlan, R. N. Wilton: *Grouping by proximity or similarity? Competition between the Gestalt principles in vision*, *Perception*, Bd. 27, Nr. 4, 1998, S. 417–430.
- [Rae99] R. Rae, M. Fislage, H. Ritter: *Visuelle Aufmerksamkeitssteuerung zur Unterstützung gestikbasierter Mensch–Maschine Interaktion, KI – Künstliche Intelligenz, Themenheft Aktive Sehsysteme*, Bd. 99, Nr. 1, 1999, S. 18–24.
- [Rae00] R. Rae: *Gestikbasierte Mensch–Maschine Kommunikation auf der Grundlage visueller Aufmerksamkeit und Adaptivität*, Dissertation, Technische Fakultät, Universität Bielefeld, 2000.
- [Ral99] A. L. Ralescu, J. G. Shanahan: *Perceptual organization for inferring object boundaries in an image*, *Pattern Recognition*, Bd. 32, Nr. 11, November 1999, S. 1923–1933.
- [Ray95] B. Ray, K. Ray: *A new split-and-merge technique for polygonal approximation of chain coded curves*, *Pattern Recognition Letters*, Bd. 16, 1995, S. 161–169.
- [Roc91] I. Rock, S. Palmer: *Das Vermächtnis der Gestaltpsychologie*, *Spektrum der Wissenschaft*, Februar 1991, S. 68–75.
- [Ros89] P. L. Rosin, G. A. West: *Segmentation of edges into lines and arcs*, *Image and Vision Computing*, Bd. 7, Nr. 2, 1989, S. 109–114.
- [Ros91] P. Rosin, G. West: *Extracting surfaces of revolution by perceptual grouping of ellipses*, in *Proc. Computer Vision and Pattern Recognition*, Maui, Juni 1991, S. 677–678.
- [Ros93] P. L. Rosin: *Multiscale Representation and Matching of Curves Using Codons*, *Computer Vision, Graphics, and Image Processing*, Bd. 55, Nr. 4, Juli 1993, S. 286–310.
- [Roy00] S. Roy, V. Govindu: *MRF Solutions for Probabilistic Optical Flow Formulations*, in *Proc. International Conference on Pattern Recognition*, Bd. 3, Barcelona, Sep. 2000, S. 1053–1059.
- [Sag01] G. Sagerer, C. Bauckhage, E. Braun, G. H. F. Kummert, H. Ritter, D. Schlüter: *Integrating Recognition Paradigms in a Multiple-path Architecture*, in *Proc. International Conference on Advances in Pattern Recognition*, IAPR, Springer-Verlag Ltd., Rio de Janeiro, Brasil, März 2001, to appear.

- [San98] T. Sanocki, K. W. Bowyer, M. D. Heath, S. Sarkar: *Are Edges Sufficient for Object Recognition*, *Journal of Experimental Psychology: Human Perception and Performance*, Bd. 24, Nr. 1, 1998, S. 1–10.
- [Sar94a] S. Sarkar: *Tracking 2D Structures using Perceptual Organizational Principles*, TR 11–94–01, Department of Computer Science and Engineering, University of South Florida, Tampa, FL 33620, Nov. 1994.
- [Sar94b] S. Sarkar, K. L. Boyer: *Computing Perceptual Organization in Computer Vision*, World Scientific Publishing Co. Pte. Ltd., 1994.
- [Sch95] D. Schlüter: *Perzeptives Gruppieren von Kontursegmenten mit Hilfe von Markov Random Fields*, Diplomarbeit, Universität Bielefeld, Technische Fakultät, AG Angewandte Informatik, 1995.
- [Sch96] J. Schürmann: *Pattern Classification*, Wiley Interscience, 1996.
- [Sch98] D. Schlüter, S. Posch: *Combining Contour and Region Information for Perceptual Grouping*, in P. Levi, R.-J. Ahlers, F. May, M. Schanz (Hrsg.): *Mustererkennung 1998. Proceedings 20. DAGM-Symposium*, Informatik Aktuell, Springer, 1998, S. 393–401.
- [Sch00a] D. Schlüter, F. Kummert, G. Sagerer, S. Posch: *Integration of regions and contours for object recognition*, in *Proc. 15th International Conference on Pattern Recognition*, Bd. I, IEEE, Barcelona, Sep. 2000, S. 944–947.
- [Sch00b] D. Schlüter, S. Wachsmuth, G. Sagerer, S. Posch: *Towards an integrated framework for contour-based grouping and object recognition using markov random fields*, in *Proc. International Conference on Image Processing*, Bd. II, IEEE, Vancouver, Sep. 2000, S. 100–103.
- [SCZ98] Y. W. Song-Chun Zhu, D. Mumford: *Filters, Random Fields and Maximum Entropy (FRAME): Towards a Unified Theory for Texture Modeling*, *International Journal of Computer Vision*, Bd. 27, Nr. 2, 1998, S. 107–126.
- [Sha88] A. Sha’ashua, S. Ullman: *Structural saliency: The detection of global salient structures using a locally connected network*, in *Proc. International Conference on Computer Vision*, 1988, S. 321–327.
- [Ste93] R. Steinbrecher: *Bildverarbeitung in der Praxis*, Oldenbourg, München, 1993.
- [Tau91] G. Taubin: *Estimation of Planar Curves, Surfaces, and Nonplanar Space Curves Defined by Implicit Equations with Application to Edge and Range Image Segmentation*, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Bd. 13, Nr. 11, 1991, S. 1115–1138.

- [Tho96] K. Thornber, L. Williams: *Analytic Solution of Stochastic Completion Fields*, *Biological Cybernetics*, Bd. 75, 1996, S. 141–151.
- [Ura95] S. Urago, J. Zerubia, M. Berthod: *A Markovian Model for Contour Grouping*, *Pattern Recognition*, Bd. 28, Nr. 5, 1995, S. 683–693.
- [Wac00] S. Wachsmuth, H. Brandt-Pook, F. Kummert, G. Socher, G. Sagerer: *Integration of Vision and Speech Understanding using Bayesian Networks*, *Videre: A Journal of Computer Vision Research*, Bd. 1, Nr. 4, 2000, S. 62–83.
- [Wal84] K. Wall, P. Danielsson: *A fast sequential method for polygonal approximation of digitized curves*, *Computer Vision, Graphics, and Image Processing*, Bd. 28, 1984, S. 220–227.
- [Wer23] M. Wertheimer: *Untersuchungen zur Lehre von der Gestalt II*, *Psychologische Forschung*, Bd. 4, 1923, S. 301–350.
- [Wil98] L. Williams, K. Thornber: *A Comparison of Measures for Detecting Natural Shapes in Cluttered Backgrounds*, in *Proc. European Conference on Computer Vision*, Bd. 1407 von *Lecture Notes in Computer Science*, Springer-Verlag, 1998.
- [Wit83] A. P. Witkin, J. M. Tenenbaum: *On the Role of Structure in Vision*, in J. Beck, B. Hope, A. Rosenfeld (Hrsg.): *Human and Machine Vision*, Academic Press, Orlando, 1983, S. 481–543.
- [Xia92] Y. Xiaohan, J. Ylä-Jääski: *Image Segmentation Combining Region Growing and Edge Detection*, in *Proc. International Conference on Pattern Recognition*, Bd. III, 1992, S. 481–484.
- [Xua95] J. Xuan, T. Adali, Y. Wang: *Segmentation of magnetic resonance brain image: Integrating region growing and edge detection*, in *Proc. International Conference on Image Processing*, Bd. C, 1995, S. 544–547.
- [YJ92] A. Ylä-Jääski, F. Ade: *Line Segment and Their Grouping*, in *Proc. International Conference on Pattern Recognition*, Bd. I, 1992, S. 750–754.
- [Zhu99] S. Zhu: *Embedding Gestalt Laws in Markov Random Fields*, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Bd. 21, Nr. 11, Nov. 1999, S. 1170–1187.

A. Bilder und Bildsegmentierung

In diesem Teil sind 12 Bilder mit ihren initialen Bildsegmentierungen abgebildet. Sie beschreiben die Teststichprobe, die zur Diskussion der initialen Bildsegmentierungsergebnisse und zur Evaluation der übrigen Gruppierungsprozesse verwendet werden. Auf der linken Seite ist jeweils das Originalfarbbild im RGB-Farbraum gezeigt. In der mittleren Spalte sind die Konturmodelle dargestellt, die aus dem Grauwertauszug des Originalbildes ermittelt werden. Für die Approximation wird das Verfahren von Leonardis eingesetzt (vergleiche Abschnitt 2.1.3). Auf der rechten Seite sind schließlich die Ergebnisse einer Regionensegmentierung mit pixelweiser Farbklassifikation gezeigt (vergleiche Abschnitt 2.2.1). Dabei erfolgt die Darstellung mit dem RGB-Farbwert der entsprechenden Klasse der **baufix**[®] Farben.

