

Institut für Umweltwissenschaften der Hochschule Vechta

---

# Identifikation und Rekonstruktion von Gebäuden in Luftbildern mittels unscharfer Constraints

**Inauguraldissertation**

zur

Erlangung des Grades

Doktor der Naturwissenschaften (Dr. rer. nat.),

angenommen vom Fachbereich 1

der Hochschule Vechta

Vorgelegt am 31. 3. 1999 von

**Thomas Heinrich Kolbe**

aus Recklinghausen

---

Referent: Prof. Dr. rer. nat. Lutz Plümer  
Institut für Kartographie und Geoinformation, Universität Bonn

Korreferenten: Prof. Dr. rer. nat. Armin B. Cremers  
Institut für Informatik III, Universität Bonn  
Prof. Dr.-Ing. Manfred Ehlers  
Institut für Umweltwissenschaften, Hochschule Vechta

Tag der mündlichen Prüfung: 17. 5. 1999

## Geleitwort

Daten bilden in Geoinformationssystemen den größten Kostenfaktor. Die Bereitstellung aktueller Gebäudedaten für dreidimensionale Stadtmodelle ist ökonomisch nur realisierbar, wenn der Prozeß der Datenerfassung durch automatische Werkzeuge unterstützt wird. Angesichts der Vielfalt der Gebäudemodelle und der Schwierigkeit der Erkennung komplexer Objekte in Bildern ist die automatische Identifikation und Rekonstruktion von Gebäuden eine technologische Herausforderung. Vieles spricht dafür, daß sie nur zu bewältigen ist, wenn die Semantik der zu erkennenden Objekte explizit repräsentiert und für Algorithmen verfügbar ist. Diese These bildet den Ausgangspunkt der vorliegenden Arbeit. Sie bildete auch den gemeinsamen Ansatz eines großen Forschungsverbundes von 14 Arbeitsgruppen aus Instituten der Photogrammetrie, Kartographie, Geodäsie, Nachrichtentechnik, Umweltwissenschaften und der Informatik, der von der Deutschen Forschungsgemeinschaft unter dem Titel „Semantische Modellierung“ über 6 Jahre gefördert wurde und in dem diese Arbeit entstanden ist.

Die automatische Erkennung von Gebäuden ist auf Werkzeuge angewiesen, die die explizite, verfahrensunabhängige Repräsentation der Semantik unterstützen und die gleichzeitig die Implementierung effizienter Algorithmen ermöglichen. Nur so ist es möglich, die Schwierigkeiten zu meistern, die aus der Vielfalt der Gebäude in der Realität und der Fragmentierung, Unschärfe und Unvollständigkeit ihrer Repräsentation im Bild resultieren.

Im Sinne eines deklarativen Vorgehens entnimmt die vorliegende Monographie ihre Verfahren der constraint-basierten logischen Programmierung, einer Technologie, mit der ich mich selbst über viele Jahre theoretisch und praktisch auseinandergesetzt habe, bevor ich mich den Geoinformationssystemen verschrieben habe. In den ersten Kapiteln wird eindrucksvoll dargelegt, wie elegant sich im Formalismus der logischen Programmierung aus den Invarianten von Beobachtungen Modelle von Gebäuden konstruieren lassen, und wie effizient constraint-basierte Verfahren das in diesen Modellen enthaltene Wissen nutzen können. Wichtige, aus dem Gebäudemodell resultierende Invarianten sind die Parallelität von Kanten, die Kollinearität von Linien und Knoten und die Adjazenz von Linien und Flächen. Die im Bild beobachteten Merkmale halten diese Invarianten aber in der Regel nicht exakt, sondern nur annähernd ein. Gebäudeteile können infolge von Verdeckungen oft nicht direkt beobachtet werden, sondern müssen indirekt erschlossen werden. Das Problem ist also nicht, eine konsistente Zuordnung von Beobachtungen und Modell zu finden, sondern eine möglichst gute. Hier stößt die logische Programmierung an ihre Grenzen, und die vorliegende Arbeit entwickelt mehrere substantielle Weiterentwicklungen und Adaptionen an die Erfordernisse der Bildverarbeitung.

Ein weiteres schwieriges Problem resultiert aus der Vielfalt der Gebäudemodelle, die zur Interpretation der Beobachtungen eines Bildes herangezogen werden müssen. Hier geht es um die Bewertung der Zuordnung eines Modells zu den vorgefundenen Beobachtungen, die verschiedene Modelle untereinander vergleichbar macht. Dieser Vergleich wird dadurch erschwert, daß der Abstand zwischen Modell und Beobachtung durch die Zunahme der Komplexität des Modells verringert werden kann — im Gegensatz zu der Maxime, daß einfache Erklärungen besser sind als komplizierte. Zur Lösung dieses Problems wird ein Verfahren entwickelt, welches das Prinzip der „minimalen Beschreibungslänge“ umsetzt. Kernidee ist die Bereitstellung eines minimalen Codes zur Beschreibung von Bildmerkmalen und Gebäudemodellen, durch den die Beschreibungslänge der Differenz zwischen Beobachtung und Modell mit der Beschreibungslänge des Modells in Beziehung gesetzt werden kann. Die vorliegende Arbeit beschreibt das darauf basierende und in lauffähige Software umgesetzte Verfahren im Detail und illustriert an zahlreichen eindrucksvollen Beispielen Techniken zur Komprimierung des verwendeten Codes, welche die für realistische Bilddaten erforderliche Trennschärfe zwischen verschiedenen Modellen hervorbringen.

Diese Dissertation ist ein Stück Grundlagenforschung im Bereich Geoinformationssysteme. Sie trägt zum Verständnis des Problems bei, wie man durch die Verwendung von Modellen und die Formalisierung von Semantik die Erfassung komplexer räumlicher Objekte für Geoinformationssysteme voranbringen kann. Sie löst ein Problem, das in der Photogrammetrie entsteht und mit der

Anwendung und Weiterentwicklung von Methoden und Werkzeugen der Informatik gelöst wird. Sie ist im echten Sinne interdisziplinär. Zahlreiche Beispiele und sorgfältige Illustrationen ebnet dem Leser den Zugang.

Es war für mich eine große Freude, mit Thomas Kolbe in den verschiedenen Phasen seiner (und meiner) akademischen Entwicklung und in verschiedenen Projekten zusammenzuarbeiten. Ausgehend vom Fachbereich Informatik der Universität Dortmund und dem Institut für Informatik an der Universität Bonn in meiner Zeit als wiss. Mitarbeiter bei A. B. Cremers führte sie über das Institut für Umweltwissenschaften an der Hochschule Vechta schließlich zum Institut für Kartographie und Geoinformation an der Universität Bonn, wo Thomas den Aufbau des neu gegründeten Lehrstuhls für Geoinformation und unseres GIS-Labors maßgeblich mitgestaltet. Die wiederverwendbaren Module und Komponenten, die im Zuge seiner Dissertation entstanden sind, bilden eine wichtige Plattform für künftige GIS-Entwicklungen des Instituts.

Lutz Plümer

im Mai 2000

## Danksagung des Autors

Als erstes möchte ich mich bei meinem Mentor und Betreuer dieser Arbeit Prof. Dr. Lutz Plümer bedanken, der mir während der Entstehung dieser Arbeit in vielen ergebnisreichen Diskussionen mit konstruktiver Kritik, Anregungen und Rat zur Seite stand.

Mein weiterer Dank gilt den Korreferenten der Arbeit, Prof. Dr. Armin B. Cremers und Prof. Dr. Manfred Ehlers, an deren Instituten diese Arbeit unter guten Forschungsbedingungen in einer stets angenehmen Atmosphäre entstanden ist.

Diese Arbeit ist weitestgehend aus dem Teilprojekt „Extraktion von Gebäuden aus digitalen Luftbildern“ des DFG-Bündelprojektes „Semantische Modellierung“ hervorgegangen. Ich möchte an dieser Stelle deshalb zunächst der DFG für die Förderung des Projektes danken. Für die fruchtbaren Diskussionen und die gemeinsam entwickelten Konzepte, auf denen diese Arbeit aufbaut, möchte ich mich herzlich bei den Projektleitern Prof. Dr. Lutz Plümer, Prof. Dr. Armin B. Cremers, Prof. Dr. Wolfgang Förstner und Dr. Volker Steinhage sowie meiner Kollegin Felicitas Lang und meinem Kollegen André Fischer bedanken. Ein spezieller Dank geht an André Fischer für die Implementierung und Bereitstellung des Ausgleichsprogramms und des Programms zur Generierung von Gebäudehypothesen und ihrer zweidimensionalen Ansichten, sowie an Felicitas Lang für die Bereitstellung rekonstruierter 3D-Gebäudeecken für die verwendeten Datensätze.

Für hilfreiche Diskussionen, Verbesserungsvorschläge und sorgfältiges Lesen dieser Arbeit danke ich ferner Dr. Jochen Schiewe, Ulrich Rhein, Stephan Heuel, Ingo Petzold und Gerhard Gröger.

Ich möchte mich bei meinen Eltern Hannelore und Heinrich Kolbe bedanken, die mir dies alles ermöglicht haben.

Mein besonderer Dank geht an meine Frau Antje, die mich während der verschiedenen Arbeitsphasen inspiriert, motiviert und immer wieder aufgebaut hat.

# Inhaltsverzeichnis

<b>1</b>	<b>Einleitung</b>	<b>1</b>
1.1	Anwendungen digitaler Gebäudemodelle . . . . .	1
1.2	Gebäudeextraktion aus Luftbildern . . . . .	2
1.3	Automatische Gebäudeextraktion . . . . .	2
1.4	Ziele der Arbeit . . . . .	4
1.5	Gliederung der Arbeit . . . . .	6
<b>2</b>	<b>Automatische Gebäudeextraktion</b>	<b>9</b>
2.1	Komplexität der Daten . . . . .	9
2.2	Modelle und Strategien zur Gebäudeerkennung . . . . .	10
2.2.1	Anforderungen an die Modellierung . . . . .	10
2.2.2	Objekt- und Bildmodellierung . . . . .	11
2.2.3	Strategien zur 3D-Objekterkennung . . . . .	12
2.3	Vergleichbare Forschungsansätze . . . . .	13
2.3.1	Prismatische Polyedermodelle . . . . .	14
2.3.2	Allgemeine Polyedermodelle . . . . .	15
2.3.3	Parametrisierte Volumenprimitive . . . . .	16
2.3.4	Diskussion der Konzepte . . . . .	17
2.4	Verwendete Modellierung . . . . .	18
2.4.1	Generisches, parametrisiertes Gebäudemodell . . . . .	19
2.4.2	Verbindung von Objekt- und Bildmodell . . . . .	21
2.4.3	Akquisition der Gebäudemodelle . . . . .	22
2.5	Strategie zur Gebäudeextraktion . . . . .	22
2.5.1	Bildsegmentierung und Merkmalsextraktion . . . . .	23
2.5.2	Fokussierung auf einzelne Bildbereiche . . . . .	24
2.5.3	3D-Rekonstruktion von Gebäudeecken . . . . .	25
2.5.4	Generierung von Gebäudehypothesen . . . . .	26
2.5.5	Berechnung der Gebäudeansichten . . . . .	27
2.5.6	Identifikation der richtigen Gebäudeansicht . . . . .	28
2.5.7	Schätzung der Gebäudeparameter . . . . .	30

<b>3</b>	<b>Effiziente Objektidentifikation mit Constraint-Techniken</b>	<b>35</b>
3.1	Constraint Satisfaction Probleme . . . . .	36
3.1.1	Das Subgraph-Isomorphismus-Problem als CSP . . . . .	37
3.1.2	Globale Lösungsverfahren für das Consistent Labeling Problem . . . . .	38
3.1.3	Lokale Lösungsverfahren: Konsistenztechniken . . . . .	40
3.2	Constraint Logic Programming . . . . .	41
3.2.1	Logische Programmierung und PROLOG . . . . .	41
3.2.2	Von LP zu CLP . . . . .	41
3.2.3	CLP auf endlichen Wertebereichen . . . . .	43
3.3	Modellierung von Gebäudeansichten durch Constraint-Mengen . . . . .	43
3.3.1	Geometrische und topologische Constraints . . . . .	44
3.3.2	Bestimmung der Schwellwerte . . . . .	51
3.3.3	Eigenschaften der Constraint-Graphen . . . . .	53
3.4	Realisierung des Konzepts in CLP . . . . .	54
3.4.1	Fakten, Regeln und Constraints . . . . .	54
3.4.2	Transformation von Gebäudeansichten in CLP-Klauseln . . . . .	55
3.4.3	Praktische Evaluation . . . . .	56
3.5	Gruppierung fragmentierter Liniensegmente . . . . .	59
3.5.1	Gruppierungskonzepte im Kontext von Zuordnungsverfahren . . . . .	60
3.5.2	Geometrische Gruppierung . . . . .	62
3.5.3	Einbettung der Gruppierung in das CSP . . . . .	64
<b>4</b>	<b>Erweiterung der Constraint-Techniken um Unbeobachtbarkeit und Unschärfe</b>	<b>67</b>
4.1	Fehlerfälle . . . . .	68
4.2	Überbeschränkte Constraint-Probleme . . . . .	71
4.2.1	MaxCSP . . . . .	72
4.2.2	HCLP . . . . .	72
4.2.3	Dynamic CSP . . . . .	73
4.2.4	Die generischen Konzepte PCSP und GOCS . . . . .	73
4.2.5	Diskussion der Konzepte . . . . .	73
4.3	Erweiterung der Constraint-Modellierung . . . . .	74
4.3.1	Relaxation von Constraints . . . . .	74
4.3.2	Eliminierung von Variablen durch Fehlwerte . . . . .	75
4.3.3	Inferenzregeln . . . . .	76
4.3.4	Optimierung mittels Branch & Bound . . . . .	78
4.3.5	A priori und aktive Beschränkungen des Suchraumes . . . . .	79
4.3.6	Praktische Evaluation . . . . .	79
4.3.7	Übertragung auf andere überbeschränkte Probleme . . . . .	82

<b>5</b>	<b>Bestimmung der wahrscheinlichsten relationalen Zuordnung</b>	<b>85</b>
5.1	Inexakte relationale Zuordnungen . . . . .	86
5.1.1	Ein einfaches Distanzmaß für die strukturelle Abweichung . . . . .	87
5.1.2	Relational Matching als Kommunikationsproblem . . . . .	87
5.2	Berechnung der gemeinsamen Information für die verwendeten Constraints . . . . .	90
5.2.1	Berechnung der a priori Wahrscheinlichkeiten . . . . .	92
5.2.2	Training der bedingten Wahrscheinlichkeiten . . . . .	93
5.3	Kopplung mit den erweiterten Constraint-Techniken . . . . .	95
5.3.1	Suchraumeinschränkungen . . . . .	96
5.3.2	Heuristiken . . . . .	96
5.3.3	Praktische Umsetzung . . . . .	97
<b>6</b>	<b>Bestimmung des wahrscheinlichsten Modells</b>	<b>99</b>
6.1	MAP-Klassifikation mit Hilfe des Prinzips der minimalen Beschreibungslänge . . . . .	100
6.1.1	Das Prinzip der minimalen Beschreibungslänge . . . . .	101
6.1.2	Optimale Codes . . . . .	102
6.1.3	Anwendungen des MDL-Prinzips zur Klassifikation und Objekterkennung . . . . .	102
6.2	Codierung relationaler Gebäudestrukturen . . . . .	105
6.2.1	Allgemeine Codierung von Graphen . . . . .	106
6.2.2	Codierung der Relationen des Gebäudemodells . . . . .	109
6.2.3	Codierung der gesamten Gebäudeansicht . . . . .	113
6.3	Praktische Evaluation . . . . .	117
6.3.1	Gebäudeerkennung in synthetischen Bildern . . . . .	117
6.3.2	Gebäudeerkennung in realen Bildern . . . . .	122
<b>7</b>	<b>Zusammenfassung</b>	<b>125</b>
7.1	Rückblick und Bewertung . . . . .	126
7.2	Beiträge der Dissertation . . . . .	128
7.3	Ausblick . . . . .	130
	<b>Literaturverzeichnis</b>	<b>133</b>
	<b>Abbildungsverzeichnis</b>	<b>143</b>
	<b>Tabellenverzeichnis</b>	<b>145</b>
	<b>Index</b>	<b>147</b>



# Kapitel 1

## Einleitung

Das Thema dieser Arbeit ist die automatische Erkennung und Rekonstruktion von Gebäuden in digitalen Luftbildern. Das Ziel besteht darin, Gebäude im Bild zu identifizieren, zu lokalisieren und anschließend als dreidimensionale Modelle zu rekonstruieren. Der Fokus dieser Arbeit liegt auf der Entwicklung einer geeigneten Repräsentation von Gebäudemodellen, die die Anwendung effizienter Suchverfahren erlaubt, sowie der Konstruktion eines Bewertungsschemas, das trotz Störungen in den Bilddaten Gebäude vom Bildhintergrund unterscheiden kann und gleichzeitig die Klassifikation erkannter Gebäude in verschiedene Gebäudetypen ermöglicht.