A. Bilder und Bildsegmentierung

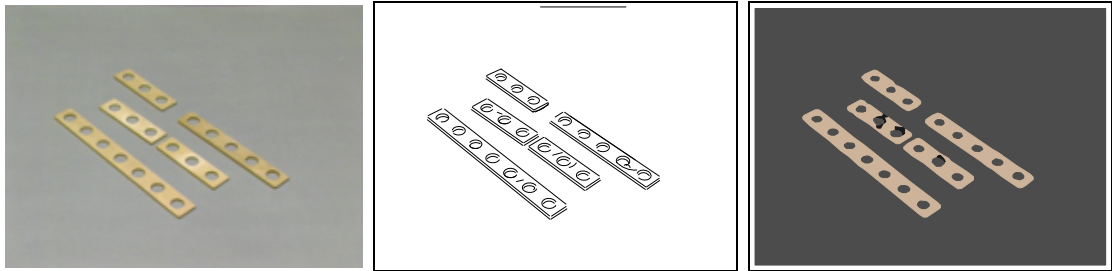


Abbildung A.1.: Bild 1

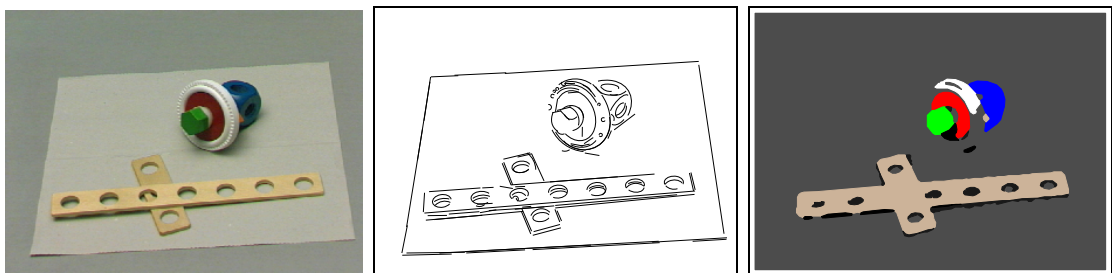


Abbildung A.2.: Bild 2

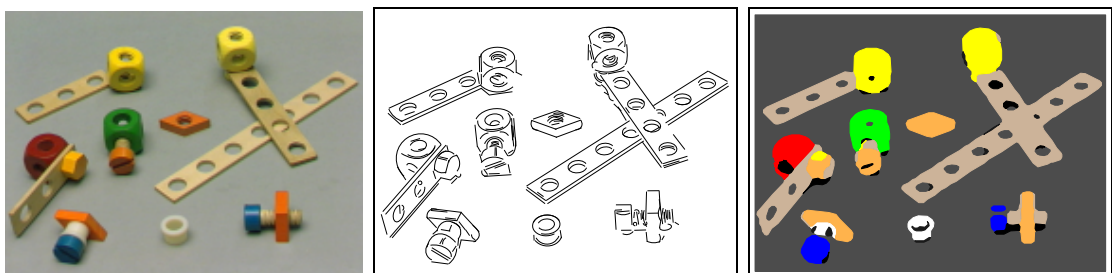


Abbildung A.3.: Bild 3

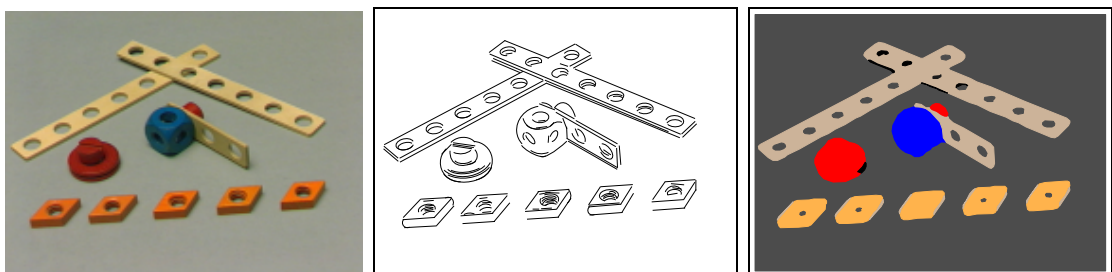


Abbildung A.4.: Bild 4

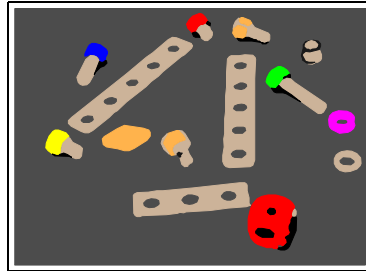
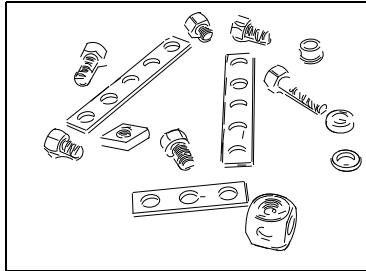
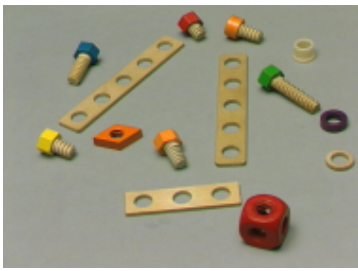


Abbildung A.5.: Bild 5

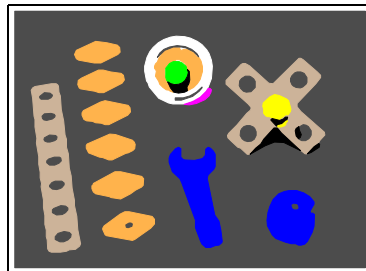
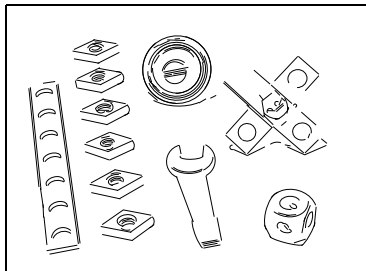


Abbildung A.6.: Bild 6

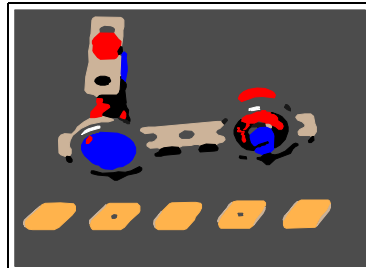
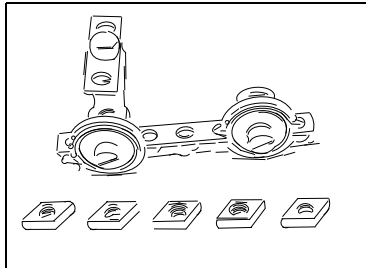
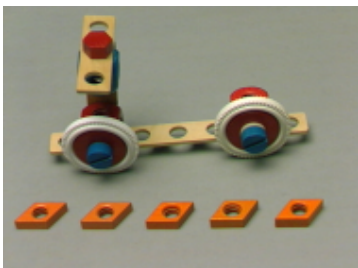


Abbildung A.7.: Bild 7

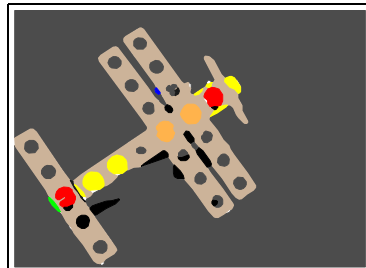
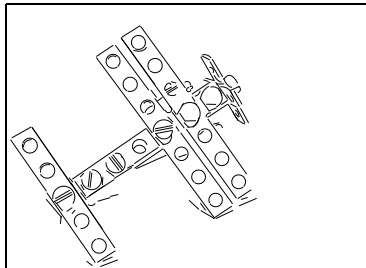
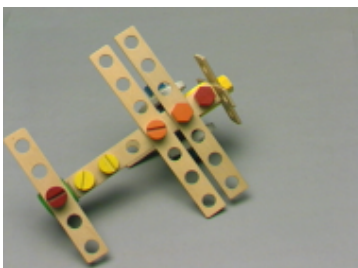


Abbildung A.8.: Bild 8

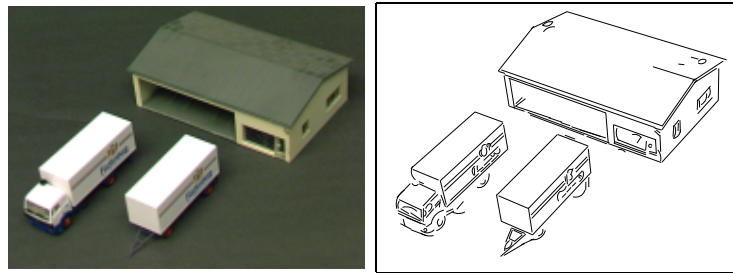


Abbildung A.9.: Bild 9

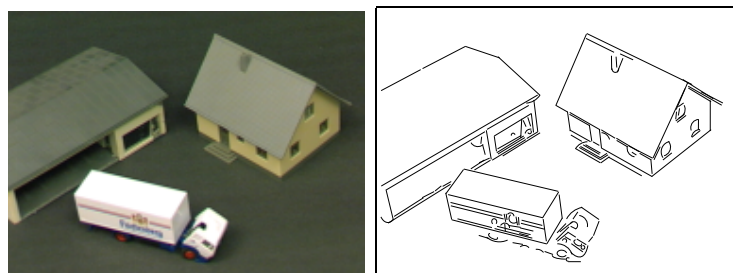


Abbildung A.10.: Bild 10

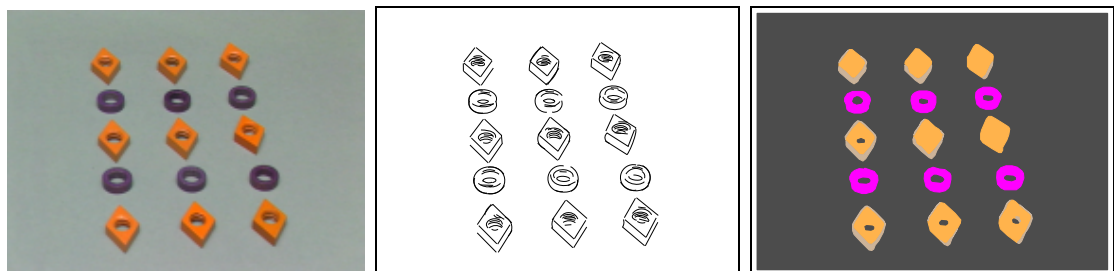


Abbildung A.11.: Bild 11

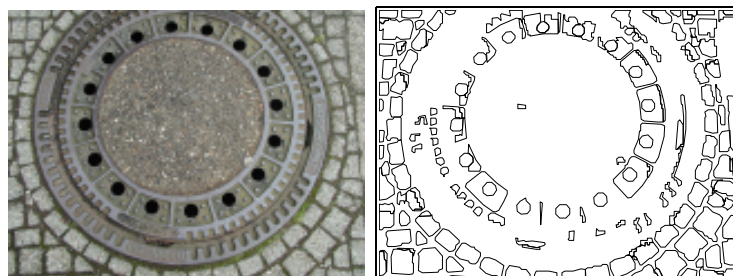


Abbildung A.12.: Bild 12

B. Gruppierungshypothesen und Kombination von Regionen und Konturgruppen

Auf den folgenden Seiten sind die Ergebnisse der Gruppierungsprozesse für die Bilder der Teststichprobe aus Anhang A dargestellt. In den Abbildungen B.1–B.8 werden jeweils die generierten Kol- und Kurvilinearitäten, Parallelitäten, Zuordnungen von linearen Gruppen und Regionen, sowie die Geschlossenheiten dargestellt. Bei Parallelitäten sind die konstituierenden Geradenstücke durch eine gepunktete Linie in ihren Mittelpunkten miteinander verbunden. Bei den Zuordnungen sind die unterschiedlichen Regionen in ihrem RGB-Farbwert der entsprechenden Klasse der *baufix*[®] Farben eingefärbt. Die zugeordneten linearen Gruppen sind schwarz gezeichnet. Die Geschlossenheiten sind jeweils durch grau schattierte Flächen mit entsprechenden Umrandungen abgebildet. Die Trennung überlappender Geschlossenheiten ist daher teilweise nur eingeschränkt möglich.

Für die Bilder 9 und 10 in den Abbildungen B.9 und B.10 entfällt die Darstellung von Zuordnungen, da hier keine Regioneninformation verwendet wurde.

Ab Seite 211 sind die Ergebnisse der Gruppierung linear angeordneter Geschlossenheiten ($\mathbf{N} \times 2\mathbf{D}$ -Ebene) für die Bilder 1 bis 8 und für die Bilder 11 und 12 dargestellt. Dabei sind die generierten Konturmodelle rot eingefärbt. Alle betrachteten Geschlossenheiten sind ebenfalls dargestellt. Die erzeugten Konturmodelle verlaufen durch die Schwerpunkte der Geschlossenheiten, die sie gruppieren (nicht explizit markiert). Damit kann grossteils nachvollzogen werden, welche Geschlossenheiten gruppiert werden. Soweit es zu rekursiven Gruppierungen in Kol- und Kurvilinearitäten kommt, sind diese entsprechend extra dargestellt.

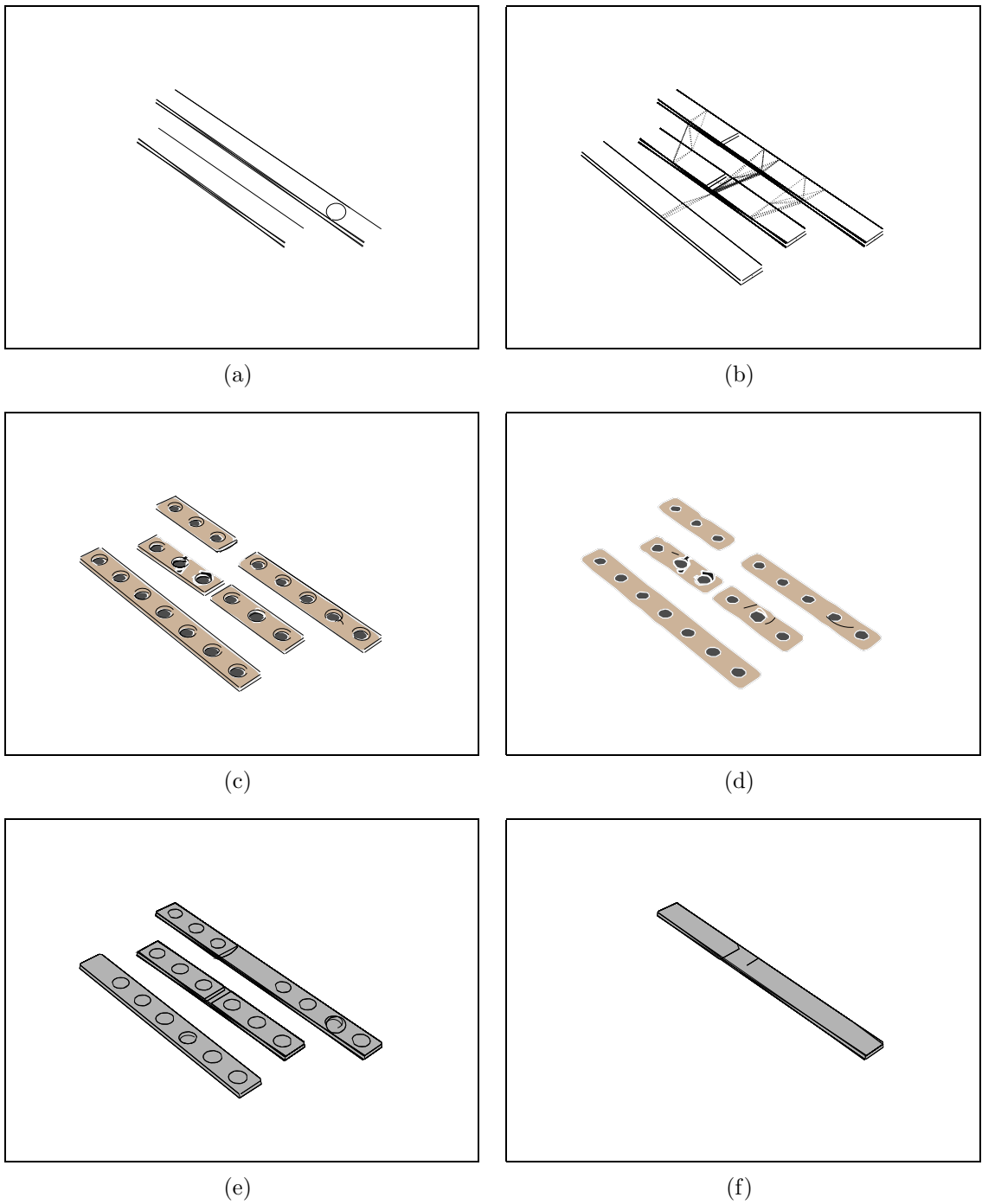
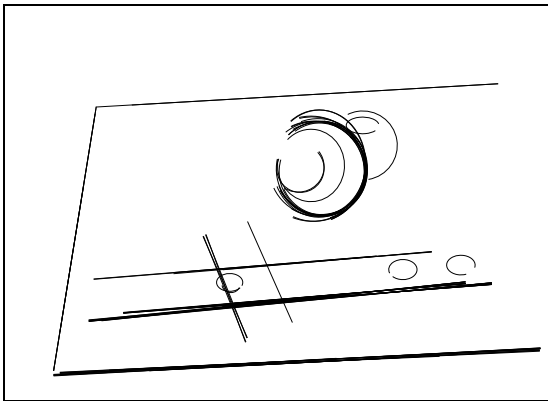
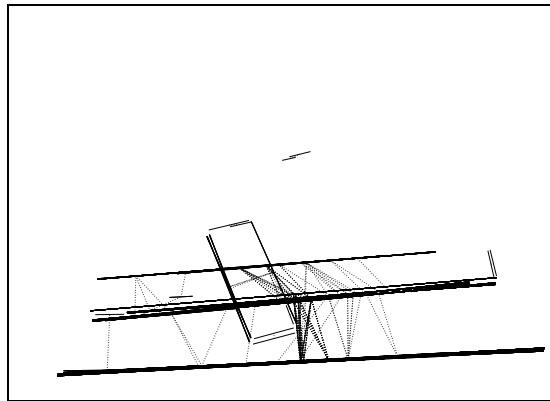


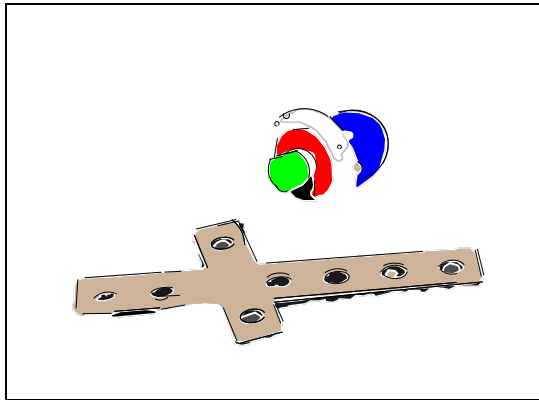
Abbildung B.1.: Bild 1: (a) Kollinearitäten und Kurvilinearitäten, (b) Parallelitätshypothesen, (c) umriss- und (d) strukturbeschreibende Zuordnungen, (e) rein konturbasierte Geschlossenheiten und (f) zusätzliche Geschlossenheiten mit regionenbasierten Proximitäten.



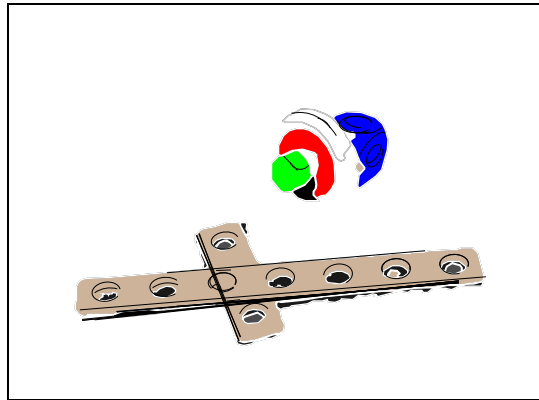
(a)



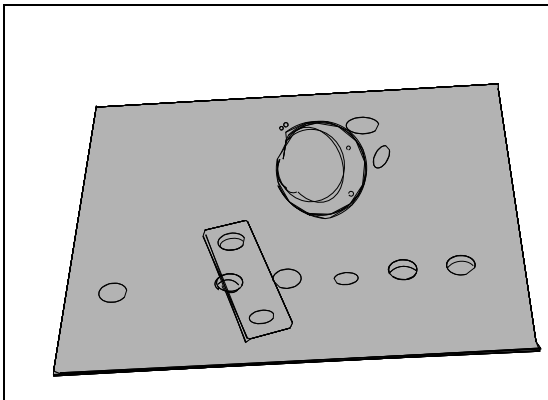
(b)



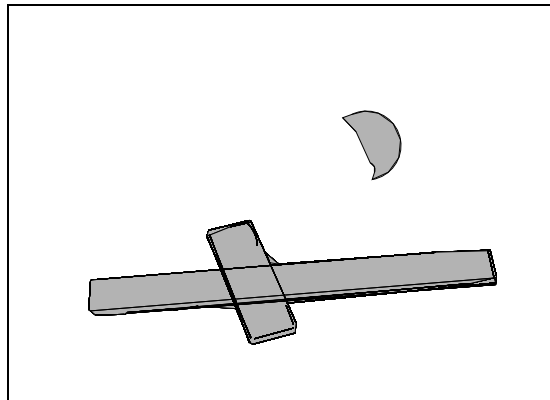
(c)



(d)



(e)



(f)

Abbildung B.2.: Bild 2: (a) Kollinearitäten und Kurvilinearitäten, (b) Parallelitätshypothesen, (c) umriss- und (d) strukturbeschreibende Zuordnungen, (e) rein konturbasierte Geschlossenheiten und (f) zusätzliche Geschlossenheiten mit regionenbasierten Proximitäten.

B. Gruppierungshypothesen und Kombination von Regionen und Konturgruppen

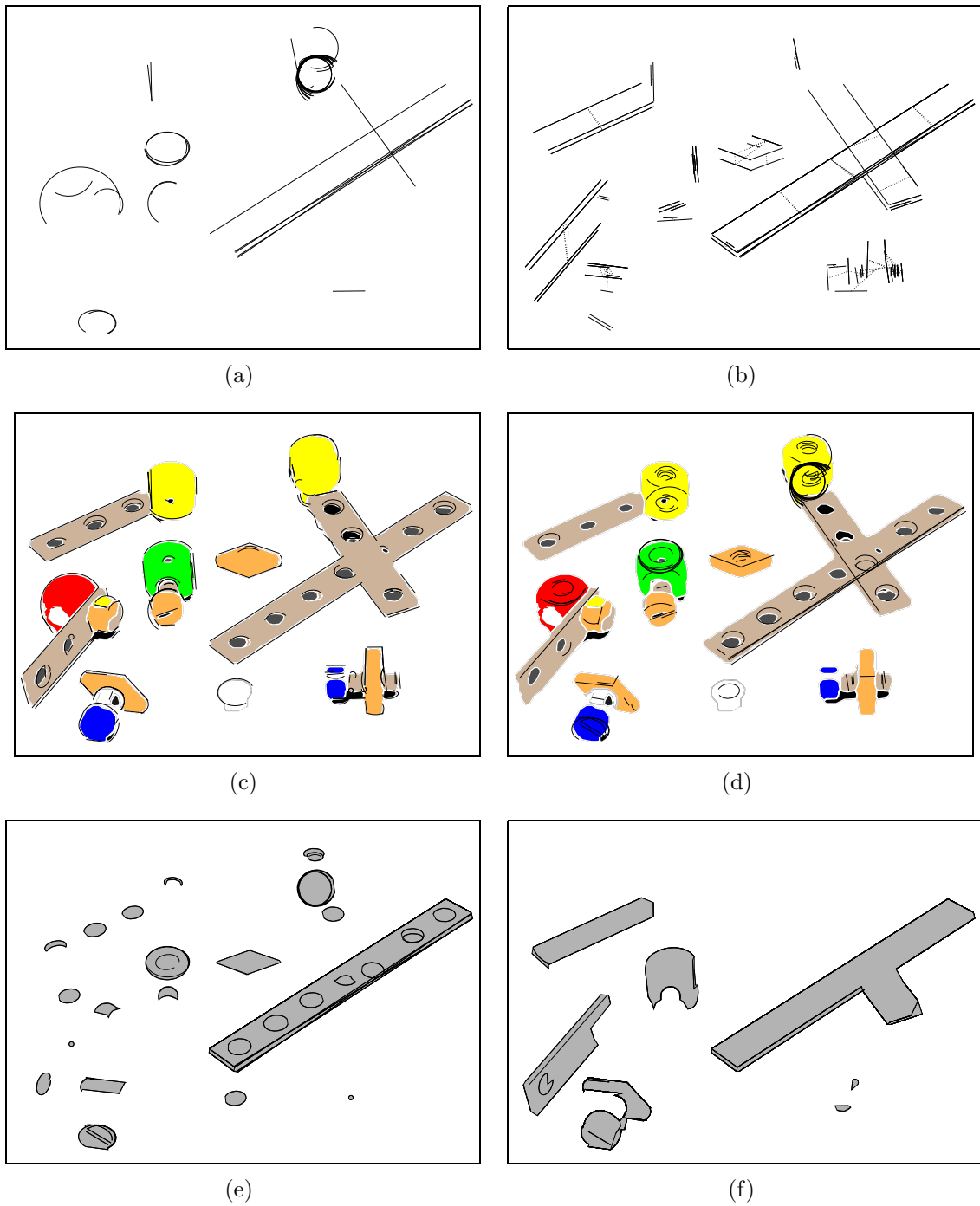
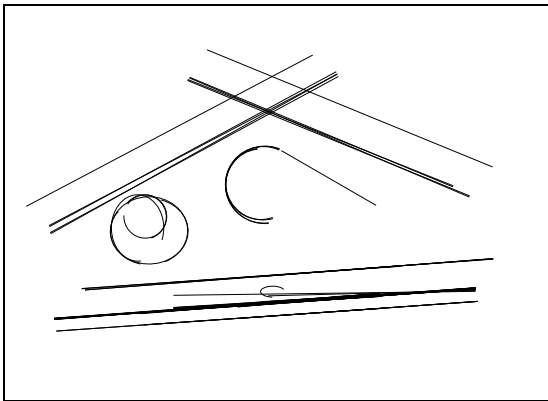
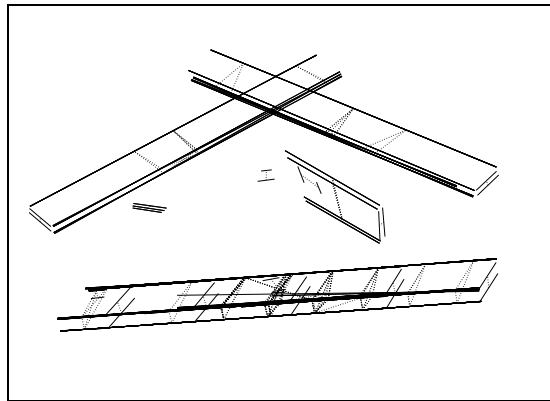


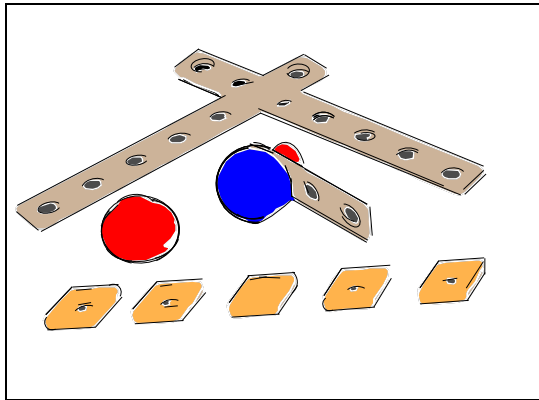
Abbildung B.3.: Bild 3: (a) Kollinearitäten und Kurvilinearitäten, (b) Parallelitätshypothesen, (c) umriss- und (d) strukturbeschreibende Zuordnungen, (e) rein konturbasierte Geschlossenheiten und (f) zusätzliche Geschlossenheiten mit regionenbasierten Proximitäten.