### 1.1 Anwendungen digitaler Gebäudemodelle

In den letzten zehn Jahren ist der Bedarf an dreidimensionalen Stadtmodellen stark angestiegen, wobei das Hauptinteresse auf den Objekten liegt, die eine Stadt im wesentlichen ausmachen: den Gebäuden. Dreidimensionale Modelle von Gebäuden werden für eine ganze Reihe verschiedener Anwendungen in den Bereichen Architektur, Stadt- und Raumplanung, Vermessung, Telekommunikation, Transportwesen und Umweltschutz benötigt. Darüberhinaus bilden 3D-Gebäudemodelle einen wesentlichen Bestandteil von Stadtinformationssystemen, die von der Vermittlung touristischer Eindrücke und Auskünfte bis hin zur Einsatzplanung von Rettungs- und Katastrophendiensten eingesetzt werden. Zur Verdeutlichung der verschiedenen Anforderungen an die Daten, werden im folgenden zunächst einige dieser Anwendungen näher betrachtet.

Im Bereich der Städte- und Raumplanung geht die Entwicklung bei der Durchführung konkreter Bauvorhaben in die Richtung, daß der Stadtplaner in seinem Büro am PC einen umfassenden Überblick über die geplanten und bereits durchgeführten Maßnahmen erhält, und sich dadurch die Anzahl der kostenintensiven Besichtigungstermine vor Ort reduzieren läßt. Zu den für die Überwachung und Abwicklung der Bauvorhaben notwendigen Informationen gehören insbesondere auch die Daten der umgebenden Gebäude wie z.B. Art des Gebäudes, Dachtyp und Abmessungen (vgl. [Bau97]).

Ein wichtiger Aspekt bei der Planung von Bauvorhaben ist die geeignete Aufbereitung und Präsentation der geplanten Maßnahmen zur Unterstützung der Entscheidungsfindung. Während dazu früher Modelle aus Holz und Gips der geplanten Gebäude und ihrer Umgebung angefertigt wurden, werden heutzutage die Entwürfe mittels CAD-Systemen konstruiert, welche eine Ansicht des geplanten Gebäudes aus beliebigen Perspektiven und darüberhinaus auch Animationen ermöglichen. Eine wesentliche Aufgabe der Visualisierung besteht darin, einen Eindruck zu vermitteln, wie sich das neue Gebäude in seine Nachbarschaft einfügt. Dazu müssen die Gebäude der Umgebung jedoch vorher erfaßt worden sein (vgl. [BB98]).

Eine große Nachfrage an Gebäudedaten kommt von den Telekommunikationsunternehmen, insbesondere aus dem Geschäftsfeld der mobilen Telefonie. Mobiltelefone besitzen nur eine geringe Sendeleistung. Da die Bausubstanz von Gebäuden die Funksignale stark dämpft, muß je nach Bebauung eine entsprechende Anzahl von Sendestationen installiert werden, um einen flächendeckenden Kontakt zu einer der festinstallierten Sendestationen zu gewährleisten. Für die Planung der Senderstandorte werden Simulationen durchgeführt, deren hauptsächliche Eingangsdaten aus den dreidimensionalen Modellen der Gebäude bestehen (siehe [SB97]).

Im Bereich der Raumplanung und des Umweltschutzes werden 3D-Gebäudemodelle für die Berechnung des Stadtklimas und der -belüftung eingesetzt. Am Städtebaulichen Institut der Universität Stuttgart wird beispielsweise z.Zt. an einem Forschungsprojekt gearbeitet, bei dem es um die umweltverträgliche Bau- und Verkehrsplanung geht (Projekt WUMS). Dabei wird die Lärm- und Schadstoffausbreitung entlang einzelner Straßenzüge für einzelne Stadtteile simuliert, wofür wiederum 3D-Modelle der entsprechenden Gebäude benötigt werden. Zur Berücksichtigung von Verwirbelungen müssen auch die jeweiligen Dachformen der Gebäude erfaßt sein.

Für die Vermessung und Kartographie spielen Gebäudemodelle künftig eine größere Rolle, weil die Kommunen ihre Katasterdaten zunehmend dreidimensional vorhalten. Zu den Gebäuden werden nicht wie bislang beim ALK<sup>1</sup> üblich nur der Grundriß und die mittlere Gebäudehöhe erfaßt, sondern auch die Dachform, die Unterteilung in verschiedene Gebäudeteile und deren Abmessungen.

## 1.2 Gebäudeextraktion aus Luftbildern

Die Akquisition dreidimensionaler Gebäudedaten geschieht im wesentlichen auf der Basis von digitalisierten Luftbildern. Luftbilder eignen sich insbesondere zur Gebäudeerfassung, weil sie auf der einen Seite großflächige Gebiete überdecken, und auf der anderen Seite dabei gleichzeitig eine hohe Auflösung bieten, die eine gute Erkennung sowie detaillierte Rekonstruktion von Gebäuden mit hoher Genauigkeit ermöglicht. Ferner erlaubt die Stereoauswertung mehrerer Bilder, die dieselbe Szene aus unterschiedlichen Perspektiven zeigen, eine unmittelbare 3D-Rekonstruktion.

Die Auswertung der Luftbilder erfolgt bislang noch hauptsächlich manuell unter Verwendung von Techniken der digitalen Photogrammetrie. Semiautomatische Systeme unterstützen dabei den Operateur durch die Bereitstellung vordefinierter Gebäudetypen oder -teile, die interaktiv am Bildschirm plziert und vermessen werden (vgl. [LS93, LF96a, EG96, Gül97, Gül98]). Aktuelle Systeme führen bereits einfache Klassifikationen und Parameterbestimmungen durch [Grü97, LG98] und erreichen dabei eine Genauigkeit, die der bei einer manuellen Erfassung am (analogen) analytischen Plotter entspricht (siehe [GMLR98]).

## 1.3 Automatische Gebäudeextraktion

Der große Bedarf an dreidimensionalen Stadt- und Gebäudeinformationen kann aus Zeit- und Kostengründen nicht mehr allein mit Hilfe von manuellen und semiautomatischen Erfassungsmethoden gedeckt werden. Es werden deshalb Verfahren zur automatischen Extraktion von Gebäuden benötigt. Neben der Ersterfassung können diese Verfahren später auch zur automatisierten Fortführung eingesetzt werden, was weiter zur Erhöhung der Aktualität der 3D-GIS-Datenbestände beiträgt.

Die automatische Gebäudeextraktion aus Luftbildern stellt sich als schwieriges Problem der Objekterkennung und Bildinterpretation dar (vgl. [SFH92, Rad93]), weil zum einen die zu erkennenden Objekte, d.h. die Gebäude, eine komplexe Struktur aufweisen, die sich durch einen hohen Grad an Variabilität auszeichnet. Gebäude sind unterschiedlich groß, haben vielfältige Dachformen

---

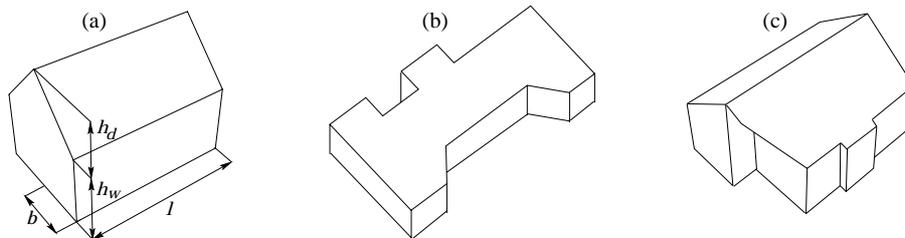
<sup>1</sup>ALK steht für Automatisierte Liegenschaftskarte und bezeichnet das Modell, das zur Repräsentation der für Katasterpläne relevanten 2D-Geo- und Sachdaten verwendet wird.

und -farben, und können sehr komplexe Grundrisse besitzen. Zum anderen ist die Komplexität der Bilddaten ebenfalls sehr hoch: Luftbilder sind sehr groß und enthalten auf der einen Seite neben den interessierenden Gebäuden noch eine große Menge irrelevanter Informationen wie z.B. Vegetation und Verkehrswege. Auf der anderen Seite fehlen jedoch auch Informationen, weil Bilder nur zweidimensionale Projektionen der Szene darstellen und keine direkte Tiefeninformation mehr beinhalten, und darüberhinaus immer Teile der Gebäude verdeckt sind.

In den letzten zehn Jahren sind bereits verschiedene automatische Extraktionsverfahren vorgeschlagen worden. Die Konzepte konzentrieren sich dabei im wesentlichen auf die Gebäudeerfassung in Vorstadtbereichen, Kleinstädten, Dörfern aber auch Industriegebieten, weil die dort auftretenden Gebäude meistens freistehend sind und sich somit problemlos von anderen Gebäuden separieren lassen, was die Aufgabe der Gebäudeextraktion auf die Erkennung von Einzelgebäuden reduziert. Für Altstadt- oder Innenstadtgebiete hingegen ist es schwierig überhaupt zu definieren, was ein Gebäude ist; oftmals handelt es sich um große, verschachtelte Baublöcke. Das in dieser Arbeit vorgestellte Verfahren wird sich deshalb ebenfalls auf die erstgenannten Gebiete beschränken.

Im Unterschied zur semiautomatischen Gebäudeerfassung müssen bei der automatischen Gebäudeextraktion die Gebäude zunächst im Bild erkannt werden. Dazu wird ein Modell benötigt, daß auf der einen Seite die wesentlichen Gebäudeeigenschaften widerspiegelt und auf der anderen Seite mit den Mitteln der digitalen Bildverarbeitung in den Bildern identifizierbar ist. Allen bislang entwickelten Konzepten liegen dabei Gebäudemodelle zugrunde, die sich letztlich in zwei Kategorien einteilen lassen, wobei mit jeder Kategorie auch eine Grundstrategie zur Erkennung und 3D-Rekonstruktion verbunden ist:

Bei der ersten Sorte von Modellen werden Gebäude durch parametrische Volumenprimitive dargestellt, deren Ausdehnung, Orientierung und Position zwar variabel sind, nicht jedoch deren Form. Es können somit nur Gebäude erkannt werden, die a priori in einer Bibliothek von Gebäudetypen explizit aufgezählt sind. Die Erkennungsstrategie ist im wesentlichen modellgetrieben (Top-Down), wobei die Luftbilder nach den dem System bekannten Gebäuden durchsucht werden. Da die Luftbilder nur zweidimensional, die Modelle aber dreidimensional sind, werden Projektionen der 3D-Modelle berechnet, um den Vergleich durchführen zu können.



**Abbildung 1.1:** Verschiedene Gebäudemodellierungen vergleichbarer Ansätze zur automatischen Gebäudeextraktion: (a) parametrisches Volumenmodell, (b) prismatisches Polyedermodell und (c) allgemeines Polyedermodell.

Die zweite Sorte von Modellen basiert auf Polyedern, die z.T. auf Prismen eingeschränkt sind<sup>2</sup>. Die Strategien der Ansätze beruhen auf der datengetriebenen, schrittweisen Aggregation von aus dem Bild extrahierten Merkmalen (Bottom-Up) nach den Prinzipien der perceptiven Gruppierung ([Moh89]). Durch Einbeziehung der Schatteninformation oder durch Mehrbildzuordnung werden dabei 3D-Merkmale generiert, die schließlich zu dreidimensionalen polyedrischen Objekten gruppiert werden. Da sich die Modellierung in vielen Arbeiten nur auf die Geometrie (und z.T. auch die Radiometrie) beschränkt, und gebäudespezifisches Wissen allenfalls implizit in die verwendete Erkennungsstrategie eingeht, handelt es sich eher um allgemeine Verfahren zur 3D-

<sup>2</sup>Prismatische Modelle werden zur Modellierung von Gebäuden mit flachen Dachstrukturen verwendet. Dabei wird der Dachumriß durch ein ebenes Polygon modelliert und dann auf den Boden projiziert.

Formrekonstruktion, weil die Modellierung nicht ausreicht, um die rekonstruierten Polyeder näher und insbesondere als Gebäude bestimmter Arten zu interpretieren.

Abbildung 1.1 zeigt jeweils ein Beispiel für die verschiedenen Modellkategorien. Das Problem bei diesen Modellen jedoch ist, daß sie entweder so beschränkt sind, daß nur wenige aufgezählte Gebäudetypen erkannt werden können (parametrische Volumenmodelle), oder aber sie sind so generisch, daß sie zwar auf viele Gebäude passen, aber nicht ausreichend gebäudespezifisch sind und auch auf Autos, Hundehütten und andere künstliche Objekte passen (prismatische und allgemeine Polyedermodelle).

Um die Lücke zwischen den Modellierungen an den beiden Seiten des Spektrums aufzufüllen, wurde im Rahmen des DFG-Bündelprojekts „Semantische Modellierung“, insbesondere im Teilprojekt „Extraktion von Gebäuden aus Digitalen Bildern“, ein Konzept entwickelt, das dem Anspruch der Generik sowie der Gebäudespezifität gerecht wird (vgl. [BKL<sup>+</sup>95b, FKL97, FKL<sup>+</sup>98]). In diesem Konzept wird die Generik durch eine komponentenbasierte, volumetrische Objektmodellierung ermöglicht. Gebäude setzen sich dabei aus Gebäudeteilen zusammen, die nach wohldefinierten Regeln kombiniert werden können. Die Teile besitzen gebäudecharakteristische Eigenschaften wie z.B. den Dachtyp und sind entsprechend parametrisiert (z.B. Breite, Höhe bis zum Dach, Höhe vom Dachansatz zum Giebel usw.). Auf den Parametern sind ferner Einschränkungen definiert wie z.B. die Mindestgeschoßhöhe oder das minimale und maximale Höhen/Breitenverhältnis bestimmter Dachformen. Diese Einschränkungen sorgen dafür, daß andere dreidimensionale Objekte wie die bereits erwähnte Hundehütte oder auch Autos nicht als Gebäude interpretiert werden.

Diese Modellierung erfordert nun eine besondere Strategie, weil sie auf der einen Seite beliebig viele Gebäudekonstruktionen erlaubt und somit nicht einfach alle möglichen Gebäudetypen von vornherein aufgezählt und im Bild gesucht werden können. Auf der anderen Seite reicht es aber zur Konstruktion eines Modells auch nicht aus, einfach nur Bildmerkmale nach geometrischen Regeln sukzessive zu gruppieren, weil für die verschiedenen Gebäudeteile unterschiedliche geometrische Regularitäten gelten, und am Anfang nicht klar ist, welche Merkmale zu welchem Gebäudeteil bzw. überhaupt zum Gebäude gehören. Aus diesem Grund wird eine kombinierte Strategie verwendet, die zunächst Bildmerkmale anhand geometrischer und topologischer Eigenschaften nur soweit gruppiert, daß die entstehende Aggregate noch zu allen Teilen eines Gebäudes passen können. Anhand dieser Beobachtungen werden dann verschiedene Hypothesen für Gebäude gebildet, die das Auftreten dieser Merkmalsaggregate im Bild erklären können.

Nach der Generierung der Gebäudehypothesen müssen diese dann im Bild verifiziert werden. Dabei treten eine Reihe von Problemen auf, auf deren Lösung sich diese Arbeit im wesentlichen konzentriert: Aufgrund der unvollständigen Beobachtungen weisen die Gebäudehypothesen meistens noch Freiheitsgrade bzgl. ihrer Größe, Lage und Orientierung im Bild auf. Das bedeutet, daß die Gebäudehypothesen zunächst im Bild gesucht und identifiziert werden müssen. Da die Gebäude auch trotz teilweiser Verdeckungen und anderer Bildstörungen z.B. durch Rauschen erkannt werden sollen, müssen die Gebäudehypothesen für die Identifikation in ihre Bestandteile zerlegt und entsprechend nach diesen Teilen gesucht werden, wobei einzelne Teile dann fehlen dürfen. Schließlich muß aus den verschiedenen Gebäudehypothesen die zutreffende ausgewählt werden. Das heißt, es muß ein Bewertungsmaß gefunden werden, das trotz verrauschter und fehlerhafter Daten sowie unterschiedlich komplexer Gebäudehypothesen eine objektivierbare und damit vergleichbare Bewertung ermöglicht.

## 1.4 Ziele der Arbeit

Das Hauptziel der Arbeit ist die automatische Identifikation und Rekonstruktion von Gebäuden. Wie im vorigen Abschnitt erläutert, besteht das Konzept zur Lösung dieser Aufgabe im wesentlichen aus den beiden Schritten der Hypothesengenerierung und -verifikation. Für die Generierung der Hypothesen wird dabei auf das in [FKL97, FKL<sup>+</sup>98] beschriebene Verfahren zurückgegriffen.

Diese Arbeit konzentriert sich auf den Aspekt der Verifikation, wobei sich entsprechend der obigen Problemendarstellung die folgenden Teilziele ergeben:

**Identifikation geeigneter Relationen zur Gebäudeerkennung:** Für die Erkennung der Gebäudehypothesen im Bild muß nach Bildmerkmalen gesucht werden, die in ihren Eigenschaften mit denen der Modellmerkmale übereinstimmen. Zum Zeitpunkt der Suche sind aufgrund der freien Parameter der Gebäudehypothese noch nicht alle Modelleigenschaften wie z.B. die Position im Bild, die Längen einzelner Linien oder die Größen der Flächen bekannt. Darüberhinaus sind viele Eigenschaften der Bildmerkmale nicht besonders robust gegenüber Störungen. Hier spielen die geometrischen und topologischen Relationen zwischen den Merkmalen des 3D-Objektmodells eine wichtige Rolle, weil diese größtenteils invariant gegenüber der konkreten Größe, Lage und Perspektive der Gebäudehypothesen sind.