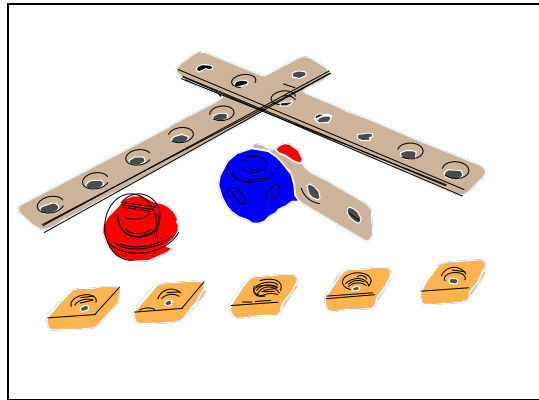
(a)



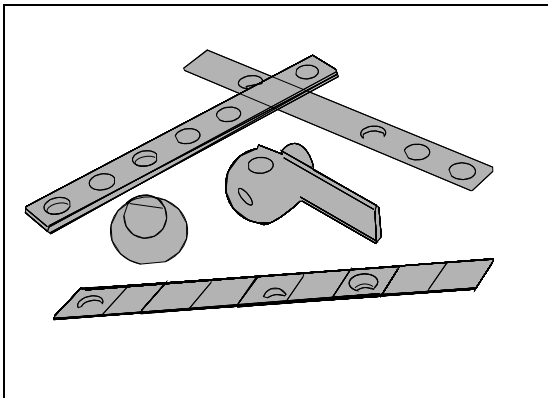
(b)



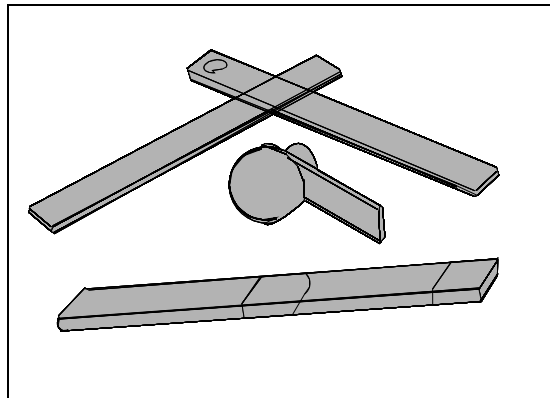
(c)



(d)



(e)



(f)

Abbildung B.4.: Bild 4: (a) Kollinearitäten und Kurvilinearitäten, (b) Parallelitätshypothesen, (c) umriss- und (d) strukturbeschreibende Zuordnungen, (e) rein konturbasierte Geschlossenheiten und (f) zusätzliche Geschlossenheiten mit regionenbasierten Proximitäten.

B. Gruppierungshypothesen und Kombination von Regionen und Konturgruppen

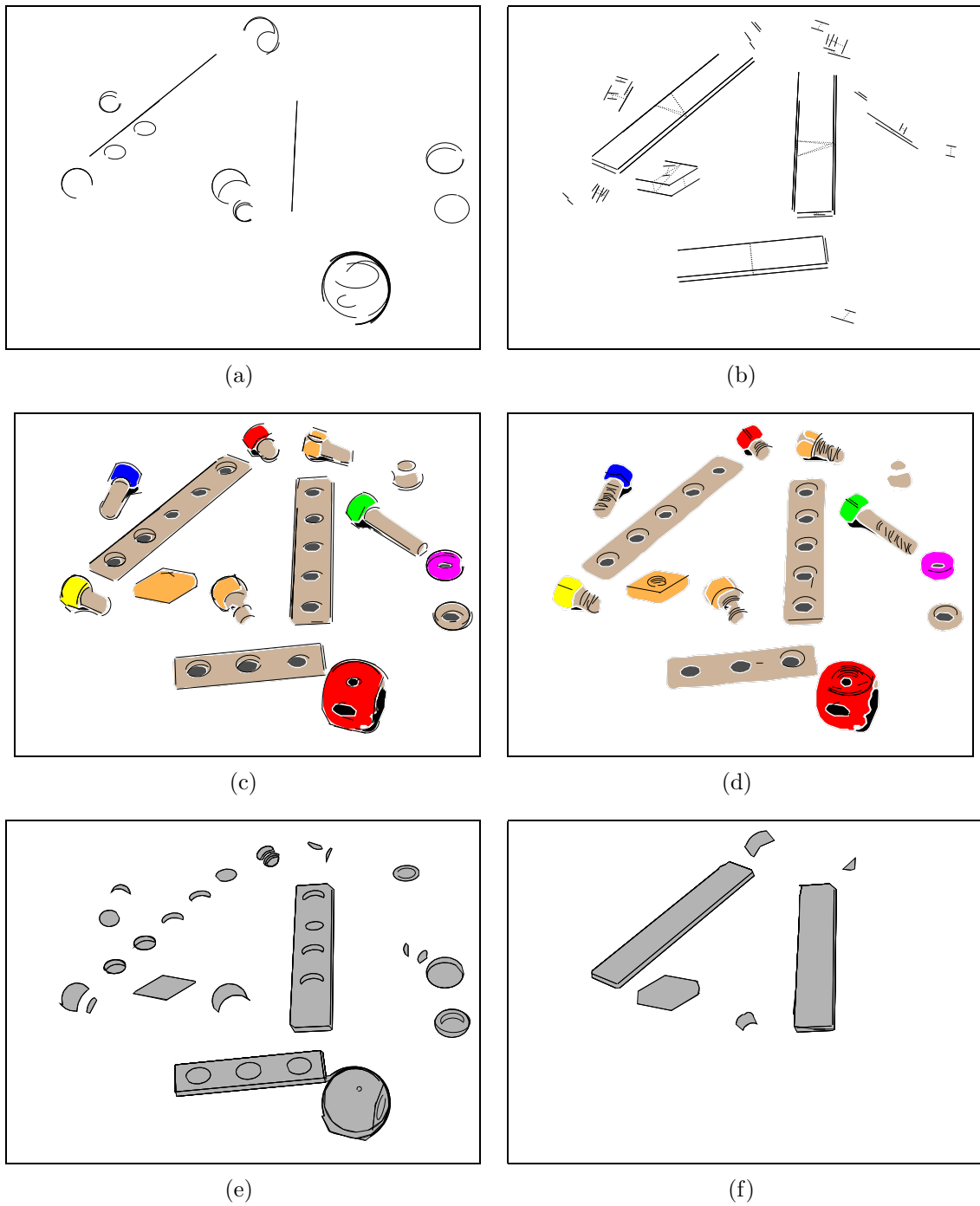
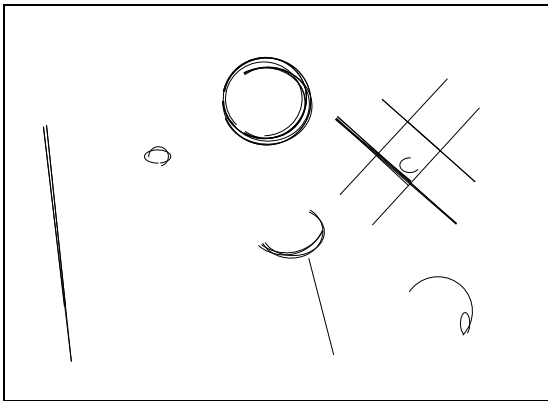
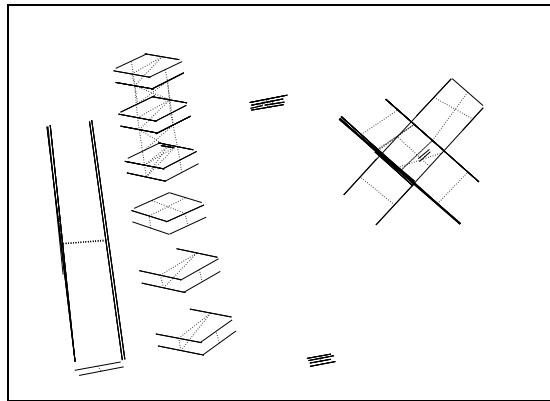


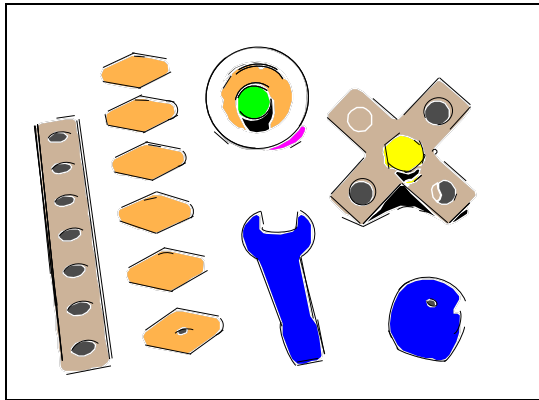
Abbildung B.5.: Bild 5: (a) Kollinearitäten und Kurvilinearitäten, (b) Parallelitätshypothesen, (c) umriss- und (d) strukturbeschreibende Zuordnungen, (e) rein konturbasierte Geschlossenheiten und (f) zusätzliche Geschlossenheiten mit regionenbasierten Proximitäten.



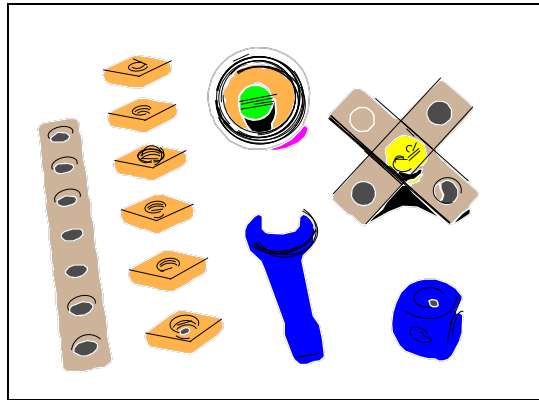
(a)



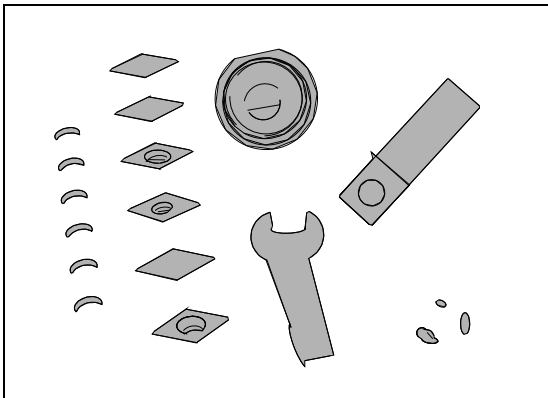
(b)



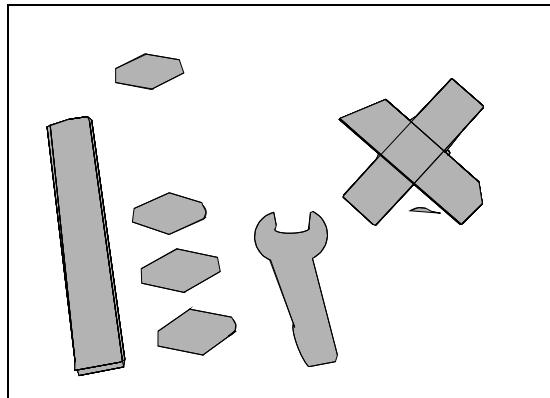
(c)



(d)



(e)



(f)

Abbildung B.6.: Bild 6: (a) Kollinearitäten und Kurvilinearitäten, (b) Parallelitätshypothesen, (c) umriss- und (d) strukturbeschreibende Zuordnungen, (e) rein konturbasierte Geschlossenheiten und (f) zusätzliche Geschlossenheiten mit regionbasierten Proximitäten.

B. Gruppierungshypothesen und Kombination von Regionen und Konturgruppen

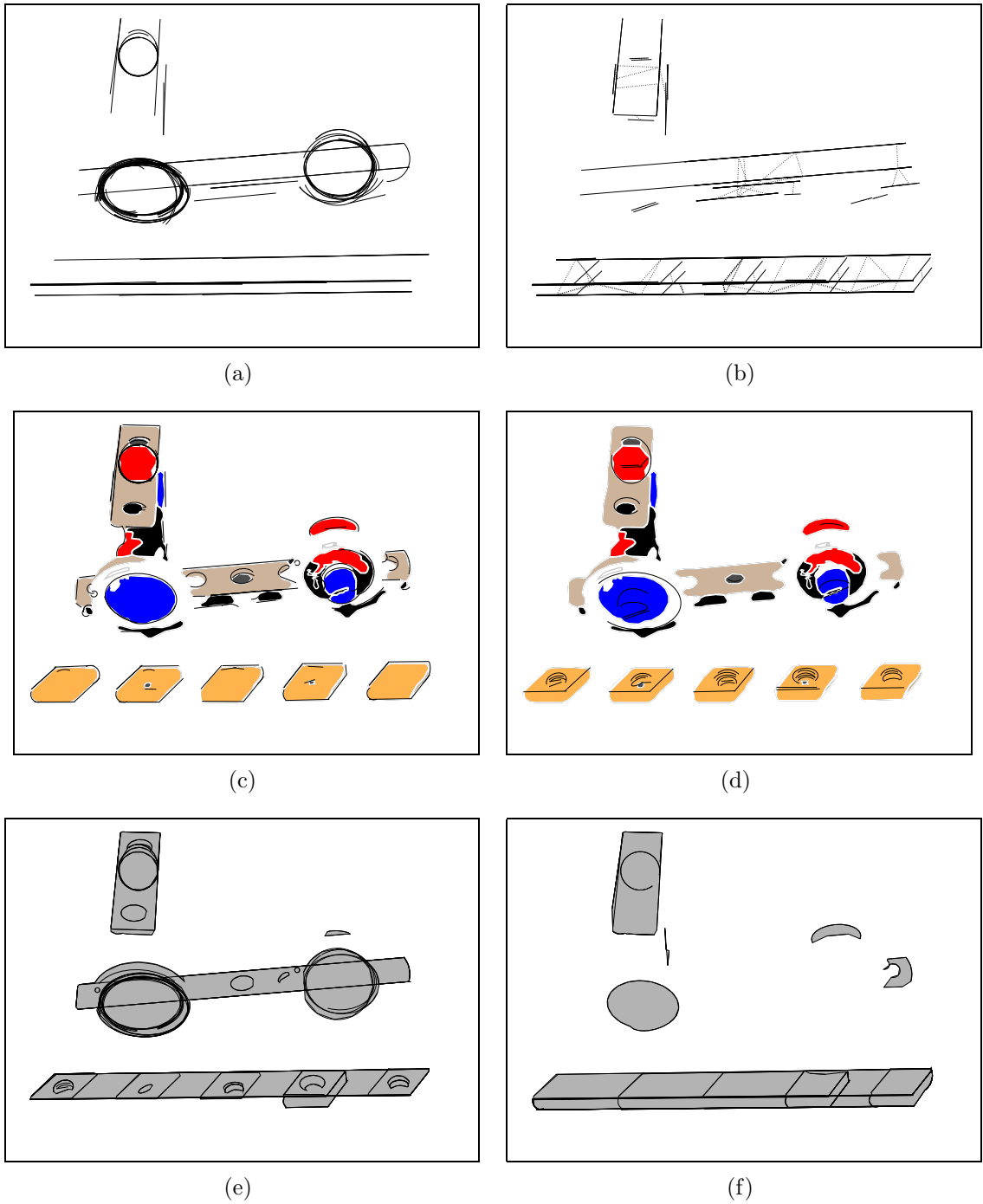
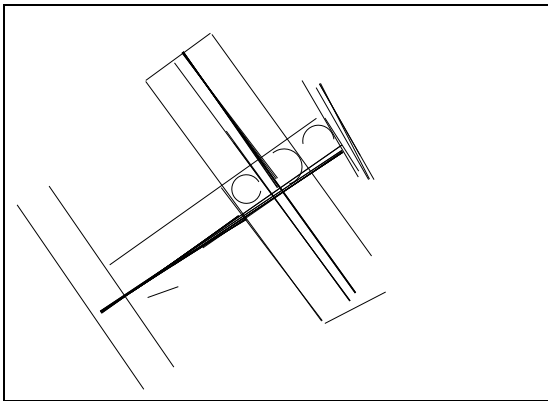
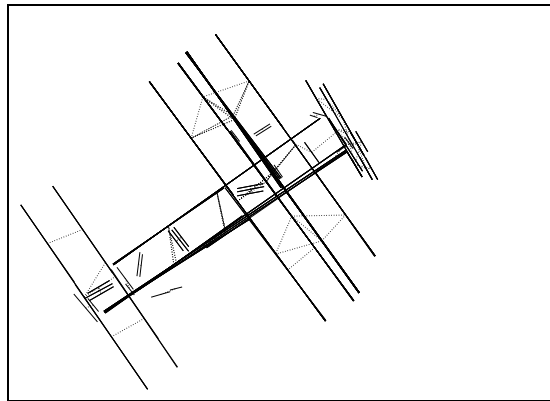


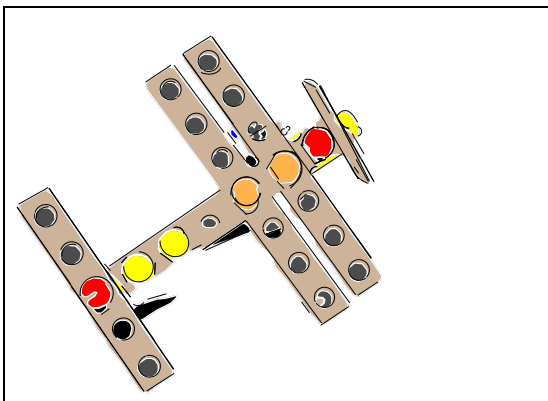
Abbildung B.7.: Bild 7: (a) Kollinearitäten und Kurvilinearitäten, (b) Parallelitätshypothesen, (c) umriss- und (d) strukturbeschreibende Zuordnungen, (e) rein konturbasierte Geschlossenheiten und (f) zusätzliche Geschlossenheiten mit regionenbasierten Proximitäten.



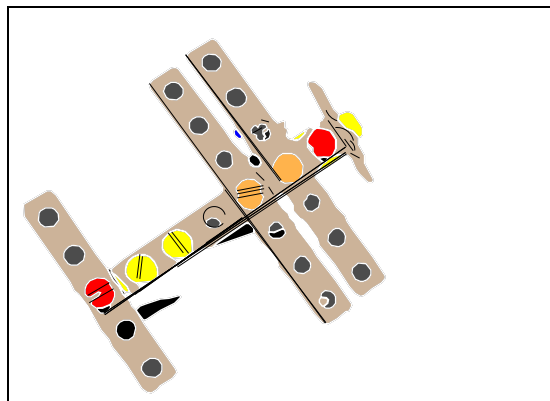
(a)



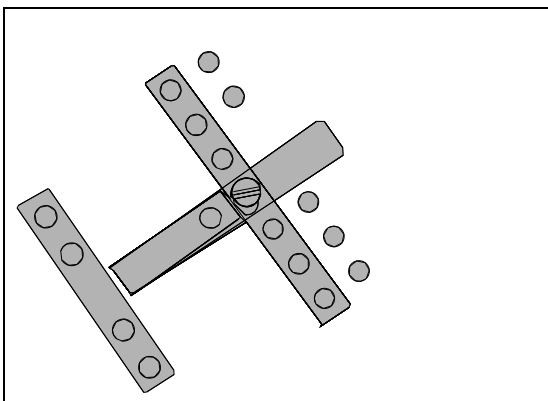
(b)



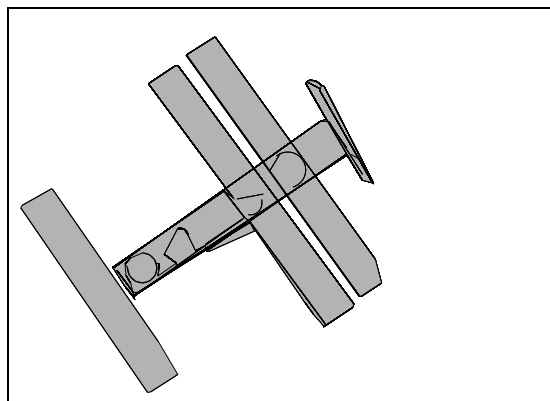
(c)



(d)



(e)



(f)

Abbildung B.8.: Bild 7: (a) Kollinearitäten und Kurvilinearitäten, (b) Parallelitätshypothesen, (c) umriss- und (d) strukturbeschreibende Zuordnungen, (e) rein konturbasierte Geschlossenheiten und (f) zusätzliche Geschlossenheiten mit regionenbasierten Proximitäten.

B. Gruppierungshypothesen und Kombination von Regionen und Konturgruppen

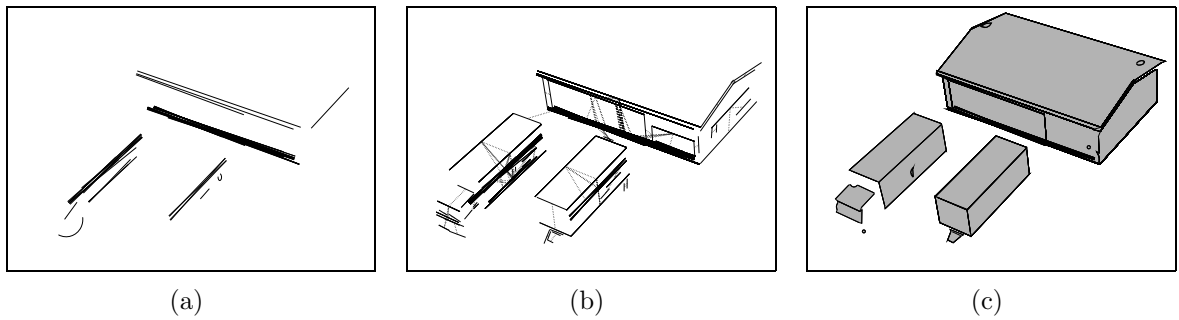


Abbildung B.9.: Bild 9: (a) Kollinearitäten und Kurvilinearitäten, (b) Parallelitätshypothesen und (c) konturbasierte Geschlossenheiten.