Ein Ziel dieser Arbeit besteht nun in der Identifikation geeigneter Relationen, die sich auf das zweidimensionale Bildmodell übertragen lassen und dort ebenfalls invariant bzgl. einzelner Projektionen (Aspekte) des Gebäudemodells sind. Die verschiedenen Relationen müssen insgesamt eine so spezifische Modellierung erlauben, daß die 2D-Gebäudehypothesen nur anhand ihrer relationalen Repräsentation vom Bildhintergrund diskriminiert werden können. Da die erkannten Gebäude auch klassifiziert werden sollen, müssen die zu bestimmenden Relationen darüberhinaus eine ausreichende Trennschärfe besitzen, um die verschiedenen Gebäude anhand ihrer relationalen Repräsentationen voneinander unterscheiden zu können.

**Verwendung von Constraint-Techniken zur Suche nach Gebäuden:** Die Dekomposition von Gebäudehypothesen in eine relationale Struktur legt es nahe, Constraint-Techniken zur Auffindung solcherart repräsentierter Modelle im Bild einzusetzen. Die relationale Modellrepräsentation kann nämlich als Menge von Bedingungen (Constraints) aufgefaßt werden, die spezifizieren, wann eine aggregierte Menge einzelner Merkmale insgesamt als Gebäude interpretiert werden kann. Die Aufgabe der Erkennung bildet sich damit auf die Suche einer Zuordnung von extrahierten Bildmerkmalen zu den Merkmalen des Modells ab, wobei die zugeordneten extrahierten Bildmerkmale alle Bedingungen erfüllen müssen, die zwischen den Modellmerkmalen bestehen. Die bereits erwähnten Constraint-Techniken bieten dabei Methoden zur effizienten Lösung dieses harten kombinatorischen Problems.

Die Effizienz der Constraint-Techniken beruht im wesentlichen darauf, daß die Lösung eines entsprechend formulierten Suchproblems alle Constraints erfüllen muß, und die Constraints dabei kategorisch in erfüllt bzw. nicht erfüllt unterschieden werden. Aufgrund von Störungen im Bild werden in bestimmten Situationen nicht alle Constraints eines Gebäudemodells erfüllbar sein, weshalb in diesen Fällen die betreffenden Gebäude mit Standard-Constraint-Techniken nicht erkannt werden können. Die Constraint-Modellierung soll nun so erweitert werden, daß auf der einen Seite die verschiedenen Störungen durch Unbeobachtbarkeit und Unschärfe im Bild angemessen berücksichtigt werden können, und sie auf der anderen Seite dennoch eine aktive Suchraumeinschränkung ermöglicht.

**Entwicklung einer Bewertungsfunktion,** die sowohl das Objekt- als auch das Bildmodell einbezieht, und dabei folgende Eigenschaften aufweist:

- Das Bewertungsschema soll für die Suche nach einzelnen Gebäudehypothesen in den Bildern verwendet werden. Dabei geht es um die Bestimmung der besten möglichen Zuordnung zwischen einem Modell und den Bilddaten.
- Die Bewertungsfunktion soll darüberhinaus die Klassifikation der Gebäude ermöglichen, d.h. aus einer Menge möglicher Gebäudehypothesen die zutreffendste auswählen.
- Aufgrund von Störungen in den Bildern z.B. durch Verdeckungen, Rauschen oder schlechtem Kontrast ist nicht damit zu rechnen, daß die Modelle vollständig beobachtbar sind. Ferner führen Unschärfe und Unsicherheit z.B. durch die Digitalisierung zu mehr oder minder starken Abweichungen der Bilddaten von den Modellen. Dies ist bei der Bewertung explizit zu berücksichtigen.

- Die Bewertung sollte selbstkalibrierend sein, in dem Sinne, daß keine von außen abzustimmenden und schwer motivierbaren Parameter notwendig sind. Alle innerhalb der Bewertung benötigten Gewichte sollten analytisch bestimmt oder anhand von Trainingsdaten gelernt werden können.
- Die Ergebnisse der Bewertung sollten objektivierbar sein, d.h. eine nachvollziehbare Semantik besitzen. Da mit unsicheren Aussagen gearbeitet wird, bietet sich eine Wahrscheinlichkeitstheoretische Interpretation in Form der Maximum-Likelihood- und der Maximum a posteriori-Schätzung an.
- Die Bewertung sollte die Möglichkeit der Selbstdiagnose bieten, d.h. eine Aussage über die Qualität der Entscheidungen treffen können.

**Kopplung der Bewertungsfunktion mit den Constraints:** Die Constraint-Techniken sind so mit der Bewertung zu verbinden, daß sie erstens die beste Lösung des Constraint-Problems den obigen Ansprüchen an die Bewertungsfunktion genügt, und zweitens die Techniken zur Suchraumeinschränkung weiterhin genutzt werden können. Vorwissen aus den Trainingsdaten soll zur a priori Beschränkung des Suchraumes genutzt werden.

**Anwendbarkeit der Konzepte auf reale, verrauschte Daten:** Die zu entwickelnden Verfahren sollen sich insbesondere für die praktische Anwendung mit realen Daten eignen. Um dies zu überprüfen, sollen sie an realen Luftbildern exemplarisch evaluiert werden.

## 1.5 Gliederung der Arbeit

Die Arbeit gliedert sich im wesentlichen in drei Teile, wobei jeder Teil aus zwei Kapiteln besteht. Im ersten Teil werden nach dieser Einleitung in Kapitel 2 die Probleme der automatischen Gebäudeextraktion aus Luftbildern näher erläutert, die Konzepte anderer Forschungsgruppen diskutiert, und schließlich das verwendete Modell und die Strategie im einzelnen vorgestellt. Es werden die für diesen Ansatz besonderen Aufgabenstellungen vorgestellt, und detailliert die Problemstellung definiert, der sich diese Arbeit widmet. Darüberhinaus wird gezeigt, wie im Bild erkannte Gebäude anhand des Gebäudemodells lokalisiert und dreidimensional rekonstruiert werden.

Wie oben ausgeführt, lassen sich die zwei wesentlichen Problembereiche, mit denen sich diese Arbeit beschäftigt, durch die zwei Begriffe Effizienz und Bewertung grob charakterisieren. Diesen Themen sind insofern auch die beiden anderen Teile der Arbeit gewidmet.

Im zweiten Teil geht es in Kapitel 3 zuerst darum zu zeigen, wie Gebäude mittels Techniken des Constraint-Solvings effizient in realen Daten (aber zunächst unter idealisierenden Annahmen) identifiziert werden können. Dazu werden als erstes die notwendigen Grundlagen zum Thema Constraint Satisfaction erläutert. Anschließend wird gezeigt, wie sich die Suche nach einem Gebäude auf die Erfüllung von Constraints abbilden läßt, und wie Gebäudehypothesen durch Constraint-Mengen repräsentiert werden. Die Eigenschaften der verwendeten topologischen und geometrischen Constraints werden dabei im Detail erörtert. In Kapitel 4 geht es um die explizite Berücksichtigung der verschiedenen möglichen Störungen in den Bildern, die dazu führen, daß Standard-Constraint-Techniken zur Identifikation nicht mehr ausreichen. Es werden zuerst die zur Lösung überbeschränkter Constraint-Systeme etablierten Verfahren auf ihre Eignung für das Problem der Gebäudeerkennung diskutiert. Da keines der bestehenden Verfahren eine adäquate Modellierung erlaubt, wird danach ein eigenes Konzept zur Erweiterung der Constraint-Modellierung vorgestellt, das die Phänomene der Unbeobachtbarkeit und der Unschärfe explizit repräsentiert.

Im dritten Teil der Arbeit wird in Kapitel 5 gezeigt, wie sich die Constraints unter Verwendung einer informationstheoretisch basierten Bewertungsfunktion so gewichten lassen, daß die beste Zuordnung zwischen Modell und Bild eine probabilistische Interpretation erhält. Es wird dabei eine Maximum-Likelihood-Schätzung zwischen Modell und Bild erreicht. Es wird dann gezeigt, wie die benötigten Wahrscheinlichkeiten analytisch aus dem Modell sowie empirisch anhand von Trainingsdaten gewonnen werden. In Kapitel 6 geht es schließlich um die Auswahl des wahrscheinlichsten

Modells aus einer Reihe möglicher Hypothesen, die zuvor den Bilddaten zugeordnet wurden. Da die Modelle unterschiedlich komplex sein können, wird zur Klassifikation das statistische Verfahren des Prinzips der minimalen Beschreibungslänge eingesetzt, das die Modellkomplexität explizit über seine Beschreibungslänge einbezieht. Da dazu die Modelle optimal codiert sein müssen, wird ein Schema zur kompakten Codierung der relational repräsentierten Gebäudehypothesen vorgestellt. Schließlich wird die praktische Anwendbarkeit des entwickelten Konzepts sowohl an synthetischen als auch realen Bilddaten demonstriert.

In Kapitel 7 werden zuletzt die erreichten Ergebnisse zusammengefaßt und bewertet. Anschließend werden die Beiträge dieser Dissertation nach den verschiedenen wissenschaftlichen Gebieten geordnet dargestellt. Eine Diskussion möglicher Erweiterungen schließt diese Arbeit ab.

**Anmerkung zur Nomenklatur:** Der Begriff Constraint wird in dieser Arbeit nicht übersetzt, weil die deutschen Übersetzungen in Zwang oder (Rand-)bedingung eher einen Zustand beschreiben und damit die beabsichtigte Konnotation nicht genau treffen. Wenn in dieser Arbeit von Constraints die Rede ist, ist damit ein aktives Element der Wissensrepräsentation gemeint, das neben der Beschreibung von Beschränkungen diese eben auch aktiv erzwingen kann.



## Kapitel 2

# Automatische Gebäudeextraktion

In diesem Kapitel wird zuerst die Komplexität des Problems der automatischen Gebäudeerkennung erörtert. Dazu wird zunächst die Komplexität der Bilddaten diskutiert. Danach werden die generellen Modellierungskonzepte und Strategien aus dem Bereich der Objekterkennung behandelt, die von den bestehenden Ansätzen und dem in dieser Arbeit vorgestellten Konzept zur automatischen Gebäudeextraktion eingesetzt werden.

Danach werden die vergleichbaren Forschungsansätze vorgestellt und in zweierlei Hinsicht diskutiert: zum einen wird die Ausdrucksstärke der verwendeten Gebäudemodelle untersucht, und zum anderen wird bei modellgetriebenen Ansätzen auf das Suchverfahren und das zur Verifikation verwendete Bewertungsschema eingegangen. Der erste Punkt soll im wesentlichen verdeutlichen, warum in dieser Arbeit ein ausdrucksstärkeres Modell eingesetzt wird. Der zweite Punkt ist wichtig im Hinblick auf das bzgl. dieser Modellierung zu lösende Problem der Verifikation von Gebäudehypothesen, dem sich diese Arbeit hauptsächlich widmet.

Im Anschluß wird die verwendete Gebäudemodellierung im Detail erläutert, und die Gesamtstrategie zur Erkennung und 3D-Rekonstruktion erklärt. Dabei werden zunächst die verschiedenen Schritte der Hypothesengenerierung vorgestellt, bevor im einzelnen die Anforderungen an die Verifikation der Gebäudehypothesen und die sich daraus ergebenden Aufgabenstellungen erörtert werden, die die Ausgangspunkte für den weiteren Verlauf dieser Arbeit darstellen. Schließlich wird gezeigt, wie die identifizierten Häuser anhand des Gebäudemodells mittels eines Parameterschätzverfahrens dreidimensional rekonstruiert werden.

### 2.1 Komplexität der Daten

Luftbilder stellen das in dieser Arbeit zur automatischen Gebäudeextraktion verwendete Datenmaterial dar. Sie besitzen eine hohe Komplexität, die nach Mayer ([May98]) von zwei wesentlichen Faktoren bestimmt wird: Dies ist erstens der *Bildinhalt*, der von der betrachteten Szene abhängt. So ist beispielsweise ein Luftbild eines Innenstadt- oder Altstadtbereiches als deutlich komplexer einzustufen als das eines ländlichen Gebietes, in dem hauptsächlich Ackerflächen und nur vereinzelte Gebäude abgebildet sind. Die in dieser Arbeit verwendete Modellierung und Erkennungsstrategie eignet sich im wesentlichen zur Gebäudeerkennung in Szenen mittlerer Komplexität, wie sie z.B. in Luftbildern von Vorstadtbereichen, Kleinstädten und Dörfern vorgefunden wird. Die in solchen Bildern auftretenden Gebäude weisen dabei bereits viele unterschiedliche Formen, Größen und Farben auf. Luftbilder beinhalten aber auch viele für die Gebäudeerkennung nicht unmittelbar relevante Informationen wie z.B. Vegetation, Gewässer und Verkehrswege.

Der zweite komplexitätsbeeinflussende Faktor ist die *Beobachtbarkeit*, die von mehreren Voraussetzungen abhängt, von denen die erste die Auflösung ist, die hauptsächlich durch den Bildmaßstab und die Digitalisierauflösung bestimmt wird. Luftbilder weisen typische Bildmaßstäbe zwischen

1:100 000 und 1:4000 auf (vgl. [Kra94, May98]). Wird beispielsweise ein Bild im Maßstab 1:5000 mit einer Scan-Auflösung von  $15 \mu\text{m}$  digitalisiert, besitzt das digitale Rasterbild eine Bodenauflösung von 7,5 cm. Für eine Gebäudeerkennung, die sich nicht auf die Detektion beschränkt, sondern eine Klassifikation und insbesondere eine 3D-Rekonstruktion durchführen soll, sollten nach eigener Erfahrung Bilder mit mindestens einer Bodenpixelauflösung von 25 cm zur Verfügung stehen (siehe auch [SSF96, HB97]).

Die Beobachtbarkeit wird darüberhinaus durch die Tatsache beschränkt, daß Luftbilder nur zwei-dimensionale Projektionen der abgebildeten dreidimensionalen Szene darstellen. Dabei sind immer nur die der Kamera zugewandten Teile der Gebäude sichtbar; die hinteren Teile werden von den vorderen verdeckt (Eigenverdeckung). Da die Aufnahmerichtung bei Luftbildern fast immer orthogonal zum Boden ist, sind in den meisten Fällen zumindest die Dächer vollständig beobachtbar. Aufgrund der Projektion beinhalten Luftbilder keine explizite Tiefeninformation mehr. Die Erkennung wird dadurch insofern erschwert, daß die Projektionen nicht eindeutig sind, weil verschiedene Objekte in unterschiedlicher Perspektive dieselbe Projektion besitzen können.

Gebäudeteile können aber nicht nur eigenverdeckt, sondern durch andere Objekte wie z.B. Bäume, nebenstehende Gebäude und deren Schatten fremdverdeckt sein. Einen weiteren Einfluß auf die Beobachtbarkeit hat die Beleuchtung in Verbindung mit der Aufnahmeperspektive. So können bei ungünstigen Verhältnissen benachbarte Flächen im Bild dieselben radiometrischen Ausprägung besitzen und deshalb nicht voneinander unterschieden werden.

Ein Teil der Probleme der Beobachtbarkeit läßt sich durch die Verwendung mehrerer Bilder, die dieselbe Szene aus unterschiedlichen Perspektiven zeigen, vermeiden. Bei heutigen Luftbildbefliegungen wird oftmals bereits eine vierfache Überdeckung erreicht, d.h. jedes Objekt der überflogenen Szene ist in bis zu vier verschiedenen Aufnahmen abgebildet. Durch eine Stereobildauswertung kann die bei Einzelbildern fehlende Tiefeninformation wiedergewonnen werden (vgl. [HS93a, Kra94]). Ferner erhöht sich durch die Redundanz die Wahrscheinlichkeit, daß in einem Bild verdeckte oder aufgrund des schlechten Kontrasts undifferenzierbare Gebäudeteile in den anderen beobachtbar sind bzw. unterschieden werden können.

## 2.2 Modelle und Strategien zur Gebäudeerkennung

Die Identifikation und Rekonstruktion von Gebäuden in Luftbildern stellt ein schwieriges Problem der Objekterkennung dar. Es wird eine Modellierung benötigt, die die charakteristischen, in Luftbildern beobachtbaren Ausprägungen von Gebäuden repräsentieren kann, sowie eine Strategie, die beschreibt wie Gebäude in den Bilddaten identifiziert werden können. Da Gebäude nicht nur detektiert, sondern insbesondere auch geographisch lokalisiert, klassifiziert und dreidimensional rekonstruiert werden sollen, reichen einfache Modelle und Verfahren der Mustererkennung wie die *merkmalsvektorbasierte Klassifikation* (vgl. [DH73]) oder *Template Matching* (vgl. [SFH92]) nicht aus.