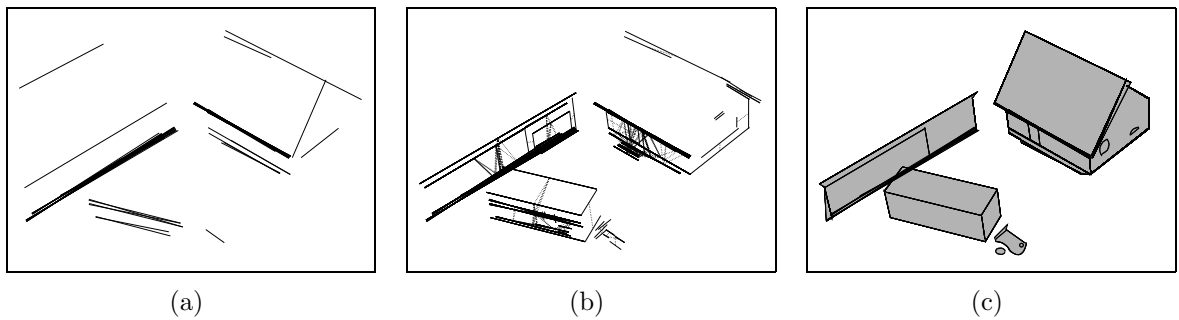
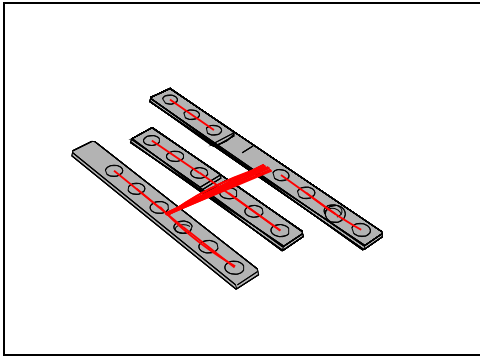
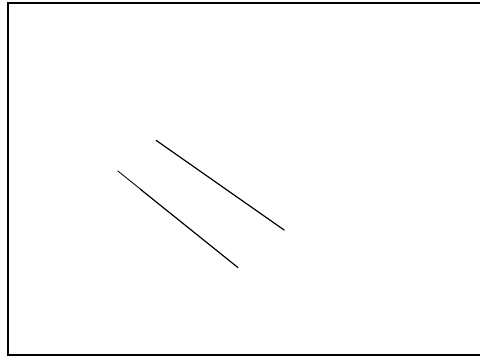


Abbildung B.10.: Bild 10: (a) lineare Gruppen, (b) Parallelitätshypothesen und (c) konturbasierte Geschlossenheiten.

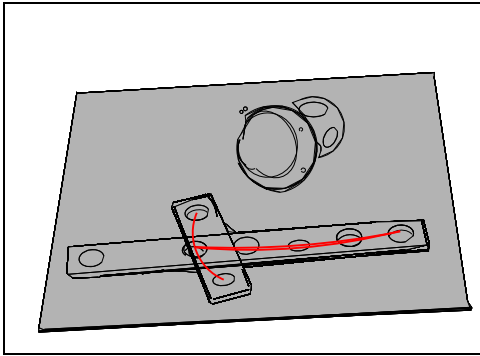


(a)

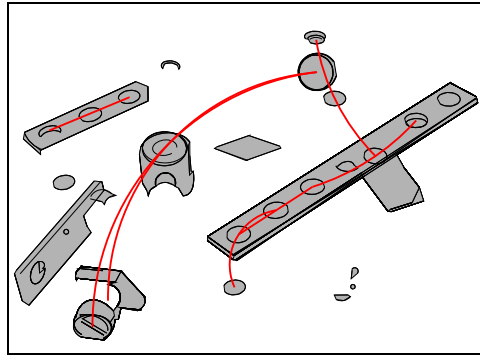


(b)

Abbildung B.11.: Bild 1: (a) Gruppierungen der $\mathbf{N} \times 2\mathbf{D}$ -Ebene, (b) rekursiv gruppierte Konturmodelle aus (a) in der $\mathbf{1D}$ -Ebene.

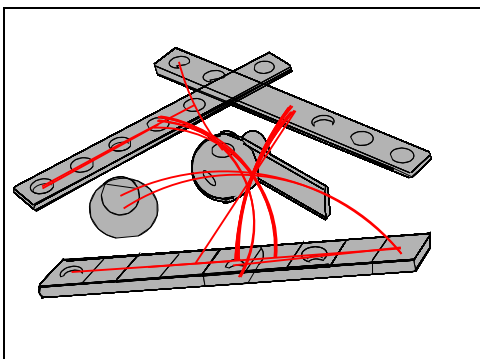


(a)

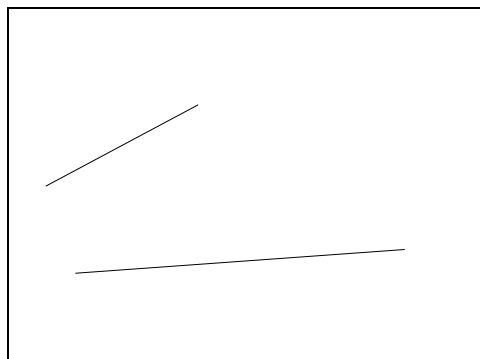


(b)

Abbildung B.12.: Gruppierungen der $\mathbf{N} \times 2\mathbf{D}$ -Ebene: (a) Bild 2, (b) Bild 3.



(a)



(b)

Abbildung B.13.: Bild 4: (a) Gruppierungen der $\mathbf{N} \times 2\mathbf{D}$ -Ebene, (b) rekursiv gruppierte Konturmodelle aus (a) in der $\mathbf{1D}$ -Ebene.

B. Gruppierungshypothesen und Kombination von Regionen und Konturgruppen

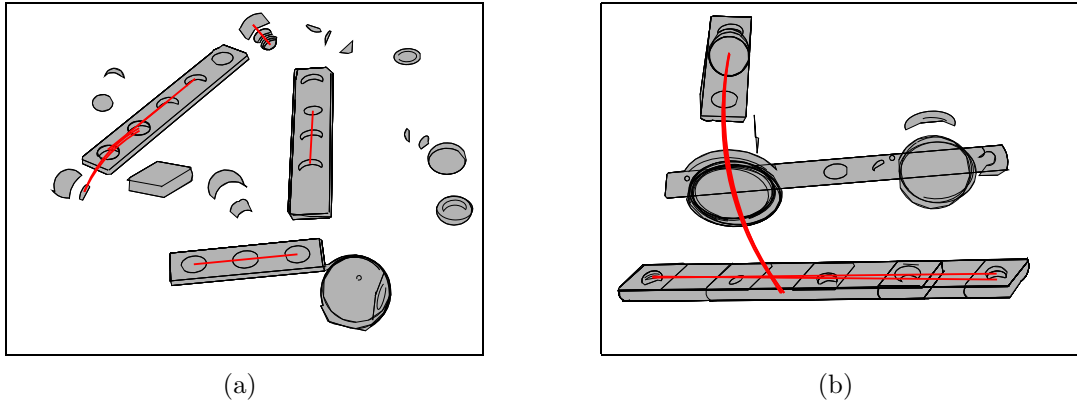


Abbildung B.14.: Gruppierungen der $\mathbf{N} \times 2\mathbf{D}$ -Ebene: (a) Bild 5, (b) Bild 7.

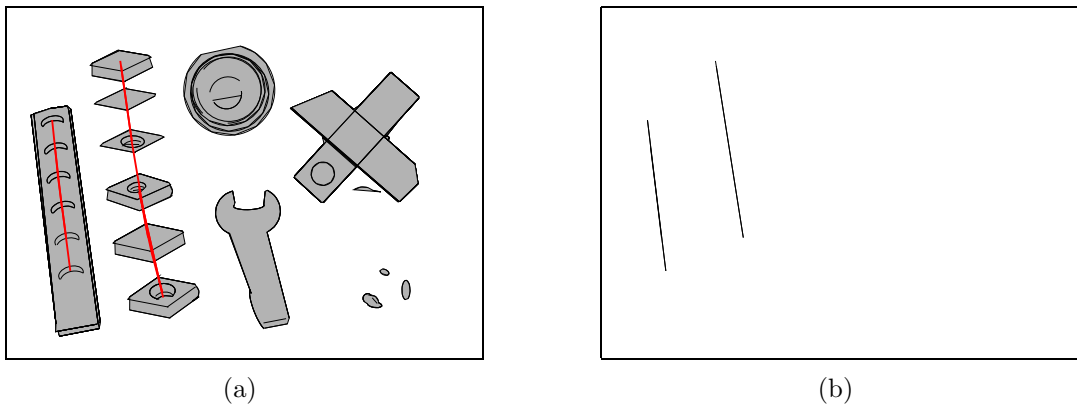


Abbildung B.15.: Bild 6: (a) Gruppierungen der $\mathbf{N} \times 2\mathbf{D}$ -Ebene, (b) rekursiv gruppierte Konturmodelle aus (a) in der $\mathbf{1D}$ -Ebene.

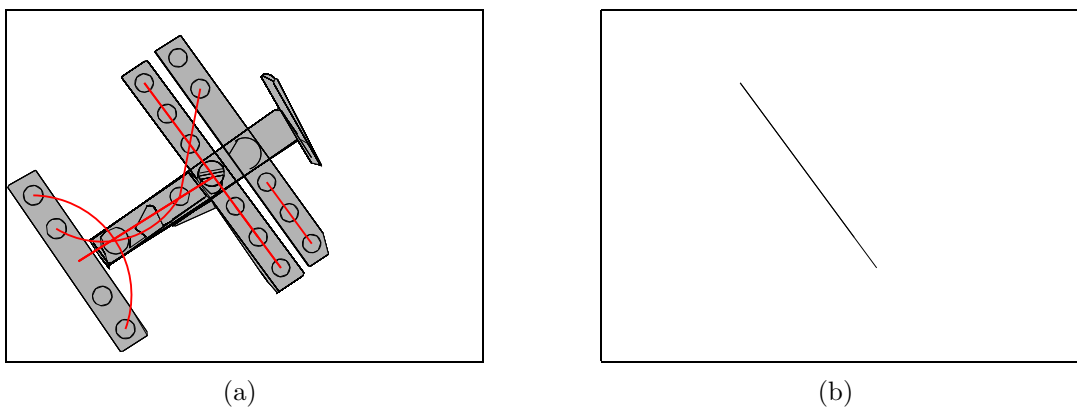
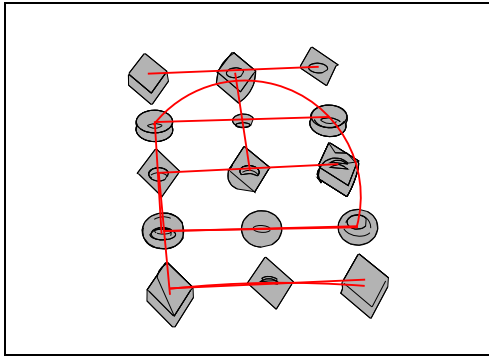
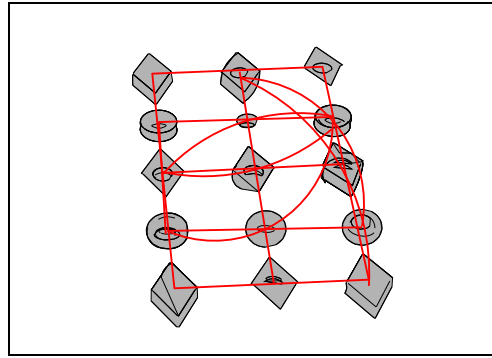


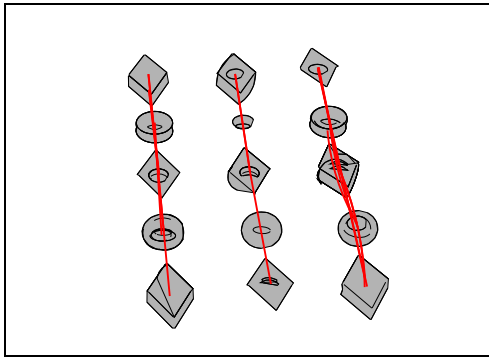
Abbildung B.16.: Bild 8: (a) Gruppierungen der $\mathbf{N} \times 2\mathbf{D}$ -Ebene, (b) rekursiv gruppierte Konturmodelle aus (a) in der $\mathbf{1D}$ -Ebene.