### 2.2.1 Anforderungen an die Modellierung

Eine geeignete Modellierung muß zunächst eine ausreichende Generik aufweisen, um die verschiedenen Gebäudeformen repräsentieren zu können. Darüberhinaus muß sie eine ausreichende Spezifität besitzen, um Gebäude von anderen abgebildeten Objekten wie z.B. Straßen, Autos und der Vegetation unterscheiden zu können. Eng damit verbunden ist die Frage nach dem Detailgrad bzw. der Granularität der Modellierung. Diese muß nämlich auf der einen Seite ausreichend groß sein, um den Anforderungen der Anwendungen gerecht zu werden, und sich auf der anderen Seite aber auch an der verfügbaren Bildauflösung und der Komplexität der Suchverfahren orientieren. Beispielsweise macht es keinen Sinn, Erker und Dachaufbauten zu modellieren, wenn diese bei der gewählten Bildauflösung gar nicht identifiziert werden können. Es ist aber auch nicht praktikabel,

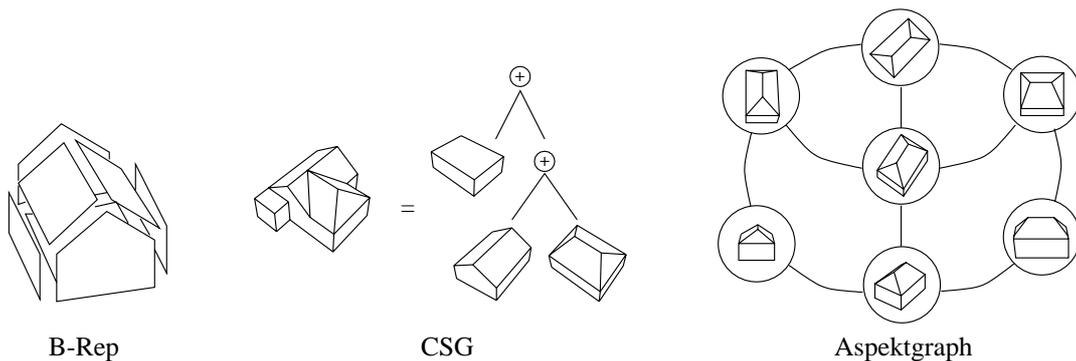
in sehr hochauflösenden Bildern mit einer filigranen Modellierung zu versuchen, Gebäude zu finden, weil der Suchraum aufgrund der vielen Daten und Freiheitsgrade des Modells viel zu groß, und eine Zuordnung von Modell und Bild nicht in akzeptabler Zeit zu erwarten wäre.

Aus den Anwendungen der 3D-Gebäudemodelle, insbesondere aus den Bereichen Stadtplanung, Vermessung und Stadtinformationssysteme, stellt sich die Anforderung, daß Gebäude weiter klassifiziert werden müssen z.B. in Satteldachhaus, L-, T- oder S-förmiges Gebäude usw. Dabei sollten die einzelnen Gebäudebestandteile benannt werden wie z.B. Dach und Seitenwand. Eine Unterscheidung der Gebäudeteile, beispielsweise in Seitentrakt und Haupttrakt, ermöglicht erst die gebäudetypische Parametrisierung, wie z.B. durch die Angabe von Breite und Höhe, oder der Länge des Seitentraktes. Um diese Klassifizierung zu ermöglichen, müssen die semantischen Eigenschaften explizit modelliert werden. Ferner resultieren daraus weitere Anforderungen an die Granularität der Modellierung, denn diese muß so fein sein, daß sie eine zur Klassifikation der verschiedenen Gebäudetypen ausreichende Trennschärfe besitzt.

## 2.2.2 Objekt- und Bildmodellierung

Für die Erkennung und Rekonstruktion von Objekten in Bildern wird zunächst ein *Objektmodell* benötigt, welches die dreidimensionale Ausprägung der gesuchten Objekte beschreibt (vgl. [SFH92, Rad93]). Objektmodelle zur Erkennung dreidimensionaler Objekte basieren in den meisten Fällen auf den allgemeinen geometrischen Modellierungskonzepten der *Randflächendarstellung* (engl.: *Boundary Representation*, kurz: B-Rep) oder der *konstruktiven Festkörpergeometrie* (engl.: *Constructive Solid Geometry*, CSG) (siehe [FvDFH90]).

Die Randflächendarstellung basiert auf der Beobachtung, daß in den Bildern nur die Oberflächen der Objekte wahrgenommen werden können. Zur Objektmodellierung mittels B-Rep werden deshalb auch nur die Objektoberflächen explizit repräsentiert. Die konstruktive Festkörpergeometrie stellt hingegen geometrische Volumenprimitive zur Verfügung, die mittels der mengentheoretischen Operationen Vereinigung, Schnitt und Mengendifferenz zu komplexen Strukturen verknüpft werden können. Die Ausprägung der einzelnen Primitive wird neben ihrem Typ wie z.B. Quader oder Prisma ferner durch die typspezifischen Parameter wie beispielsweise bei einem Quader die Breite, Länge und Höhe bestimmt. Abbildung 2.1 verdeutlicht die Konzepte der B-Rep und des CSG in dem linken und dem mittleren Bild. Auf das rechte Bild wird weiter unten eingegangen.



**Abbildung 2.1:** Verschiedene Arten der 3D-Objekt- und 2D-Bildmodellierung. Von links nach rechts: Randflächendarstellung (*Boundary Representation*, kurz: B-Rep), konstruktive Festkörpergeometrie (*Constructive Solid Geometry*, CSG) und Aspektgraphen (aus [FS97a]).

Anwendungsspezifische Objektmodelle, wie in diesem Fall die Gebäudemodelle, setzen auf diesen Repräsentationsformalismen auf und ergänzen sie um semantische Attribute und Merkmale wie z.B. Flächentyp (Dach oder Wand), Linienorientierung im Raum (horizontal, vertikal, schräg) oder maximale Linienlänge (vgl. [Sto92, Gül92, LF96b]). Ferner wird Gebäudesemantik durch

Relationen ausgedrückt wie z.B. „eine Seitenfläche ist benachbart zu einer Dachfläche“, oder „ein Haus ist immer länger als breit“ (vgl. [BKL<sup>+</sup>95b, FKL97, FKL<sup>+</sup>98]). Der Grad dieses zusätzlichen Wissens entscheidet mit über die Trennschärfe des Modells in bezug auf andersartige Objekte und den Bildhintergrund.

Komplexe Objektmodelle sind oftmals hierarchisch im Sinne einer Bestandteilshierarchie aufgebaut. Die Objekte setzen sich aus Objektteilen zusammen, die wiederum aus kleineren Einzelteilen aggregiert sein können usw. Die Verwendung einer Hierarchie besitzt eine Reihe von Vorteilen: Komplexe Objekte wie Gebäude setzen sich in der Realität ebenfalls aus kleineren Teilen zusammen. Eine komponentenbasierte Modellierung der Teile ermöglicht die logische und semantische Trennung dieser Objektteile. Des weiteren setzen sich Objekte oftmals aus Komponenten der gleichen Art zusammen. Das bedeutet, daß mit der richtigen Auswahl der Komponenten eine Vielzahl verschiedener komplexer Objekte zusammengesetzt werden können. Die für die Erkennung der verschiedenen Teile benötigten Methoden sind zudem in bezug auf das mehrfache Vorkommen der Teile im Objektmodell wiederverwendbar. Schließlich erlaubt die hierarchische Modellierung die Verwendung der sogenannten *Recognition-by-Component*-Erkennungsstrategie (vgl. [SFH92]), bei der nicht das Modell als Ganzes, sondern separat nach den einzelnen Teilen gesucht wird.

Während es sich bei der konstruktiven Festkörpergeometrie um ein inhärent hierarchisches Modellierungsprinzip handelt (vgl. Abbildung 2.1), muß die Zerlegung von B-Rep-Modellen in einzelne Objektteile durch die explizite Aufzählung der zu den Objektteilen gehörenden Randflächen erfolgen. Allerdings ermöglicht die Randflächendarstellung eine natürliche Dekomposition in die geometrischen Primitive Fläche, Linie und Punkt, aus denen die Randflächen gebildet werden.

Neben der Repräsentation der Objekte im 3D-Objektraum wird auch ein Modell für ihre Repräsentation im 2D-Bildraum, das *Bildmodell* benötigt, das beschreibt, wie sich die Objekte konkret im Bild ausprägen. Die meisten Verfahren zur Bildinterpretation setzen nicht bei den einzelnen Bildpunkten (Pixel) auf, sondern führen vor der Suche eine Merkmalsextraktion durch, bei der Bildbereiche nach bestimmten Kriterien zusammengefaßt werden (vgl. [SFH92, Rad93, HS93b]). Meistens werden gerade Liniensegmente, homogene Flächen und auch punktförmige Merkmale extrahiert (vgl. [För94, Hen96, Fuc97]), wobei viele Gebäudeextraktionsverfahren, wie später gezeigt wird, nur eine oder zwei dieser Merkmalsklassen verwenden. Da die extrahierten Bildmerkmale den Bildinhalt durch einzelne Objekte beschreiben, wird auch von einer *symbolischen Bildbeschreibung* bzw. *symbolischen Bildmerkmalen* gesprochen. Die dabei unterschiedenen Merkmalsklassen wie z.B. Punkt, Linie und Fläche stellen die Primitive des Bildmodells dar.

Das Bildmodell muß außer der radiometrischen Ausprägung der Objekte insbesondere auch die Modellierung der Kamera und damit die Projektion der dreidimensionalen Objekte in zweidimensionale Bilder berücksichtigen. Ein Konzept, das diese Verbindung zwischen dem 3D-Objektraum und dem 2D-Bildraum explizit modelliert, ist der *Aspektgraph* (siehe [KvD76, KvD79]). Der grundlegende Gedanke ist dabei, daß es zwar unendlich viele verschiedene Ansichten eines 3D-Objekts gibt, diese sich jedoch in endlich viele qualitativ verschiedene Klassen, die sogenannten *Aspekte*, einteilen lassen. Dabei werden alle topologisch isomorphen Ansichten eines dreidimensionalen Objekts jeweils zu einem Aspekt zusammengefaßt. Die obige Abbildung 2.1 zeigt rechts den Aspektgraph eines Walmdachhauses, wobei nur die für Luftbilder in Frage kommenden Perspektiven berücksichtigt sind. Aspektgraphen können analog zu den Objektmodellen hierarchisch zerlegt werden. Man spricht dann von *Aspekthierarchien* (vgl. [DPR92b]), wobei die einzelnen Aspekte entsprechend in ihre Bestandteile, den Flächen, Linien und Punkten zerlegt werden. Dabei werden oftmals auch Zwischenstufen modelliert, die aus bestimmten Merkmalsaggregaten wie z.B. parallelen Linien, Eckenstrukturen etc. gebildet werden (vgl. [DPR92a, Ste97, FKL<sup>+</sup>98, Ste98]).

### 2.2.3 Strategien zur 3D-Objekterkennung

Die Strategien zur Erkennung von Gebäuden unterscheiden sich im wesentlichen darin, daß sie entweder datengetrieben (bottom-up), modellgetrieben (top-down) oder kombiniert daten- und modellgetrieben vorgehen (vgl. [SFH92, HS93a, BKL<sup>+</sup>95b, May98]).

*Bottom-Up-Strategien* basieren auf der schrittweisen Gruppierung von Bildmerkmalen zu Merkmalsaggregaten, die dann weiter zu Objektteilen und schließlich zu ganzen Objekten aggregiert werden. Das verwendete Objektmodell definiert dabei die Bedingungen, wann eine bestimmte Menge einzelner Merkmale zu einem komplexeren Objekt aggregiert werden darf. Der Vorteil dieser Strategie liegt im wesentlichen darin, daß zur Gruppierung nur lokale Kriterien, wie z.B. auf der untersten Modellebene die Parallelität zweier Linien oder die räumliche Nachbarschaft mehrerer Merkmale, berücksichtigt werden, und deshalb die Bestimmung der zu gruppierenden Elemente auf jeder Modellebene sehr effizient erfolgen kann. Dieses Vorgehen bringt jedoch die folgenden Probleme mit sich: Da zur Gruppierung immer nur Wissen der entsprechenden Modellebene verwendet wird, können erstens Gruppierungen von Merkmalen gebildet werden, die zufällig in das Schema passen, aber nicht zu einem gesuchten Objekt gehören. Zweitens können verschiedene, sich einander widersprechende Gruppierungen gebildet werden. Diese Widersprüche können zwar auf der nächsthöheren Hierarchieebene festgestellt werden, jedoch ist die Auflösung dieser Widersprüche problematisch, solange dazu nicht auf das gesamte Modellwissen zurückgegriffen werden kann. Dies ist aber erst auf der höchsten Aggregationsstufe zugänglich.

*Top-Down-Strategien* basieren auf der Suche von Instanzen der Modelle im Bild. Dabei werden die Bilder nach allen in Frage kommenden Modellen durchsucht. Das bedeutet, daß diese Strategie nur dann verwendbar ist, wenn erstens die verschiedenen Modelle aufzählbar sind, und zweitens ihre Anzahl dabei nicht zu groß wird. Aus diesem Grund ist für das verwendete Objektmodell nur eine stark eingeschränkte Generik erlaubt. Bei einer hierarchischen Objektmodellierung und Verwendung von symbolischen Bildmerkmalen stellt sich diese Strategie zudem als komplexes kombinatorisches Zuordnungsproblem dar, bei dem zu jedem Merkmal auf der untersten Hierarchieebene des Modells das entsprechende Bildmerkmal gefunden werden muß (vgl. [Vos92]). Der wesentliche Vorteil gegenüber der Bottom-Up-Strategie besteht darin, daß durch die erfolgreiche Identifikation eines Objekts im Bild auch unmittelbar bekannt ist, um welchen Modelltyp es sich handelt, d.h. es muß nicht mehr weiter klassifiziert werden.

Bei der Strategie der *Generierung und Verifikation von Hypothesen* handelt es sich um eine spezielle Realisierung der Top-Down-Strategie. Dabei wird zunächst datengetrieben nach signifikanten Beobachtungen in den Bilddaten gesucht, die jedoch unvollständig und auch widersprüchlich sein können. Anschließend wird versucht, global konsistente Instanzen des Objektmodells zu generieren, die diese Beobachtungen so gut wie möglich erklären können. Um festzustellen, ob die Beobachtungen tatsächlich von einem entsprechenden Objekt stammen, werden die Hypothesen danach modellgetrieben verifiziert. Dabei muß neben der Entscheidung, ob die Beobachtungen überhaupt zu einem Objekt gehören, im Falle mehrerer Hypothesen auch die Auswahl der zutreffendsten Hypothese getroffen werden.

## 2.3 Vergleichbare Forschungsansätze

In dem letzten Jahrzehnt wurde eine Reihe von Verfahren zur automatischen Extraktion von Gebäuden entwickelt und prototypisch implementiert. Als Datenquellen wurden dazu hauptsächlich digitale Luftbilder und digitale Oberflächenmodelle (DOM) benutzt. Die Liste der im folgenden Unterabschnitt vorgestellten Arbeiten ist nicht vollständig. Die Arbeiten wurden im Hinblick auf die Repräsentativität für bestimmte Modellierungen und Strategien ausgewählt. Sie sollen einen Überblick über die verschiedenen Vorgehensweisen zur automatischen Extraktion von Gebäuden aus Luftbildern geben. Einen umfassenden Überblick über die aktuellen Konzepte zur automatischen Gebäuderekonstruktion geben die Tagungsbände der beiden Ascona-Workshops „Automatic Extraction of Man-Made Objects from Aerial and Space Images“ in 1995 und 1997 ([GKA95, GBH97]) sowie des Bonner Workshops „Semantic Modeling for the Acquisition of Topographic Information from Images and Maps – SMATI’97“ ([FP97]).

In den Habilitationsschriften von Mayer ([May98]) und Steinhage ([Ste98]) werden die aktuellen Konzepte diskutiert und bewertet. Steinhage unterscheidet die Verfahren dabei primär anhand

der verwendeten Modellierung in vier Klassen: prismatische Polyedermodelle, allgemeine Polyedermodelle, parametrisierte Volumenprimitive und komponentenbasierte, generische Modelle. Die relevanten Arbeiten werden in den folgenden Unterabschnitten entsprechend dieser Einordnung vorgestellt und anschließend im Hinblick auf ihre Gebäudespezifität sowie Generik diskutiert. Dabei wird insbesondere untersucht, inwieweit die Verfahren eine Verifikation und Klassifikation der rekonstruierten Gebäudestrukturen ermöglichen und auch durchführen. Da die Modellierungen der vergleichbaren Ansätze nur in die ersten drei Kategorien fallen, wird auf die komponentenbasierte, generische Modellierung erst im Anschluß in Abschnitt 2.4 bei der Vorstellung der dieser Arbeit zugrundeliegenden Gebäudemodellierung eingegangen.

### 2.3.1 Prismatische Polyedermodelle

Prismatische Polyedermodelle erlauben die Repräsentation von Gebäuden mit flachen Dachstrukturen und variablem Grundriß. Dabei wird angenommen, daß der Grundriß des Gebäudes dem Umrißpolygon des Daches entspricht. Durch die Projektion des Polygons auf den Boden und die Annahme senkrechter Wände dazwischen entsteht ein prismatischer Polyeder.

Nevatia, Huertas, Lin und Mohan haben verschiedene Ansätze zur Extraktion von Gebäuden aus Luftbildern vorgestellt (siehe [NP82, HN88, MN88, LHN95, NLH97]), die sich hauptsächlich nur in der Art der Bildsegmentierung, Validierung der Hypothesen und in Einzelbild- oder Mehrbildverfahren unterscheiden. Allen Ansätzen gemein ist die Modellierung und größtenteils die Strategie, bei der im wesentlichen zuerst gerade Linien aus dem Bild bzw. den Bildern extrahiert werden. Diese werden dann anhand geometrischer Bedingungen wie Kollinearität, Orthogonalität, Parallelität und räumlicher Nähe erst zu L-, dann zu U-förmigen Strukturen und schließlich zu Rechtecken gruppiert. Bei dem Einzelbildansatz wird die Gebäudehöhe über die Schatteninformationen der Gebäude bestimmt, und bei den Mehrbildverfahren über die Zuordnung der Rechtecke in den verschiedenen Bildern. Wegen der Orthogonalitätsbedingung können jedoch nur Gebäude mit rechtwinkliger und flacher Dachstruktur erkannt werden. Einen zu diesem Konzept ähnlichen Ansatz beschreiben Collins, Hanson und Riseman in [CHRS95].

Bei dem in [FH87, FH89a] von Fua und Hanson vorgestellten Konzept wird, wie beim vorigen Ansatz, zunächst eine Kantenextraktion auf einem Luftbild durchgeführt. Nun werden nach den geometrischen Kriterien Orthogonalität und Kollinearität mögliche Kantengruppierungen durchgeführt. Anschließend werden anhand der Kantengruppen konsistente Zyklen konstruiert, die zu Polygonen geschlossen werden. Die Polygone stellen Hypothesen für die Konturlinien rechtwinkliger Dächer dar, die anschließend mit Hilfe eines an das Snakes-Konzept angelehnten Verfahrens verbessert werden (siehe [KWT88, SFH92]). Dabei werden die Konturlinien „gummibandartig“ mit Hilfe einer Energiefunktion flexibel an die Radiometrie eingepaßt. Auf diese Weise werden zum einen Fehler der Kantenextraktion behoben, und zum anderen die radiometrischen Eigenschaften der an die Linien grenzenden Flächen berücksichtigt. Da bei diesem Vorgehen für eine Szene verschiedene, sich überlappende und damit widersprechende Hypothesen entstehen können, wird zur Selektion für jede Hypothese auf der einen Seite ihre radiometrische Homogenität und auf der anderen Seite ihre geometrische Komplexität bewertet. Durch Bildung der Differenz der beiden Einzelbewertungen erfolgt eine Balancierung zwischen der radiometrischen Übereinstimmung und der Komplexität einer Dachhypothese. Die Hypothese mit der höchsten Gesamtbewertung wird schließlich als Dach interpretiert. Bei der Vorstellung des Konzepts wird zwar erwähnt, daß bei Vorliegen mehrerer Bilder diese zur Eliminierung falscher Dachhypothesen verwendet werden können, aber es wird nicht näher erläutert, wie die in verschiedenen Bildern extrahierten und u.U. voneinander abweichenden Polygone einander zugeordnet werden, um eine 3D-Rekonstruktion durchführen zu können.

Ein anderes Konzept zur Extraktion von Gebäuden mit prismatischer Struktur wurde von Weidner et. al. entwickelt (siehe [WF95, Wei95, Wei97a, Wei97b]). Dabei werden als Datenquelle keine Luftbilder, sondern ein digitales Oberflächenmodell (DOM) der Szene verwendet. Gebäude stellen sich in diesem Oberflächenmodell als signifikante Erhebungen dar. Diese werden in einem ersten

Schritt ausgewählt. Mittels eines Hörschwelldwertes wird dann ein Binärbild erzeugt, bei dem alle Pixel schwarz sind, die über dem Schwellwert liegen. Eine Kantendetektion auf diesem Binärbild liefert dann den Umriß möglicher Gebäude. Da Gebäude meistens glatte Umrisse besitzen, wird zur Ausgleichung der Unsicherheit und des Rauschens versucht, die einzelnen Kantenpixel durch ein möglichst einfaches Polygon mit wenigen Linien zu approximieren. Dabei gehen geometrische Bedingungen wie Kollinearität, Orthogonalität und Parallelität explizit ein (vgl. [BWF95])<sup>1</sup>. In seiner Dissertation erweitert Weidner den Ansatz um das parametrisierte Modell eines Satteldachhauses (vgl. [Wei97b]). Das Verfahren ist dabei in der Lage, aus der Hypothese eines prismatischen Modells mit beliebigem Grundriß sowie flacher Dachstruktur und der Hypothese für ein Satteldachhaus die wahrscheinlichere zu bestimmen.

### 2.3.2 Allgemeine Polyedermodelle

Allgemeine Polyedermodelle zeichnen sich durch eine große Generik sowohl in bezug auf die Form als auch die Ausdehnung aus. Es werden a priori keine Beschränkungen auf bestimmte Formen gemacht. Darüberhinaus können auch unvollständige, d.h. nicht geschlossene Polyeder repräsentiert werden. Die Repräsentation dieser polyedrischen Strukturen basiert auf der Randflächendarstellung (B-Rep).

Henricsson et. al. stellen in [Hen95, BHFS96, Hen96] ein Mehrbildverfahren zur Extraktion von Gebäuden mit beliebigen polyedrischen Dachstrukturen vor. Zur Einschränkung der Bildbereiche, in denen nach Dächern gesucht wird, werden zunächst aus dem zur der Szene gehörenden digitalen Oberflächenmodell signifikante Erhebungen extrahiert. Wenn sich die Farbwerte der einzelnen digitalisierten Bilder an den entsprechenden Stellen ausreichend von denen der Umgebung unterscheiden, wird davon ausgegangen, daß sich in diesem Bereich ein Gebäude befindet. Für diese Bereiche wird eine Kantenextraktion durchgeführt, wobei zu jeder Kante die radiometrischen und chromatischen Eigenschaften der Flächen zu beiden Seiten der Kante vermerkt werden. Nun wird für die Kanten in einem Bild nach Kanten mit entsprechenden Charakteristiken in den anderen Bildern entlang der Epipolarlinien gesucht, und aus einander zugeordneten Kanten jeweils eine 3D-Kante berechnet. Koplanare 3D-Kanten werden dann gruppiert und zur Erzeugung einer 3D-Flächenhypothese verwendet. Durch Rückprojektion in die Bilder und Betrachtung der lokalen Radiometrie wird dabei versucht, geschlossene Konturlinien zu bilden. Die abgeschlossenen Bereiche werden anhand von Kriterien wie Einfachheit der Form, Flächeninhalt und 3D-Vollständigkeit bewertet und sortiert. Es werden nach absteigender Bewertung Flächen als Dachflächen ausgewählt, und diese mit ihren benachbarten Flächen zu Dächern aggregiert. Nach Abschluß der Aggregationen wird die Dachhöhe aus der Höhendifferenz des Daches und dem umliegenden Geländemodell geschätzt. Unter der Annahme senkrechter Wände wird ein Polyeder mit ebener Grundfläche und der gruppierten Dachstruktur rekonstruiert.

Der Ansatz von Frère et. al. ([FHV<sup>+</sup>97]) baut im wesentlichen auf der Arbeit von Henricsson et. al. auf, wobei ein anderer Weg zur Konstruktion der 3D-Flächen beschrritten wird. Zunächst wird auf allen Bildern eine Kantendetektion durchgeführt. Danach wird eine sogenannte Constraint-Triangulierung durchgeführt, bei der größtmögliche Dreiecke mit homogener Radiometrie aus dem Bild segmentiert werden, wobei die Constraints darin bestehen, daß die vorher extrahierten Kanten in dem Dreiecksnetz enthalten sein müssen. Nach dem Prinzip des Regionenwachstums werden nun benachbarte Dreiecke mit ähnlichen chromatischen und radiometrischen Eigenschaften miteinander verschmolzen. Im Unterschied zum Ansatz von Henricsson et. al. werden auf diese Weise von vornherein geschlossene Flächen generiert. Der Rest der Rekonstruktion entspricht weitestgehend dem obigen Ansatz.

Einen anderen Weg zur Rekonstruktion polyedrisch modellierter Gebäude geht Teleki in seiner Dissertation ([Tel97]). Die Arbeit setzt nicht mehr unmittelbar auf Luftbildern auf, sondern verwendet als Ausgangsdaten bereits dreidimensionale Eckenstrukturen, die, wie z.B. von Lang und Förstner in [LF96b] beschrieben wird, durch eine Mehrbildzuordnung gewonnen werden können. Es wird nun versucht, die Menge der vorliegenden Eckenstrukturen durch Polyeder zu erklären.

---

<sup>1</sup>Auf das Verfahren zur Polygonbestimmung wird unten in Abschnitt 6.1.3 noch näher eingegangen.

Dazu erhalten die zu den Ecken inzidenten Kanten qualitative Attribute, die ihre grobe Richtung im Objektraum benennen. Wenn zwei Kanten verschiedener Ecken zueinander zeigen (was anhand der qualitativen Richtung überprüft wird), und sie geometrisch ungefähr kollinear sind, wird eine gemeinsame Verbindungskante zwischen den beiden Ecken angenommen. Flächen werden dann aus Verbindungskanten konstruiert, die sich nicht überschneiden. Jede konstruierte Fläche wird schließlich noch auf Konsistenz bzgl. der bereits rekonstruierten Flächen überprüft. Da bei diesem Verfahren nur die vorhandenen Eckenstrukturen einbezogen werden, und keine Hypothesen für evtl. fehlende Ecken generiert werden, sind die Polyeder u.U. nicht geschlossen.

### 2.3.3 Parametrisierte Volumenprimitive

Parametrisierte Volumenprimitive zeichnen sich durch eine feste Form und variable Ausdehnung aus. Gebäude werden als geschlossene geometrische Primitive wie z.B. Würfel, Quader, oder auch als vollständiges Satteldachhaus modelliert, wobei sich die Koordinaten der Eckpunkte als Funktion der Gebäudeparameter darstellen lassen.

McGlone und Shufelt zeigen in [MS94, Shu96] ein Konzept zur Erkennung von Gebäuden aus Einzelbildern unter intensiver Nutzung eines Kamera- und Beleuchtungsmodells. Es wird dabei nur zwischen quaderförmigen Gebäuden und Satteldachhäusern unterschieden. Aus dem Luftbild werden zunächst gerade Liniensegmente extrahiert. Danach wird versucht, die Orientierung der Linien im 3D-Objektraum qualitativ über die möglichen Fluchtpunkte zu rekonstruieren. Für die Linien, die bzgl. der Modelle konsistente Orientierungen aufweisen, werden Gebäudehypothesen generiert. Die Gebäudehöhe wird dabei entweder anhand eines DGMs oder aus der Länge des Schattens ermittelt. Bei der anschließenden Verifikation werden für alle Hypothesen die erwarteten Umrisse ihrer Schatten berechnet, und diejenigen Hypothesen eliminiert, deren Schattenflächen heller sind als die Umgebung. Ferner werden die Hypothesen eliminiert, bei denen die sonnenzugewandten Flächen nicht (wie erwartet) heller sind, als die von der Sonne abgewandten Flächen. Alle restlichen Hypothesen werden bzgl. der Flächenhomogenität, der Helligkeitsverhältnisse benachbarter Flächen und dem Helligkeitsverhältnis des Schattens zur Umgebung bewertet und sortiert. Alle Hypothesen, die sich räumlich mit der bestbewerteten Hypothese überlappen, werden eliminiert. Dieser Schritt wird dann für die zweitbeste, drittbeste etc. wiederholt, bis eine räumlich konsistente Szenenbeschreibung vorliegt.

Jaynes, Hanson und Riseman stellen in ihrem Artikel [JHR97] ein Konzept zur Identifikation von Dachtypen anhand eines digitalen Oberflächenmodells vor. Das Verfahren setzt auf dem in [CHRS95] beschriebenen Ansatz zur Rekonstruktion prismatischer Gebäude aus Luftbildern auf, der eine ähnliche Strategie wie das oben beschriebene Verfahren von Mohan und Nevatia verfolgt. Für die rekonstruierten Gebäude, für die aufgrund des verwendeten prismatischen Modells zunächst nur ein Flachdach angenommen wurde, wird nun der entsprechende Ausschnitt des DOMs ausgeschnitten, und darauf eine Delaunay-Triangulation berechnet (siehe [PS85]). Es wird dann ein Orientierungshistogramm der Flächennormalen der Dreiecke erstellt. Dabei wird eine Einheitskugel im Zentrum eines lokalen 3D-Koordinatensystems aufgespannt und die Oberfläche dieser Kugel in gleichgroße Flächen partitioniert. Für jedes bei der Triangulation entstandene Dreieck wird nun geschaut, welche Kugelteilfläche durchstoßen wird, wenn der Flächennormalenvektor in den Ursprung des Koordinatensystems verschoben und entsprechend verlängert wird. Für die durchstoßene Teilfläche wird ein ihr zugeordneter Zähler erhöht. Auf diese Weise entsteht ein charakteristischer „Fingerabdruck“ des Daches, der unmittelbar mit denen verglichen wird, die vorher für eine Reihe modellierter Dachformen berechnet wurden. Die Auswahl der zutreffendsten Dachform erfolgt anhand der Ergebnisse der Kreuzkorrelationen der Gebäudehypothese mit allen bekannten Modellen. Abschließend werden die entsprechenden Dachparameter geschätzt. Der Vorteil der Modellauswahl über das Orientierungshistogramm besteht darin, daß auch runde Dächer und Dächer mit gekrümmten Flächen erkannt werden können. Der Nachteil besteht darin, daß es zu konvexen Dächern konkave Gegenstücke gibt, die dasselbe Histogramm aufweisen und somit nicht voneinander unterschieden werden können.

Der Ansatz von Schutte et. al. [SB93, Sch96, SSH97a, SSH97b] basiert auf parametrischen Gebäudeprimitiven, wobei bislang nur quaderförmige Gebäude und Satteldachhäuser verwendet werden. Um die verschiedenen Erscheinungsformen der Modelle zu berücksichtigen, wird zu den Modellen a priori jeweils der Aspektgraph berechnet. Vor der Suche wird das Luftbild zuerst mittels eines Regionenwachstumsverfahrens segmentiert, wobei homogene Flächen und ihre Nachbarschaftsrelationen extrahiert werden. Für jede extrahierte Fläche wird ein Umringspolygon approximiert, und die Eckpunkte in L- oder T-förmig klassifiziert. Nun werden aus der Menge der Modellaspekte diejenigen ausgewählt, die zu den extrahierten Flächen passen. Dies wird anhand logischer Regeln überprüft: die Anzahl der Eckpunkte muß für jede Fläche eines Aspekts ungefähr mit der Anzahl der Eckpunkte der zugeordneten extrahierten Fläche übereinstimmen, und jede Fläche muß eine bestimmte Lage zu anderen besitzen, wie z.B. links oder rechts davon. Für die ausgewählten Aspekte wird eine erste Parameterschätzung vorgenommen, wobei zur 3D-Rekonstruktion der angrenzende Schatten herangezogen wird. Die Gebäudehypothesen werden dann zurück ins Bild projiziert und die Überdeckungen der einzelnen Modellflächen zu den extrahierten Flächen gemessen. Anhand der Abweichungen werden die geschätzten Parameter adaptiert und eine neue Rückprojektion berechnet. Das ganze wird solange iteriert, bis die Abweichungen minimal werden. Aus allen Hypothesen wird schließlich diejenige mit dem kleinsten (über die Anzahl der Beobachtungen normierten) Residuum der Parameterschätzung ausgewählt.

### 2.3.4 Diskussion der Konzepte

Die prismatischen und allgemeinen Polyedermodelle erlauben eine mittel bis stark detaillierte Rekonstruktion von Gebäuden. Durch die wenig eingeschränkte Form allgemeiner Polyeder kann eine Vielzahl von Gebäuden erfaßt werden. Da aber neben der räumlichen, geometrischen und bei Henricsson et. al. auch radiometrischen Konsistenz keine semantischen Kriterien abgefragt werden, sind die Modelle insgesamt wenig gebäudespezifisch. Die darauf basierenden Verfahren zur Gebäudeextraktion sind bis auf eine Ausnahme — den Ansatz von Weidner — deshalb weder in der Lage, ein rekonstruiertes 3D-Objekt als Gebäude zu klassifizieren, noch verschiedene Gebäudetypen zu benennen (siehe [HB97]). Da alle Verfahren eine Bottom-Up-Strategie verfolgen<sup>2</sup>, bei der nur beobachtete Bildmerkmale berücksichtigt werden, erfordern sie eine hohe Datenqualität. Ferner dürfen Gebäude nicht durch andere Objekte wie z.B. Bäume teilverdeckt sein, weil das Gebäude sonst entweder gar nicht erkannt wird, oder das rekonstruierte 3D-Modell an den entsprechenden Stellen unregelmäßig geformte Aussparungen aufweist.

Die bei den Gruppierungen entstehenden widersprüchlichen Hypothesen werden bei allen vorgestellten Bottom-Up-Verfahren heuristisch auf der jeweiligen Aggregationsebene aufgelöst, um dem Gruppierungsprozeß der nächsten Modellebene eine konsistente Menge von Beobachtungen bereitzustellen. Das Problem bei diesem Vorgehen besteht darin, daß u.U. bei der Selektion korrekte Gruppierungen als fehlerhaft aussortiert werden können, und deshalb die Rekonstruktion des abgebildeten Objekts schließlich unvollständig bzw. suboptimal sein kann<sup>3</sup>.

Im Gegensatz dazu erfolgt die Klassifikation von Gebäuden bei den Verfahren, die parametrisierte Volumenprimitive verwenden, implizit mit der Identifikation einer Modellinstanz im Bild. Die vom Gebäudetyp abhängigen Parameter können danach einfach durch eine Parameterschätzung ermittelt werden. Da das Aussehen, insbesondere die Form der Modelle bekannt ist, können Gebäude auch dann vollständig rekonstruiert werden, wenn einzelne Teile aufgrund von Verdeckungen durch andere Objekte oder eine schräge Perspektive unbeobachtbar sind. Der Detailgrad parametrisierter Volumenmodelle ist gering, weil für jede Variante ein eigenes Modell benötigt wird, und die

<sup>2</sup>Damit ist die Hauptstrategie der Ansätze gemeint, die (prismatische) Polyedermodelle verwenden. Zwar werden auf den einzelnen Modellierungsebenen auch Gruppierungshypothesen gebildet, und die nach lokalen Kriterien jeweils besten ausgewählt (verifiziert), jedoch führt der Weg zur vollständigen Extraktion von Gebäuden eindeutig von den Daten zum Modell.