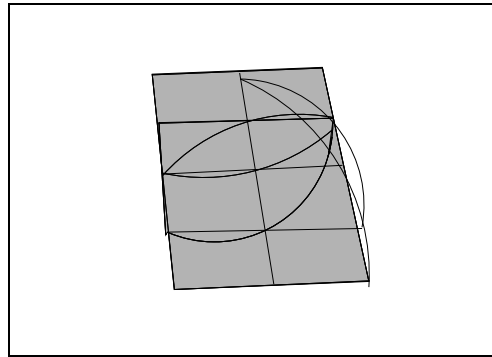
(a)



(b)

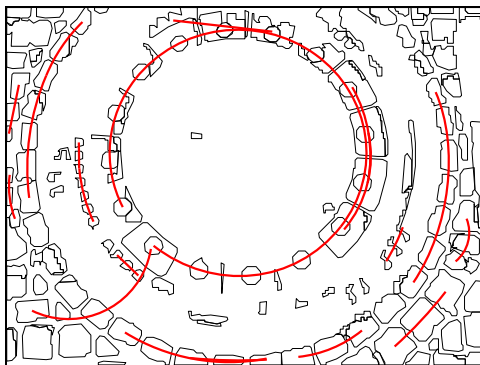


(c)

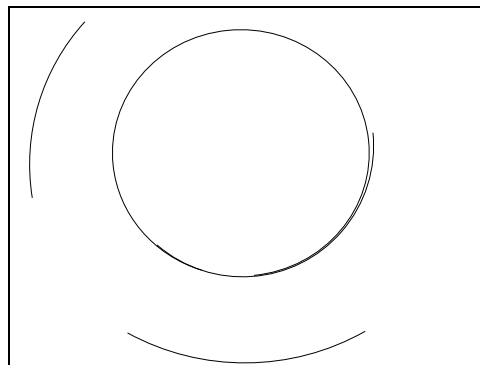


(d)

Abbildung B.17.: Bild 11: (a) Gruppierungen der $\mathbf{N} \times \mathbf{2D}$ -Ebene, (b) Ähnlichkeit bevorzugende Gruppierung ($\kappa_{sn} = 0.1$), (c) Nähe bevorzugende Gruppierung ($\kappa_{sn} = 0.9$), (d) rekursive $\mathbf{2D}$ -Gruppierung aus (b).



(a)



(b)

Abbildung B.18.: Bild 12: (a) Gruppierungen der $\mathbf{N} \times \mathbf{2D}$ -Ebene, (b) rekursiv gruppierte Konturmodelle aus (a) in der $\mathbf{1D}$ -Ebene.

B. Gruppierungshypothesen und Kombination von Regionen und Konturgruppen

C. Signifikante Gruppierungshypothesen

In diesem Anhang werden die signifikant bewerteten konturbasierten Gruppierungshypothesen für die Bilder der Teststichprobe aus Anhang A dargestellt. Die Darstellung erfolgt analog zu Anhang B, wobei hier nur die signifikant markierten initialen Konturmodelle, Kol- und Kurvilinearitäten, sowie Parallelitäten und Geschlossenheiten abgebildet sind. Initiale Konturmodelle, die zu keiner Gruppierung verwendet wurden, sind nicht dargestellt, da sie keine Bewertung erfahren. Als Signifikanzniveau wurde hier $\omega_{sig} = 0.8$ gewählt, wie in Kapitel 7.2 beschrieben. In Abbildung C.11 ist exemplarisch die Mehrfachinterpretation einer Struktur durch Geschlossenheiten für das Bild 2 dargestellt.

C. Signifikante Gruppierungshypothesen

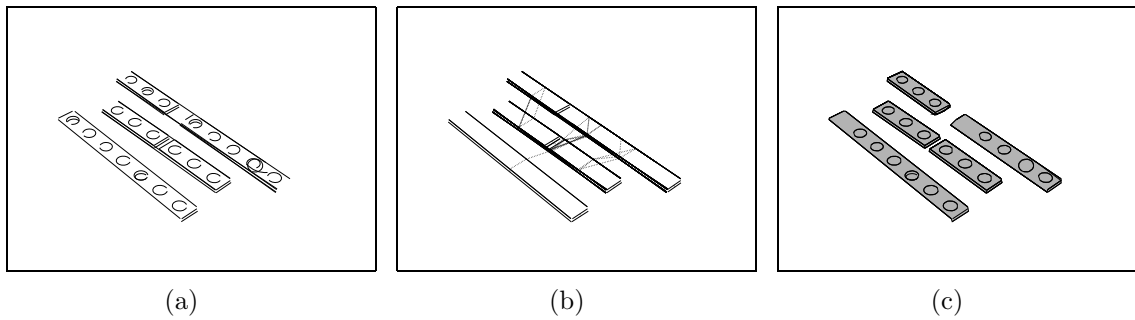


Abbildung C.1.: Bild 1: Alle signifikanten (a) linearen Gruppen, (b) Parallelitäten, (c) Geschlossenheiten.

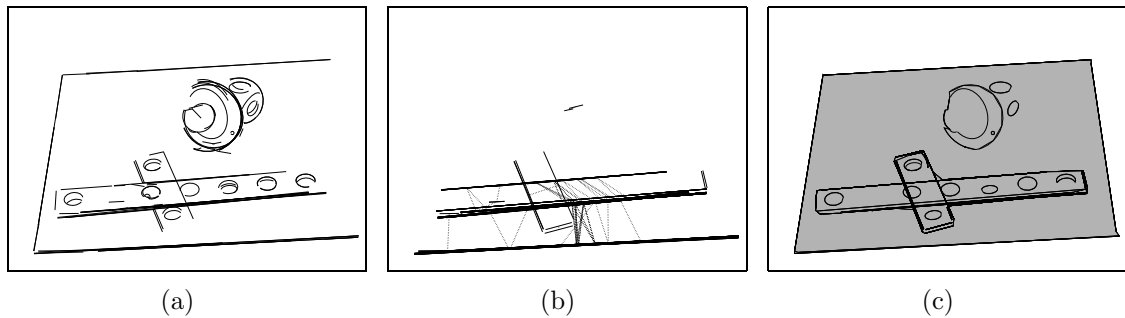


Abbildung C.2.: Bild 2: Alle signifikanten (a) linearen Gruppen, (b) Parallelitäten, (c) Geschlossenheiten.

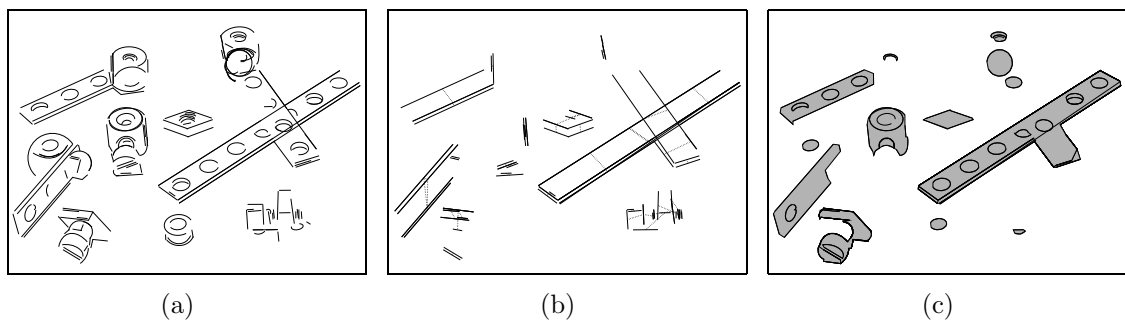


Abbildung C.3.: Bild 3: Alle signifikanten (a) linearen Gruppen, (b) Parallelitäten, (c) Geschlossenheiten.

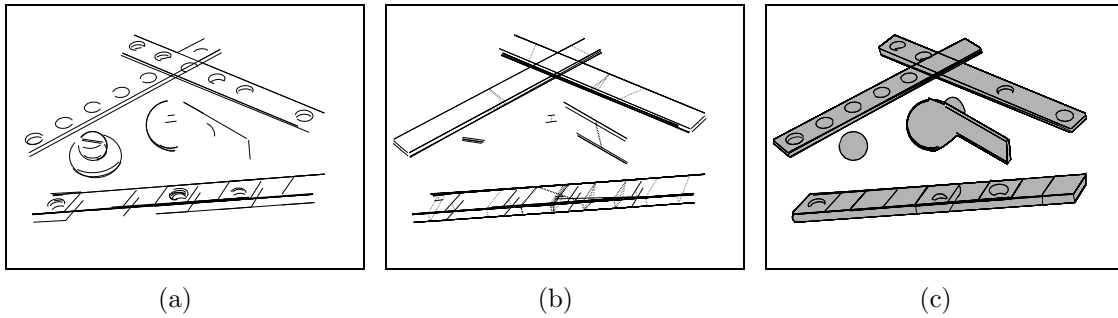


Abbildung C.4.: Bild 4: Alle signifikanten (a) linearen Gruppen, (b) Parallelitäten, (c) Geschlossenheiten.

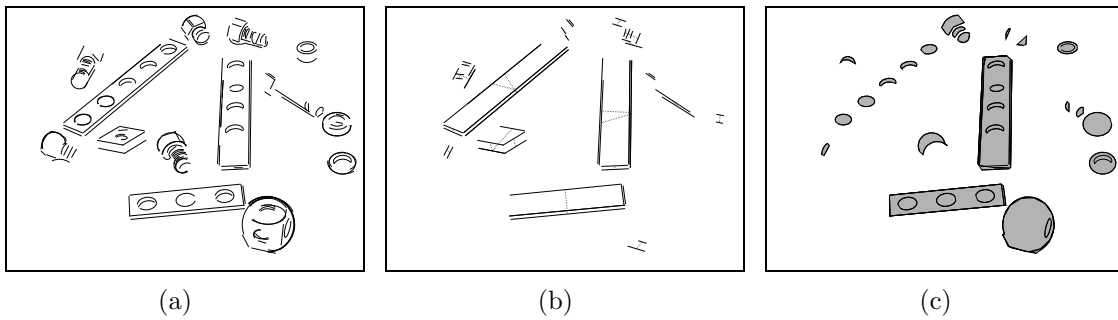


Abbildung C.5.: Bild 5: Alle signifikanten (a) linearen Gruppen, (b) Parallelitäten, (c) Geschlossenheiten.

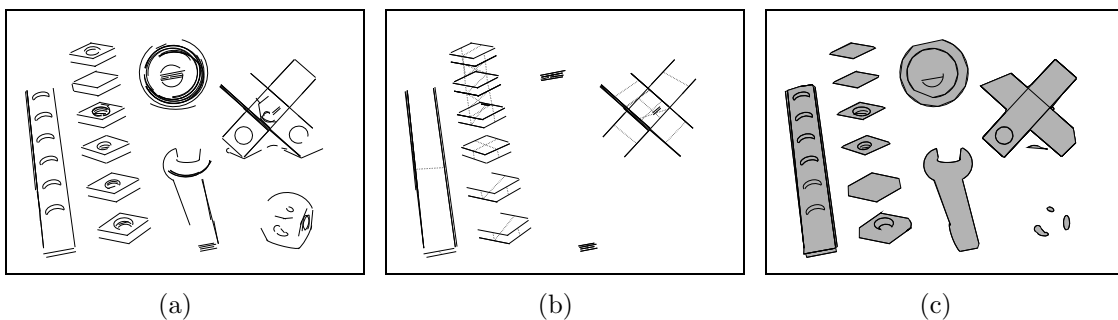


Abbildung C.6.: Bild 6: Alle signifikanten (a) linearen Gruppen, (b) Parallelitäten, (c) Geschlossenheiten.