<sup>3</sup>Ein Ansatz zur Bildinterpretation (in einem anderen Kontext), der widersprüchliche Aggregationen explizit modelliert und diese erst auf der höchsten Modellierungsebene auflöst, wird von Bringmann et. al. in [BBS97] vorgestellt.

wenigen Modelle auf möglichst viele Gebäude passen sollen. Hier liegt auch das Hauptproblem parametrisierter Volumenprimitive, denn es können mangels Generik nur Gebäude erkannt werden die a priori explizit modelliert wurden.

Alle vorgestellten Konzepte, denen parametrisierte Volumenprimitive zugrundeliegen, basieren auf der Strategie der Generierung und Verifikation von Hypothesen, wobei die Hypothesen bei McGlone und Shufelt sowie bei Schutte et. al. auf der Basis extrahierter Bildmerkmale generiert werden. Jaynes et. al. verwenden hingegen extrahierte Gebäude, für die bislang Flachdächer angenommen wurden, als Ausgangspunkt für Gebäudehypothesen mit andersartiger Dachstruktur. Die Verfahren von McGlone und Shufelt sowie Schutte et. al. unterscheiden bislang nur zwischen quaderförmigen Gebäuden, wie z.B. Bungalows oder Hochhäusern, und Satteldachhäusern. Jaynes et. al. unterscheidet zwar zwischen einer Reihe verschiedener Dachformen, beschränkt sich jedoch auf Gebäude mit rechteckigen Grundrissen. Bei der Verifikation werden in allen drei Ansätzen unterschiedliche Bewertungen verwendet. Schutte et. al. klassifizieren anhand des Residuums der Parameterschätzung. Dieses wird zwar über die Anzahl der Beobachtungen normiert und ist damit nicht von der absoluten Ausdehnung des Modells abhängig, jedoch ist es bei ähnlichen Modellen möglich, daß sie gleich gut ins Bild eingepaßt werden können, so daß diese Bewertung insgesamt als zu einfach einzustufen ist. McGlone und Shufelt verwenden eine Bewertung, in die verschiedene Faktoren unterschiedlich gewichtet eingehen, wie z.B. die radiometrische Homogenität der Flächen und der Kontrast zwischen Gebäude- und Schattenflächen. Da erstens nicht klar ist, wo die Gewichte herkommen und zweitens keine Normierung in bezug auf die Modellgröße erwähnt wird, werden wahrscheinlich komplexere Modelle generell bevorzugt. Jaynes et. al. wählen die beste Hypothese anhand einer Kreuzkorrelation aus, jedoch ist die auf dem Orientierungshistogramm basierende Bewertung nicht für komplexe Gebäude geeignet, weil sie erstens nicht zwischen konkaven und konvexen Dächern unterscheiden kann, und zweitens einen einfachen Grundriß voraussetzt.

**Fazit:** Die Verfahren, die (prismatische) Polyedermodelle verwenden, ermöglichen eine große Generik in bezug auf die Form der rekonstruierbaren Gebäude, berücksichtigen dabei aber zu wenig gebäudespezifische Semantik. Die Ansätze, die auf parametrisierten Volumenprimitiven basieren, erlauben zwar die Klassifikation abgebildeter Gebäude, jedoch sind sie auf die Erkennung weniger, fest modellierter Gebäudetypen beschränkt. Die bei der Verifikation von Gebäudehypothesen verwendeten Bewertungsmaße lassen sich darüberhinaus nicht für die Bewertung komplexerer Gebäude verwenden.

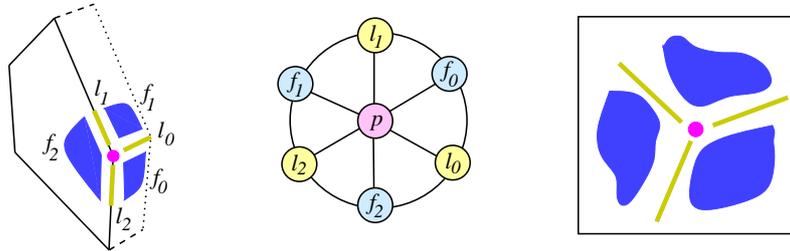
## 2.4 Verwendete Modellierung

In diesem Abschnitt wird das in dieser Arbeit zur automatischen Gebäudeerkennung verwendete Gebäude- und Bildmodell vorgestellt. Dieses wurde in dem DFG-Bündelprojekt „Semantische Modellierung“, insbesondere im Projekt „Extraktion von Gebäuden aus Digitalen Bildern“, in Kooperation der Institute für Informatik I und III und dem Institut für Photogrammetrie der Universität Bonn sowie dem Institut für Umweltwissenschaften der Hochschule Vechta entwickelt (siehe [FKL97, FKL+98]). Es handelt sich dabei um ein komponentenbasiertes, generisches Modell, das auf Gebäudeteilprimitiven basiert, die nach wohldefinierten und gebäudespezifischen Regeln miteinander zu Gebäudehypothesen kombiniert werden können. Auf diese Weise wird eine Formgenerik realisiert, die aber auf die Konstruktion von Gebäudestrukturen beschränkt bleibt. Die Aufteilung der Gebäude in einzelne Teile erfolgt auf eine Weise, die erstens eine Klassifikation in Gebäudebereiche wie Dach und Wand, und zweitens in Haupt-, Seiten und Nebentrakt ermöglicht. Dabei wird eine gebäudetypische Parametrisierung, wie z.B. Breite, Höhe bis zum Dach, Dachhöhe oder auch Dacheinzugstiefe eines Walmdaches, verwendet. Die Verwendung dieses Modells erlaubt damit auch die Klassifikation von Gebäuden und Gebäudeteilen.

Das Modell ist eingebettet in ein — ebenfalls in dem DFG-Projekt entstandenen — Rahmenkonzept zur photogrammetrischen Gebäuderekonstruktion (siehe [BKL+95a, BKL+95b]). Das dort beschriebene Rahmenmodell ordnet das Gebäudemodell noch in ein darüberliegendes Szenenmodell ein, auf das hier jedoch nicht weiter eingegangen wird.

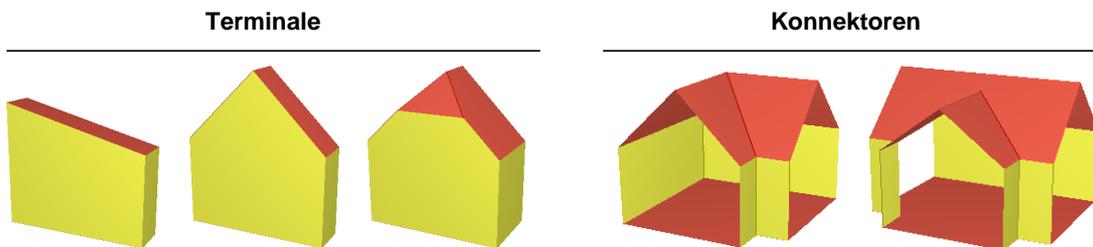


(s-), Symmetrie(s-,s-)} beschrieben. Es wurde festgestellt, daß 21 Kombinationen ausreichen, um alle Eckenstrukturen (mit drei Linien) der bislang modellierten Gebäudeteile zu repräsentieren (vgl. [Eng97b]).



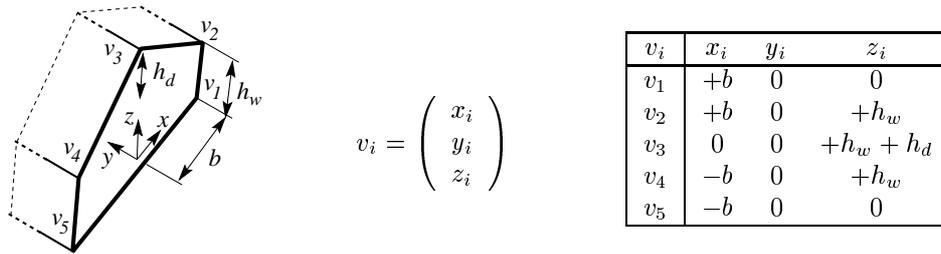
**Abbildung 2.3:** Modellierung des Merkmalsaggregats *Ecke*. Eine Ecke wird aus einem Knotenpunkt, den von ihm ausgehenden Linien und den dazwischen aufgespannten Flächen gebildet (linkes Bild). Die Anordnung der Merkmale besitzt eine typische Nachbarschaftsstruktur (Mitte), die auch in den durch eine Segmentierung extrahierten Bildmerkmalen beobachtbar ist (rechtes Bild).

Die dritte Modellierungsebene beschreibt *Gebäudeteile* durch parametrisierte Volumenprimitive. Die Gebäudeteile besitzen ein oder mehrere offene Enden, an denen sie mit anderen Gebäudeteilen verbunden werden können. Die offenen Enden werden durch sogenannte Verbindungsflächen abgeschlossen, wobei die Form dieser Flächen darüber entscheidet, ob zwei Gebäudeteile zueinander kompatibel sind und miteinander verschmolzen werden dürfen. Anhand der Anzahl der Verbindungsflächen werden die Teile in zwei Klassen unterschieden: Gebäudeteile mit genau einer Verbindungsfläche werden *Terminale* genannt, und Teile mit mit zwei oder mehr Verbindungsflächen entsprechend *Konnektoren*. Abbildung 2.4 zeigt einige Beispiele für Terminale und Konnektoren, wobei die Verbindungsflächen bei den Konnektoren transparent, d.h. nicht sichtbar sind.



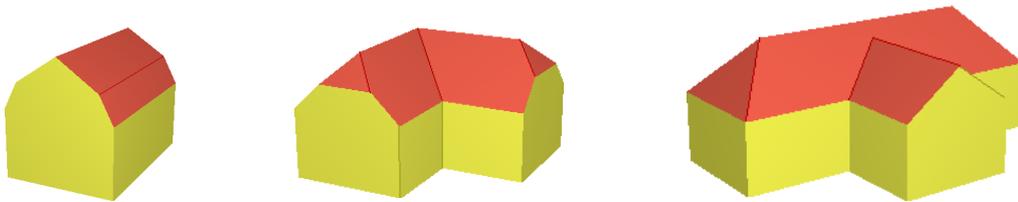
**Abbildung 2.4:** Beispiele für Gebäudeteile. Links: drei Endstücke (Pultdach-, Satteldach- und Krüppelwalmdachhaus). Rechts: zwei Verbinder (L-förmiges Satteldach- und T-förmiges Satteldachhaus).

Die Repräsentation der Gebäudeteile basiert auf einem Graph, dessen Knoten die Gebäudeecken repräsentieren und dessen Kanten beschreiben, welche von einer Ecke ausgehenden Linien mit welchen anderen verbunden sind. Die Koordinaten der Eckpunkte werden in Abhängigkeit von gebäudetypischen Parametern ausgedrückt. Abbildung 2.5 zeigt, wie die Eckpunktkoordinaten eines Satteldachhausterminals durch die drei Formparameter Breite, Höhe bis zum Dach und Dachhöhe ausgedrückt werden. Zusätzlich sind zu jedem Gebäudeteil Bedingungen auf den Parametern formuliert, wie z.B. daß die Gebäudebreite größer als 3 m, die Dachneigung eines Satteldachhauses nicht größer als 70° oder die Gebäudehöhe nicht kleiner als 2,5 m sein darf. Diese Bedingungen werden später bei der Verifikation von dreidimensional rekonstruierten Gebäuden abgeprüft, und helfen dabei zu vermeiden, daß — wie in der Einleitung erwähnt — eine Hundehütte o.ä. als Haus interpretiert wird.



**Abbildung 2.5:** Parametrisierung der Gebäudeteile (aus [FKL+98]). Die Koordinaten der fünf Eckpunkte  $v_1, \dots, v_5$  des Satteldachterminals werden mittels der drei Formparameter  $b$  (halbe Breite),  $h_w$  (Wandhöhe bis zum Dachanfang) und  $h_d$  (Dachhöhe) ausgedrückt.

Die höchste Aggregationsstufe umfaßt die vollständigen *Gebäude*, welche ebenfalls durch eine Graphstruktur dargestellt werden. Die Knoten des Graphen repräsentieren die Gebäudeteile, und die Kanten beschreiben welches Teil mit welchen anderen verschmolzen wurde. Da Gebäude vollständig geschlossen sein müssen, d.h. keine freien Verbindungsflächen mehr aufweisen dürfen, bestehen die einfachsten Gebäude aus der Verbindung zweier Terminals. Abbildung 2.6 zeigt drei Beispiele für unterschiedlich komplexe Gebäude, die sich mit der vorgestellten Modellierung konstruieren lassen.



**Abbildung 2.6:** Beispiele für Gebäude, die sich aus einfachen Konnektoren und Terminals zusammensetzen lassen (von links nach rechts: Mansarddachhaus, L-förmiges Krüppelwalmdachhaus und T-förmiges Walm-/Satteldachhaus).

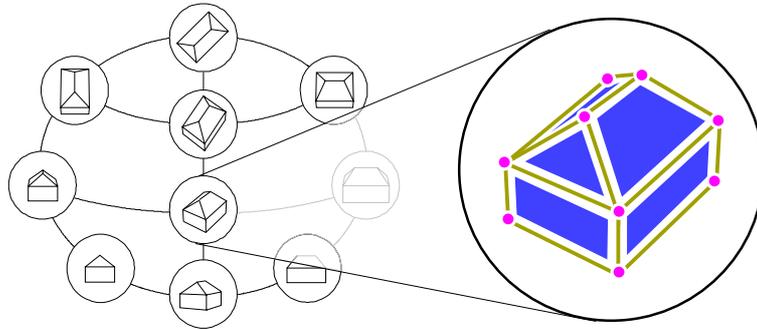
Damit die modellierten Gebäudeteile auf möglichst viele verschiedene Gebäude in den Luftbildern passen, wird bislang keine detailliertere Dachmodellierung verwendet. Dachaufbauten wie z.B. Schornsteine und Gauben sowie Dachvorsprünge werden nicht modelliert, um die Komplexität des Modells zu begrenzen.

Insgesamt orientiert sich die Objektmodellierung sowohl an dem Prinzip der konstruktiven Festkörpergeometrie als auch der Randflächendarstellung, weil zur Komposition von Gebäuden auf der einen Seite parametrisierte Primitive verwendet werden, die nach definierten Regeln zu geschlossenen Gebäuden verbunden werden, und auf der anderen Seite der Eckengraph der Gebäudeteile die Randflächen der Modelle beschreibt.

## 2.4.2 Verbindung von Objekt- und Bildmodell

Zur Verbindung des dreidimensionalen Objektmodells mit dem zweidimensionalen Bildmodell wird eine Aspekthierarchie verwendet, die analog zum Objektmodell aus vier Aggregationsstufen besteht. Die oberste Ebene der Aspekthierarchie enthält dabei die verschiedenen Aspekte ganzer Gebäude. Die Dekomposition der Aspekte in Aspektteile, Merkmalsaggregate und schließlich einzelne Merkmale realisiert die Verknüpfung zwischen dem 3D-Objektmodell und dem 2D-Bildmodell auf den anderen drei Ebenen des Gesamtmodells (vgl. Abbildung 2.2 auf Seite 19). Die Merkmale der untersten Ebene umfassen Punkte, Linien und Flächen, wobei Linienmerkmale immer durch die gerade Strecke zwischen zwei Punktmerkmalen aufgespannt werden, und Flächen durch ein

Umringspolygon beschrieben werden. Ein Umringspolygon besteht dabei aus einem einfachen Linienzyklus, d.h. aus einem geschlossenen, über Knotenpunkte miteinander verketteten und überschneidungsfreien Linienzug. Die in einem Punkt beginnenden Linienmerkmale werden als *inzident* zu diesem Punkt bezeichnet. Abbildung 2.7 zeigt die Dekomposition für einen Aspekt des Aspektgraphen eines Walmdachhauses in die einzelnen Modellmerkmale.



**Abbildung 2.7:** Aspektgraph eines Walmdachhauses und Dekomposition eines Aspekts in die einzelnen Modellmerkmale. Jeder Aspekt besteht aus Punkten, Linien und Flächen, wobei Linien immer von zwei Punkten aufgespannt und Flächen immer durch einen einfachen Linienzyklus begrenzt werden.

### 2.4.3 Akquisition der Gebäudemodelle

Bislang wurden die Gebäudemodelle manuell erstellt. Dazu wurden die Gebäudeteile mit Hilfe eines von Steinhage und Fischer entwickelten Systems zur hybriden CSG-/B-Rep-Modellierung konstruiert (siehe [FS97b]). Es werden z.Zt. Gebäudeendstücke mit Flach-, Pult-, Sattel-, Walm-, Krüppelwalm- und Mansarddach verwendet. Dazu wurden auch jeweils die passenden L-, T- und X-förmigen Konnektoren modelliert.

Einen Anhaltspunkt dafür, mit wie wenigen Gebäudeteilen bereits große Teile der in Kleinstädten vorkommenden Haustypen modelliert werden können, liefert eine Studie von Englert, in der ein Verfahren zur Feststellung von Gebäudesymmetrien vorgestellt und anschließend empirisch evaluiert wird ([Eng97a]). Die dazu durchgeführte manuelle Erfassung der Gebäude der Kleinstadt Oedekoven in der Nähe von Bonn lieferte 1846 Gebäudestrukturen. Der Arbeit läßt sich entnehmen, daß bereits die Hälfte dieser Gebäude nur durch die Kombination der drei Terminale Flachdach-, Satteldach- und Walmdachendstück sowie der entsprechenden L- und T-förmigen Konnektoren zwischen diesen Terminalen dargestellt werden können.

Zukünftig sollen die Gebäudemodelle automatisch unter Verwendung von Methoden des maschinellen Lernens akquiriert werden (vgl. [RN95]). Dabei geht es insbesondere um die Identifikation von geeigneten Gebäudeteilen. Englert stellt dazu in [EC97, Eng98] ein Verfahren vor, welches in bereits erfaßten Stadtmodellen wiederkehrende Gebäudeteilstrukturen identifiziert. Zur Bestimmung einer geeigneten und kompakten Parametrisierung von Gebäudeteilen wird darüberhinaus das Wissen um Gebäudesymmetrien benötigt. Ein entsprechendes Konzept wird dazu von Englert in [Eng97a, Eng97a] vorgestellt.

## 2.5 Strategie zur Gebäudeextraktion

Die in dieser Arbeit verwendete Gesamtstrategie zur Identifikation und Rekonstruktion von Gebäuden in Luftbildern basiert auf dem Prinzip der Generierung und Verifikation von Hypothesen (vgl. Abschnitt 2.2.3). Im einzelnen besteht sie aus den folgenden sieben Teilschritten, wobei die

Schritte 1-5 die Generierung von Hypothesen realisieren, Schritt 6 die Verifikation und Schritt 7 schließlich die 3D-Rekonstruktion:

1. Bildsegmentierung und Merkmalsextraktion.
2. Bestimmung der Bildbereiche, in denen vermutlich Gebäude abgebildet sind, und Fokussierung der Suche auf die entsprechenden Bildmerkmale. Die folgenden Schritte werden für jeden dieser Bildbereiche separat ausgeführt.
3. Rekonstruktion dreidimensionaler Eckenstrukturen unter Verwendung des Eckenmodells. Hier erfolgt der Übergang von den 2D-Bilddaten in den 3D-Objektraum.
4. Generieren von Gebäudehypothesen, die die beobachteten Ecken erklären können.
5. Berechnung der möglichen zweidimensionalen Ansichten der Gebäudehypothesen im Bild.
6. Bestimmung der bestmöglichen Zuordnung für jede Gebäudeansicht zu den extrahierten Bildmerkmalen und Zurückweisung von Hypothesen mit zu schlechter Bewertung. Anschließend wird die wahrscheinlichste der verbliebenen Hypothesen bestimmt und ausgewählt.
7. Schätzung der Gebäudeparameter.

Dieses Vorgehen entspricht weitestgehend der in [FKL97] und [FKL<sup>+</sup>98] vorgestellten Strategie zur Gebäudeextraktion. Für die Realisierung der Strategie ist aus mehreren Gründen eine Mehrbildverarbeitung notwendig: Erstens erfolgt die Bestimmung in Frage kommender Bildbereiche auf Basis eines digitalen Oberflächenmodells, das — falls es nicht bereits durch eine Befliegung mit Laser-Abtastung erfaßt wurde — mit Hilfe von Stereobildverfahren generiert werden muß (vgl. [CHRS95, Wei97b]). Zweitens erfolgt die 3D-Rekonstruktion der Gebäudeecken auf der Basis eines Mehrbildzuordnungsverfahrens. Das bedeutet, daß die zu untersuchenden Bildszenen in mindestens zwei Luftbildern verschiedener Perspektive abgebildet sein müssen. Dies stellt jedoch keine bedeutende Einschränkung dar, weil in der Praxis zu fast allen Luftbildszenen Stereoaufnahmen existieren, und bei neuen Befliegungen die Szenen oftmals sogar vierfach überdeckt sind.

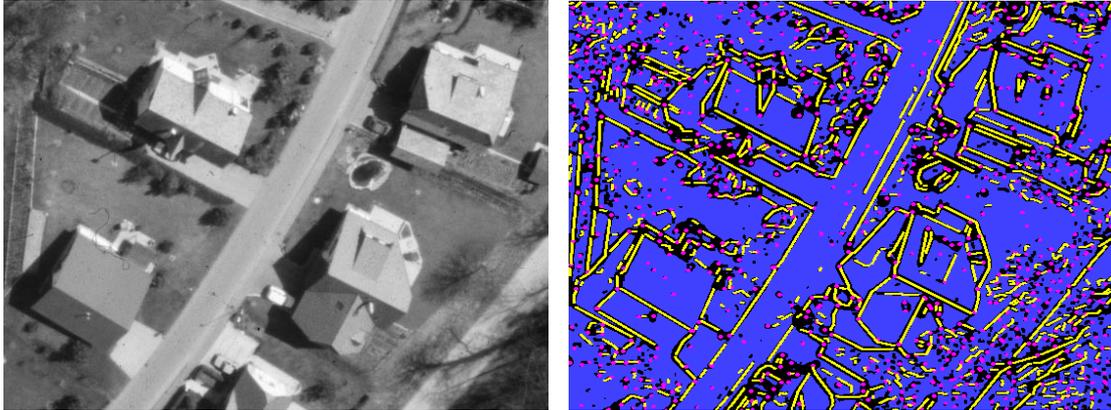
Die einzelnen Schritte werden entsprechend der obigen Reihenfolge und Numerierung in den folgenden Unterabschnitten erläutert. Dabei stellen die Merkmalsextraktion, die 3D-Rekonstruktion von Gebäudeecken, die Generierung von Gebäudehypothesen sowie die Herleitung der Ansichtenhierarchien eigene, komplexe Problembereiche dar, die jeweils Themen anderer Forschungsarbeiten sind. In dieser Arbeit werden sie deshalb nur soweit erläutert, wie es zur Verdeutlichung der Strategie notwendig ist. Es wird an den entsprechenden Stellen durch Referenzen auf die weiterführenden und detaillierteren Aufsätze verwiesen.

Diese Arbeit konzentriert sich auf den Problembereich der Verifikation der Gebäudehypothesen. In Unterabschnitt 2.5.6 werden deshalb die damit verbundenen Schwierigkeiten im Detail erörtert, woraus sich schließlich die Aufgabenstellungen für die folgenden Kapitel dieser Arbeit ergeben.

### 2.5.1 Bildsegmentierung und Merkmalsextraktion

Wie die anderen vergleichbaren Ansätze zur automatischen Gebäudeerkennung setzt auch das in dieser Arbeit beschriebene Konzept auf einer symbolischen Bildbeschreibung in Form extrahierter Bildmerkmale auf (vgl. Abschnitt 2.3). Da die unterste Ebene des Bildmodells aus Punkten, Linien und Flächen besteht, wird zur Herleitung dieser symbolischen Bildbeschreibung das von Fuchs und Förstner entwickelte Verfahren zur Bildsegmentierung und Extraktion polymorpher Merkmale eingesetzt (siehe [För94, Fuc97]). Dieses Verfahren klassifiziert alle Bildpunkte des digitalisierten Luftbildes entweder als Bestandteil von Punkt-, Linien- oder Flächenmerkmalen, oder weist sie im Fall zu großer Unsicherheit als unklassifizierbar zurück. Damit wird das Bild zunächst vollständig partitioniert. Mit der Zusammenfassung von Linienpixeln zu geraden Liniensegmenten

und Flächenpixeln zu radiometrisch homogenen Flächen entsteht als Ergebnis eine konsistente und überlappungsfreie Menge von Bildmerkmalen. Die Merkmale sind darüberhinaus in einen Merkmalsnachbarschaftsgraph (MNG) angeordnet, der aufzählt, welche Merkmale unmittelbar an welche anderen angrenzen. Abbildung 2.8 zeigt ein typisches Beispiel für das Ergebnis einer Merkmalsextraktion für einen Luftbildausschnitt.



**Abbildung 2.8:** Ausschnitt aus einem digitalisierten Luftbild mit einer Bodenauflösung von 15 cm (links) und Ergebnis der Merkmalsextraktion (rechts). Extrahierte Flächen sind in blau, Linien in gelb und Punkte in magenta dargestellt.

Die Bildmerkmale sind reichhaltig attributiert. Neben Aussagen über die Klassifikationsunsicherheit wird zu allen Merkmalen zuerst die geometrische Ausprägung beschrieben. Punktmerkmale werden nur durch ihre Koordinaten definiert. Die Linienmerkmale werden durch die Anfangs- und Endpunktkoordinaten bestimmt, und enthalten damit implizit auch die Linienlänge und Linienwinkel bzgl. der Bildachsen. Flächenmerkmale werden zum einen mittels einer Laufängencodierung beschrieben, die alle zu der Fläche gehörenden Bildpunkte aufzählt, und zum anderen durch ihre Konturlinie. Ferner werden die Flächen um weitere Attribute ergänzt, wie z.B. Koordinaten des Schwerpunkts, Flächeninhalt, Winkel der Hauptachse, Umfang, durchschnittlicher Grauwert und Grauwertvarianz.

In Abbildung 2.8 läßt sich erkennen, daß es z.T. erhebliche Abweichungen zwischen der Aufteilung und Begrenzung der dargestellten Objekte im Bild und den von der Merkmalsextraktion gelieferten Bildmerkmalen gibt. Beispielsweise sind viele Linien, die eine Abgrenzung zweier Flächen beschreiben zu kurz, fragmentiert, oder sie fehlen sogar ganz (vgl. [FLF94]). Da Flächen an vielen Stellen in benachbarte Bildbereiche „auslaufen“, eignen sich insbesondere die morphologischen Flächenattribute bei dieser Qualität der Merkmale nicht für die Bildinterpretation. Zur Verbesserung der Ergebnisse der Merkmalsextraktion wird von Fuchs in [FF95, Fuc97] ein Verfahren zur geometrisch und topologisch basierten Gruppierung vorgeschlagen. Es wird gezeigt, daß damit zwar einige Verbesserungen erzielt werden können, jedoch aufgrund eines fehlenden Objektmodells nur sehr eingeschränkte Gruppierungen vorgenommen werden dürfen<sup>4</sup>. Insgesamt bleiben die Ergebnisse der Merkmalsextraktion weiterhin fehlerbehaftet. Da die verschiedenen Fehlerfälle für die Suche nach Gebäuden eine große Rolle spielen, werden sie in Kapitel 4 ausführlich diskutiert.

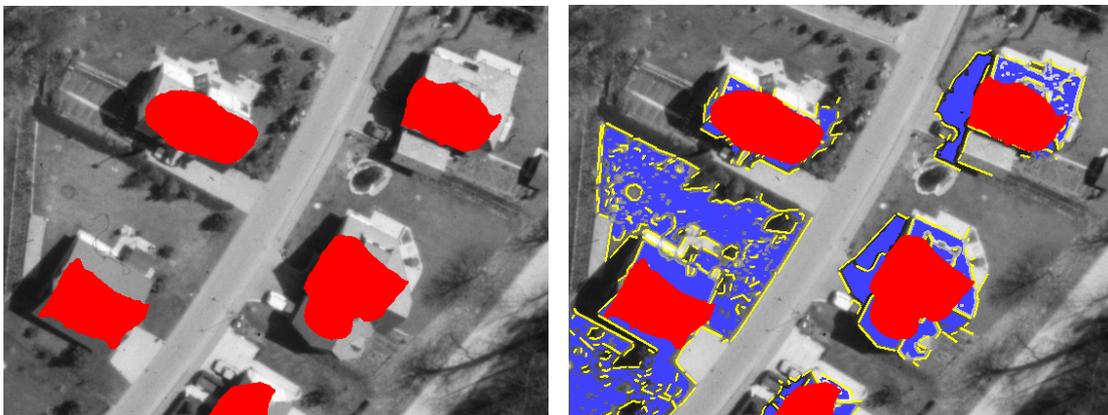
## 2.5.2 Fokussierung auf einzelne Bildbereiche

Die typischen Größen digitalisierter Luftbilder liegen im Bereich von  $10\,000 \times 10\,000$  bis zu  $18\,000 \times 18\,000$  Bildpunkten. Da die Bilder für eine Segmentierung zu groß und zu unhandlich sind, werden vor der Merkmalsextraktion (meistens manuell) Ausschnitte gebildet, die bis zu  $1000 \times 1000$

<sup>4</sup>Auf das Problem der Gruppierung wird später in Kapitel 3.5 noch näher eingegangen.

Bildpunkte groß sind und mehrere Gebäude beinhalten, wobei deren Anzahl von der Bildauflösung und der Bebauungsdichte abhängt. Eine Merkmalsextraktion auf diesen Ausschnitten kann aber immer noch einige tausend Punkt-, Linien- und Flächenmerkmale liefern, so daß der Suchraum für eine automatische Bildinterpretation immer noch deutlich zu groß ist.

Zur Fokussierung auf interessante Bereiche (engl.: *Regions of Interest*, kurz ROI), in denen Gebäude vermutet werden können, wird in Anlehnung an Baltasvias et. al. ([BMS95]) und Henricsson et. al. ([HB97]) ein digitales Oberflächenmodell (DOM) der Szene nach signifikanten Erhebungen durchsucht. Dieses DOM kann automatisch aus einem Stereoluftbildpaar hergeleitet werden. Einen Überblick über Verfahren zur automatischen Rekonstruktion von DOM aus Luftbildern gibt Weidner in [Wei97b]. Aus dem berechneten DOM der Szene werden Erhebungen anhand von Schwellwertoperationen ermittelt (siehe [Wei97b]). Die Bereiche, die über einem Schwellwert liegen und ausreichend groß sind, werden anschließend in die Bilder projiziert (vgl. Abbildung 2.9 links). Nun werden alle Bildmerkmale selektiert, die sich mit den projizierten Bereichen schneiden (vgl. Abbildung 2.9 rechts). Diese Merkmale stellen den Ausgangspunkt für die im folgenden Abschnitt beschriebene 3D-Rekonstruktion von Eckenstrukturen dar.

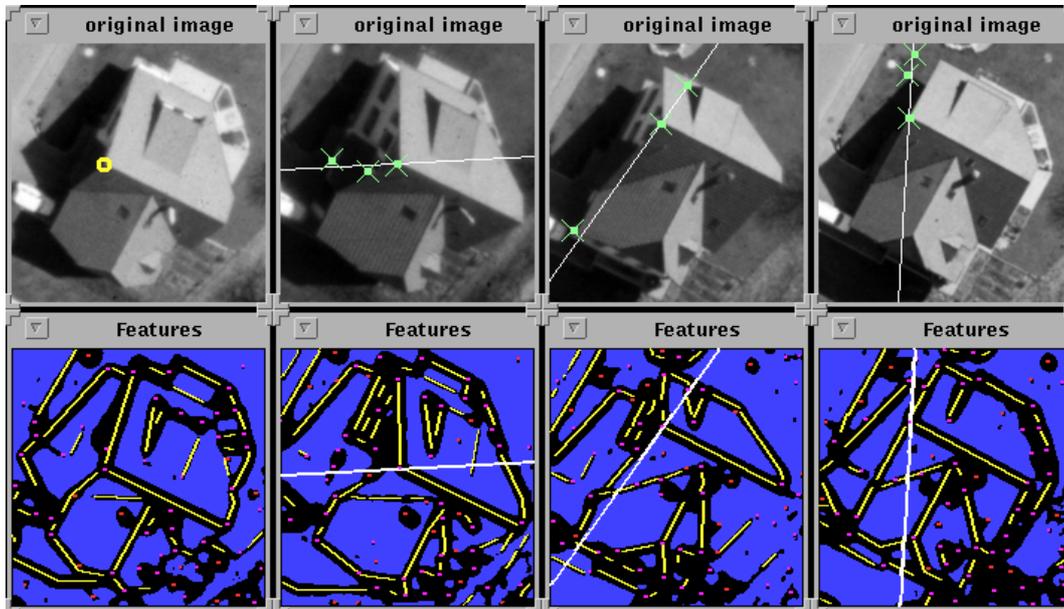


**Abbildung 2.9:** Bestimmung der Bildbereiche, in denen möglicherweise Häuser abgebildet sind, mittels Analyse eines Digitalen Oberflächenmodells (links). Anhand der ermittelten Bereiche, an denen eine lokale Erhebung festgestellt wurde (hier rot dargestellt), wird eine Vorauswahl der für die Suche nach Gebäudestrukturen in Frage kommenden extrahierten Bildmerkmale durchgeführt (rechts).

### 2.5.3 3D-Rekonstruktion von Gebäudeecken

Dreidimensionale Beobachtungen besitzen im Gegensatz zu den 2D-Bilddaten eine deutlich größere Ausdrucksstärke (u.a. wegen der Mehrdeutigkeit zweidimensionaler Projektionen). Aus diesem Grund wird in der Gesamtstrategie ein früher Übergang von den zweidimensionalen Bilddaten in den dreidimensionalen Objektraum angestrebt. Dazu werden mittels eines von Lang und Förstner entwickelten Verfahrens die für Gebäude typischen Eckenstrukturen in den Bildern gesucht und durch eine Mehrbildzuordnung dreidimensional rekonstruiert (siehe [LF96b, FKL97, FKL<sup>+</sup>98, Lan99]). Der Grund, warum Eckpunkte und nicht etwa Linien oder gar Flächen 3D-rekonstruiert werden liegt darin, daß erstens die zur Bildung von Ecken beteiligten Merkmale eine typische Nachbarschaftsstruktur aufweisen, wie bereits in Abbildung 2.3 auf Seite 20 illustriert wurde. Zweitens kann der Eckpunkt aufgrund der Epipolargeometrie in den anderen Bildern nur auf einer Linie, der *Epipolarlinie*, liegen (siehe [Kra94]). Das bedeutet, daß der Aufwand zur Suche nach den entsprechenden Knotenstrukturen in den anderen Bildern stark beschränkt ist, und die Suche damit effizient durchgeführt werden kann (vgl. [HS93a]). Abbildung 2.10 zeigt für die im ersten

Bild betrachtete Ecke die Suche nach korrespondierenden Punkten in den anderen Bildern entlang der Epipolarlinien.