C. Signifikante Gruppierungshypothesen

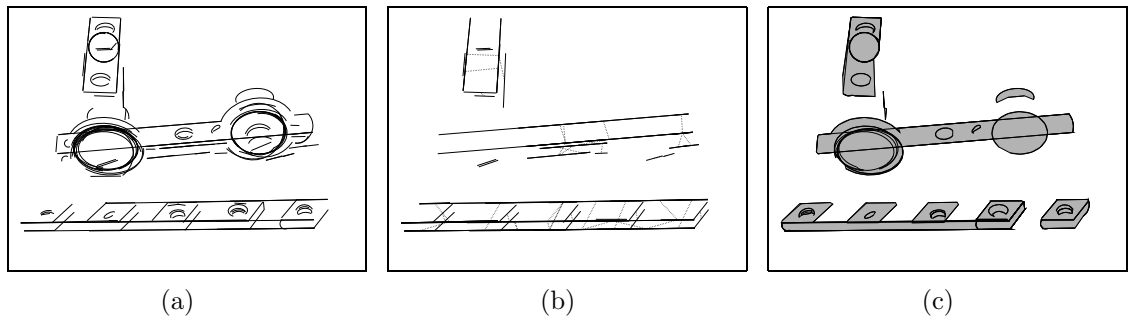


Abbildung C.7.: Bild 7: Alle signifikanten (a) linearen Gruppen, (b) Parallelitäten, (c) Geschlossenheiten.

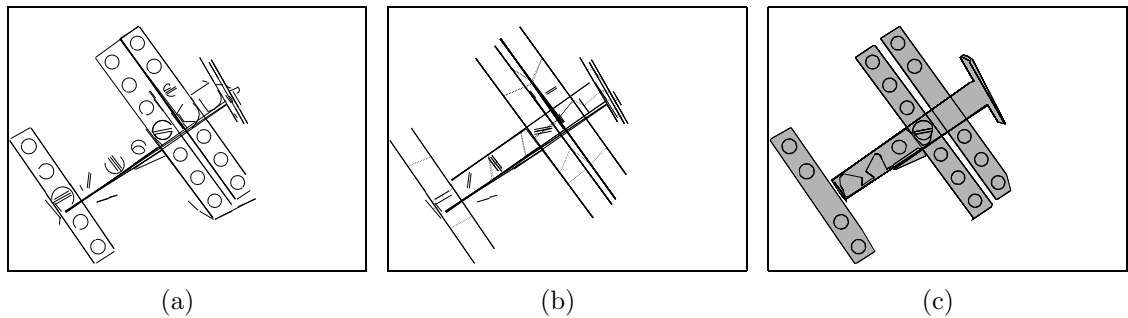


Abbildung C.8.: Bild 8: Alle signifikanten (a) linearen Gruppen, (b) Parallelitäten, (c) Geschlossenheiten.

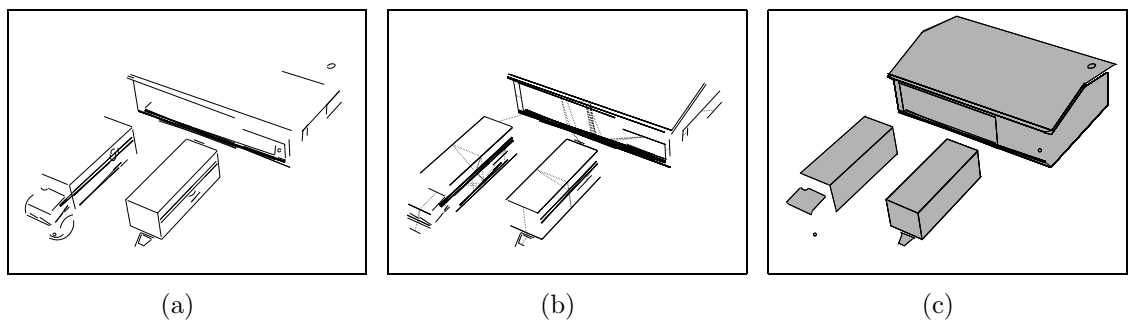


Abbildung C.9.: Bild 9: Alle signifikanten (a) linearen Gruppen, (b) Parallelitäten, (c) Geschlossenheiten.

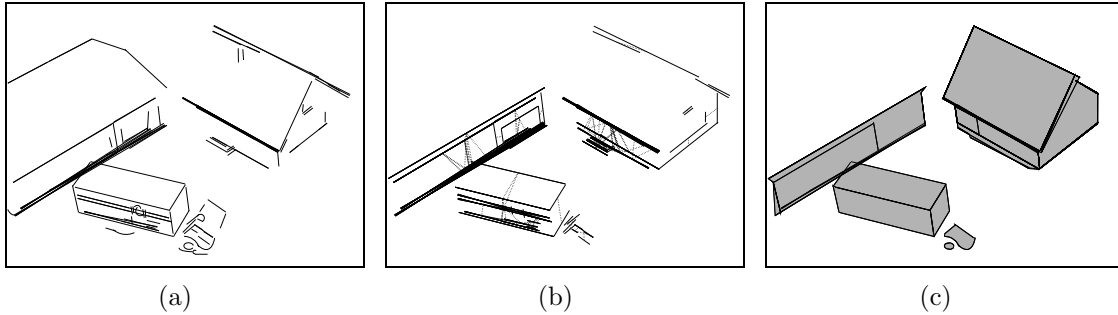


Abbildung C.10.: Bild 10: Alle signifikanten (a) linearen Gruppen, (b) Parallelitäten, (c) Geschlossenheiten.

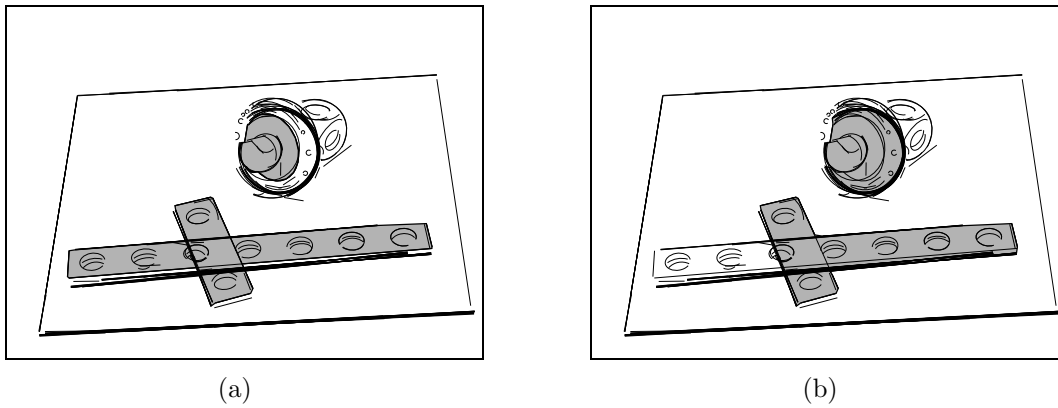


Abbildung C.11.: Bild 2: Signifikante Geschlossenheiten, die eine Mehrfachinterpretation der Leistenprojektionen aus Bild 2 vornehmen. In (a) sind jeweils Geschlossenheiten für die einzelnen Leisten dargestellt, während in (b) eine Verschmelzung beider Strukturen gezeigt ist. Auch die verschiedenen Strukturen aus Felge inklusive Schraubkopf und dem äußeren Reifen sind hier exemplarisch dargestellt.

C. Signifikante Gruppierungshypothesen

D. Integrierte Objekterkennung

In diesem Anhang werden die Ergebnisse einer integrierten Objekterkennung für die Bilder 6 und 8 dargestellt. Die interpretierten Farbregionen sind durch ihren Umriß in der jeweiligen RGB-Farbe dargestellt. Die ermittelten Objekte sind durch entsprechende Begriffe angezeigt. Für die Integration liegt der Focus auf den als *unbekannt* klassifizierten Objekten. Die konturbasierten Gruppierungen werden nur in diesen Bereichen generiert und sind in den Darstellungen entsprechend ihrem Gruppierungstyp unterschiedlich eingefärbt. Geschlossenheitshypothesen sind rot umrandet, während Parallelitäten blau, Kol- und Kurvilinearitäten gelb und die initialen Konturmodelle schwarz eingezeichnet sind.

D. Integrierte Objekterkennung

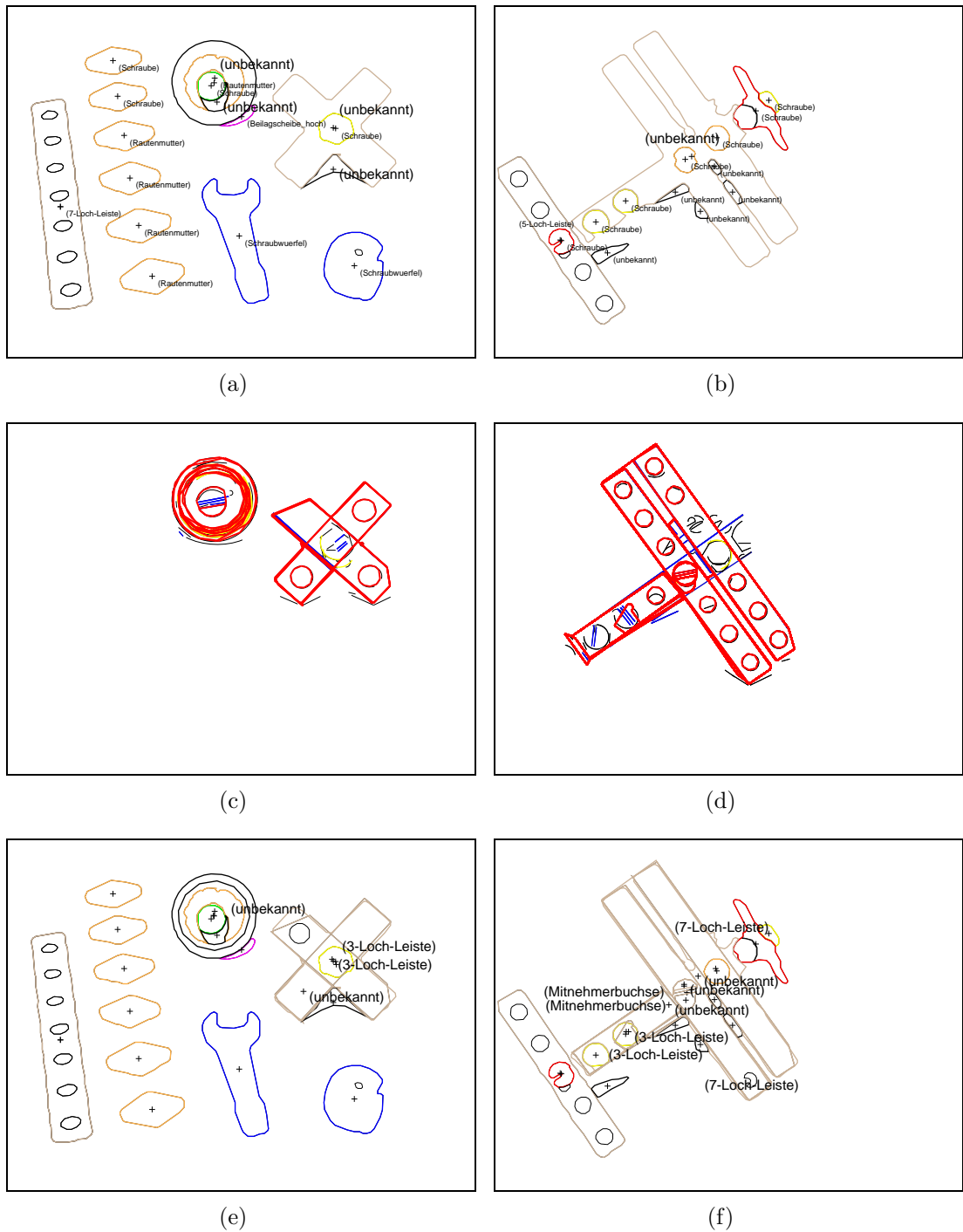


Abbildung D.1.: Integrierte Objekterkennung für die Bilder 6 und 8 . (a),(b) Regionbasierte Objekterkennung mit unbekannt klassifizierten Regionen. (c),(d) Konturbasierte Gruppierungen für unbekannt klassifizierte Bildregionen. (e),(f) Zusätzlich generierte Objekthypothesen aus signifikanten konturbasierten Geschlossenheiten.