**Abbildung 2.10:** 3D-Rekonstruktion von Eckpunkten durch Mehrbildanalyse (nach [LF96b]). Um die korrespondierenden Punkte des Giebelpunktes des linken Bildes in den anderen Bildern zu finden, wird entlang der Epipolarlinien nach Punkten mit ähnlicher Merkmalsnachbarschaftsstruktur gesucht.

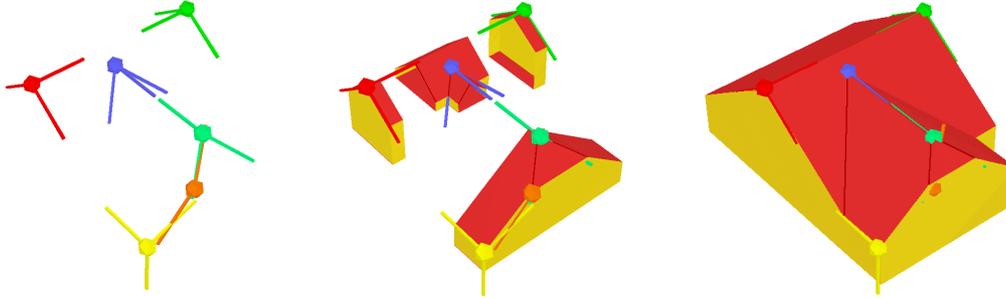
Bei der Rekonstruktion werden die gefundenen Eckenstrukturen gemäß des in Abschnitt 2.4.1 beschriebenen Eckenmodells klassifiziert, und die zu den Eckpunkten inzidenten Linien jeweils mit einer qualitativen Orientierungsklassifikation attribuiert. Zur Suche nach weiteren Eckenstrukturen werden die Richtungsvektoren der zu den gefundenen Ecken inzidenten Linien in die jeweiligen Bilder rückprojiziert und entlanggelaufen, um weitere in Frage kommende Knotenstrukturen zu entdecken. Da hierbei auch Ecken rekonstruiert werden können, die nicht zu demselben Gebäude oder überhaupt zu keinem Gebäude gehören, muß dies bei der im folgenden beschriebenen Generierung von Gebäudehypothesen berücksichtigt werden.

#### 2.5.4 Generierung von Gebäudehypothesen

Die Generierung von Gebäudehypothesen erfolgt auf Basis der rekonstruierten 3D-Eckenstrukturen. Die Aufgabe besteht darin, Gebäude so aus den modellierten Gebäudeteilen zusammensetzen, daß diese das Auftreten der beobachteten 3D-Ecken erklären können. Dabei muß allerdings berücksichtigt werden, daß die Eckenbeobachtungen einerseits unvollständig und andererseits auch fehlerbehaftet sein können.

Fischer und Steinhage haben dazu ein Verfahren entwickelt, das im wesentlichen aus den vier Schritten *Indexing*, *Merging*, *Aggregation* und *Prädiktion* besteht (siehe [FKL97, FKL<sup>+</sup>98]). Dabei werden zunächst im Indexing-Schritt für alle Eckenbeobachtungen die passenden Gebäudeteile ausgewählt. Beim Merging-Schritt werden die Gebäudeteile benachbarter Ecken zusammengefaßt, wenn das neue Gebäudeteil beide Ecken gleichzeitig erklären kann. Dies wird in Abbildung 2.11 an dem mittleren Bild illustriert: die unteren drei Ecken werden durch ein einzelnes Krüppelwalmdachterminal überdeckt. Beim Aggregationsschritt werden die zueinander kompatiblen Gebäudeteilhypothesen auf konsistente Weise miteinander verbunden, wobei die Parameter der Gebäudeteile zum

Teil unifiziert werden. Das Ergebnis der Aggregation wird im rechten Bild von Abbildung 2.11 dargestellt. Die Höhen und Breiten der drei oberen Gebäudeteile (mittleres Bild) wurden unifiziert. Bei der Aggregation konnten ferner die Längen des Haupt- und des Nebentraktes eindeutig bestimmt werden. Die Höhe des Gebäudes bleibt noch unbestimmt, weil keine Beobachtung für eine Bodenecke vorliegt.



**Abbildung 2.11:** Konstruktion einer Gebäudehypothese aus Gebäudeteilen, die die beobachteten 3D-Gebäudeecken erklären kann (aus: [FKL<sup>+</sup>98]). Links: rekonstruierte 3D-Ecken; Mitte: zueinander konsistente Auswahl der zu den Ecken passenden Gebäudeteile; rechts: aggregiertes Gebäudemodell.

Sollten nach dem Aggregationsschritt aufgrund fehlender Beobachtungen noch unvollständige, d.h. offene Gebäudehypothesen bestehen, werden diese beim Prädiktionsschritt durch die Verbindung mit passenden Endstücken geschlossen. Sollten verschiedene Terminale passen, wird für die nach empirischer Ermittlung wahrscheinlichsten Terminale jeweils eine entsprechende Gebäudehypothese generiert.

Bei der Generierung von Gebäudehypothesen können insgesamt aus drei Gründen mehrere alternative Hypothesen entstehen: Erstens können bei der Aggregation verschiedene konsistente Kombinationen möglich sein, zweitens können 3D-Ecken aufgrund unzureichender Beobachtungen den Gebäudeteilen u.U. auf verschiedene Weise zugeordnet werden, und drittens sind bei fehlenden Beobachtungen oftmals verschiedene Prädiktionen möglich. Welche und wieviele Parameter der einzelnen Hypothesen zu diesem Zeitpunkt schon bestimmt sind, hängt von dem Haustyp und den im jeweiligen Fall einbezogenen rekonstruierten 3D-Ecken ab. Das Spektrum reicht von weitestgehend unbestimmt<sup>5</sup> (wenn z.B. nur eine Eckenbeobachtung zur Verfügung stand) bis vollständig bestimmt.

### 2.5.5 Berechnung der Gebäudeansichten

Da die Bilddaten zweidimensional und die Gebäudehypothesen dreidimensional sind, werden zur Verifikation die verschiedenen möglichen Ansichten der Gebäudehypothesen in den Bildern berechnet. In Abschnitt 2.4.2 wurde erläutert, daß dazu Aspekthierarchien verwendet werden, die die topologisch verschiedenen Projektionen aufzählen. Während klassische Verfahren zur Berechnung von Aspektgraphen bzw. -hierarchien (siehe [GCS91]) davon ausgehen, daß die Formparameter der 3D-Objekte feststehen und nur die Betrachtungsrichtung frei ist, steht in diesem Fall jedoch die Betrachtungsrichtung aufgrund der bekannten Aufnahmeorientierung und der durch die beobachteten Ecken bekannte Lage der Gebäudehypothese weitestgehend fest. Allerdings können einige Formparameter wie z.B. die Gebäudehöhe oder Länge unbestimmt sein. Dies kann bei komplexeren Modellen dazu führen, daß von einem Gebäude für dieselbe Betrachterposition unterschiedliche

<sup>5</sup>Wenn für einen Bildausschnitt keine Ecken rekonstruiert werden konnten, und trotzdem bekannt ist, daß sich an der betreffenden Stelle ein Haus befindet, kann das Verfahren auch Hypothesen für die wahrscheinlichsten Gebäudetypen präzisieren. In diesem Fall sind alle Parameter unbestimmt.

Aspekte entstehen können. Beispielsweise könnte bei einer Schrägansicht eines T-förmigen Hauses je nach Länge des Seitentraktes der hintere Teil des Gebäudes verdeckt sein oder nicht.

Steinhage und Fischer haben deshalb eine entsprechende Erweiterung der Aspekthierarchien vorgeschlagen, die sie als *Ansichtenhierarchie* bezeichnen (siehe [FKL97, Ste97]). Ansichtenhierarchien berücksichtigen bei der Generierung der möglichen Ansichten sowohl die freien als auch die feststehenden Gebäudeform- und Orientierungsparameter. Da Ansichtenhierarchien speziell auf die jeweilig betrachtete Szene zugeschnitten sind, kann ein Beleuchtungsmodell einfach integriert werden, das bei bekanntem Sonnenstand der Szene u.a. die Prädikation des Schattenwurfs erlaubt (vgl. [Ste97]). Die Berechnung der Gebäudeansichten erfolgt mittels eines Sampling-Ansatzes [Ste98], bei dem für die freien Form- und Orientierungsparameter zufällige Werte (innerhalb der jeweiligen gültigen Bereiche) gewählt werden und die entsprechende Ansicht berechnet wird. Dies wird für einige hundert bis tausend Parametersätze wiederholt, wobei schließlich nur die topologisch verschiedenen Ansichten als Ergebnis ausgegeben werden. Der Begriff *Gebäudeansicht* wird im folgenden zur Bezeichnung einer speziell für die aktuelle Szene berechneten möglichen zweidimensionalen Ausprägung einer in Frage kommenden Gebäudehypothese verwendet.

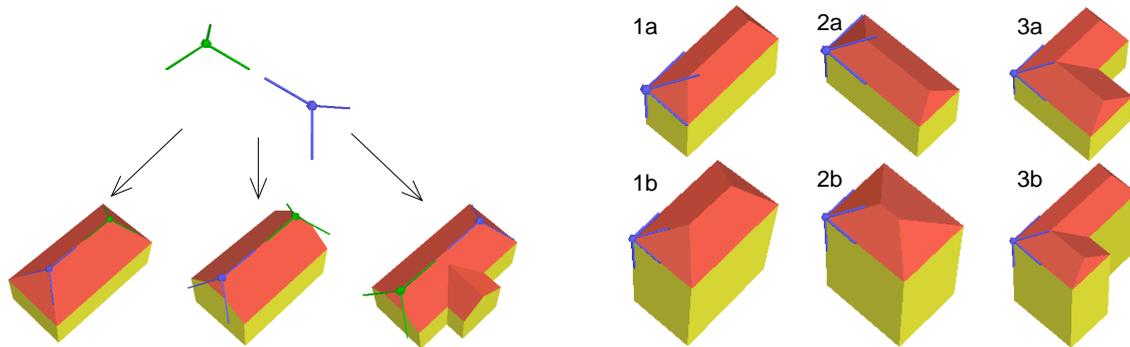
### 2.5.6 Identifikation der richtigen Gebäudeansicht

Nach der Generierung einer Menge von Gebäudehypothesen und ihrer möglichen Ansichten in den Bildern, geht es bei dem Schritt der Identifikation darum, festzustellen, ob und wo sich eine der Gebäudeansichten im Bild befindet. Gleichzeitig ist das Problem zu lösen, daß wenn mehrere alternative Gebäudehypothesen aufgrund von zufälligen Ähnlichkeiten mit unbeteiligten Bildstrukturen oder wegen starker Ähnlichkeiten der Hypothesen erkannt werden konnten, zu entscheiden, um welche der gefundenen Hypothesen es sich am wahrscheinlichsten handelt.

Im Gegensatz zu den in Abschnitt 2.3.3 vorgestellten Verfahren zur automatischen Gebäudeextraktion, die ebenfalls eine Verifikation generierter Gebäudehypothesen durchführen, treten hier zusätzlich die folgenden Probleme auf:

1. Die einzelnen Gebäudeansichten besitzen noch Freiheitsgrade bzgl. der Orientierung und Ausdehnung im Bild. Das bedeutet, daß die Gebäudeansichten erst noch durch eine Zuordnung der Modellmerkmale zu den entsprechenden extrahierten Bildmerkmalen eindeutig bestimmt werden müssen. Hierbei ist jedoch zu beachten, daß bei den Gebäudeansichten zwar die Topologie der Merkmale feststeht, nicht jedoch deren Geometrie. Das bedeutet, daß z.B. die Position, Längen und Winkel von Linien noch stark variieren können, und deshalb keine Zuordnung von Modell und Bild ausschließlich auf Basis der Merkmalseigenschaften durchgeführt werden kann.
2. Es können unterschiedlich komplexe Gebäudehypothesen generiert werden. Die Bewertungsfunktion zur Auswahl der zutreffendsten Hypothese muß also in der Lage sein, unterschiedlich komplexe Modelle miteinander vergleichen zu können.

Abbildung 2.12 illustriert diese Probleme anhand von zwei Beispielen. Im ersten Beispiel (linkes Bild) wurden zwei Giebeleckpunkte rekonstruiert, die sowohl zu einem Walm- als auch zu einem Krüppelwalmdachhaus gehören könnten. Dementsprechend wurden auch für beide Haustypen Gebäudehypothesen generiert. Würde das Verfahren zur Generierung von Gebäudehypothesen eine stärkere Prädiktionskomponente besitzen, könnte auch eine Hypothese wie die rechts dargestellte generiert werden. Das im rechten Bild von Abbildung 2.12 dargestellte Beispiel zeigt die möglichen Gebäudehypothesen und deren Freiheitsgrade, wenn nur eine Eckenbeobachtung vorliegt. Es werden drei Hypothesen generiert, wobei in jeder Spalte eine Hypothese in zwei verschiedenen Ausdehnungen gezeigt wird. Während ein Teil der Linienwinkel feststeht, sind die Linienlängen und damit die Flächengrößen noch unbestimmt. Wegen dieser Freiheitsgrade sind auch die Positionen der Linien im Bild noch nicht festgelegt. An den Bildern 3a und 3b läßt sich ferner erkennen, daß aufgrund der freien Parameter topologisch verschiedene Ansichten möglich sind.



**Abbildung 2.12:** Mögliche Hypothesen und ihre Freiheitsgrade in Abhängigkeit rekonstruierter 3D-Ecken. Links: zwei Dachecken, die sowohl von einem Walmdach- als auch einem Krüppelwalmdachhaus stammen können. Rechts: zu der einen rekonstruierten Dachecke sind viele verschiedene Gebäudehypothesen (1a, 2a, 3a) in verschiedenen Größen und Ausdehnungen (1b, 2b, 3b) möglich.

Die Verifikation der Gebäudehypothesen erfordert also die Suche nach einer passenden Zuordnung der Modellmerkmale der Gebäudehypothesen zu den extrahierten Bildmerkmalen, um zuerst die freien Parameter zu bestimmen, bevor anschließend der Grad der Übereinstimmung zwischen Modell und Bild gemessen werden kann. Da die meisten Eigenschaften der Modellmerkmale, wie z.B. die Linienlänge, Flächengröße und die absoluten Bildkoordinaten von Eckpunkten, aufgrund der Freiheitsgrade z.T. noch nicht feststehen, reicht ein einfaches merkmalsbasiertes Zuordnungsverfahren nicht aus. Geometrische Relationen wie z.B. die Kollinearität und Linienparallelität sind hingegen (in bestimmten Grenzen) genau wie die topologischen Beziehungen der Merkmale invariant gegenüber diesen Freiheitsgraden. Darüberhinaus sind Relationen oftmals robust gegenüber Störungen in den Bildern. Beispielsweise werden bei der Segmentierung Linien oftmals zu kurz extrahiert, oder sie sind in kleinere Liniensegmente fragmentiert (vgl. Abbildung 2.8 auf Seite 24). In beiden Fällen bleibt jedoch die Parallelität derartiger betroffener Linien erhalten.

Die Suche der entsprechenden Zuordnung von Modell und Bild erfolgt deshalb insbesondere anhand der Relationen zwischen den Merkmalen. Die Aufgabe besteht darin, den einzelnen Merkmalen der Gebäudeansichten jene extrahierte Bildmerkmale zuzuordnen, die die vom Modell erwarteten Relationen erfüllen. Dieses Problem wird auch als Problem der relationalen Zuordnung bezeichnet (engl.: *Relational Matching*), wobei die Zuordnung durch die in Luftbildern auftretenden Störungen z.B. durch Rauschen und Verdeckungen deutlich erschwert wird (vgl. [SH87, Vos92, SFH92, Rad93]). Im Kontext dieser Problemstellung ergeben sich zwei Aufgabenbereiche: Erstens müssen geeignete Relationen identifiziert werden, die Gebäudeansichten ausreichend spezifisch beschreiben, so daß sie sowohl vom Bildhintergrund als auch von Gebäuden anderen Typs unterschieden werden können. Zweitens besitzt das Zuordnungsproblem eine große kombinatorische Komplexität, weshalb ein effizientes Suchverfahren benötigt wird. Diese beiden Bereiche werden in dieser Arbeit in den Kapiteln 3 und 4 behandelt, wobei in Kapitel 3 die Bildstörungen zunächst teilweise vernachlässigt werden, ihre explizite Berücksichtigung dafür aber das Thema von Kapitel 4 ist. In Kapitel 5 wird dann gezeigt, wie das Zuordnungsverfahren mit einer informationstheoretisch basierten Bewertungsfunktion gekoppelt werden kann, so daß die beste Zuordnung eine probabilistische Semantik erhält, nämlich die der Maximum-Likelihood-Schätzung zwischen Modell und Bild. Ferner erlaubt die Bewertungsfunktion die Verwendung eines Hypothesentests, anhand dessen schlechte Zuordnungen und damit die unwahrscheinlichen Hypothesen eliminiert werden können.

Zur Bestimmung der wahrscheinlichsten Hypothese wird für jede Gebäudeansicht zunächst separat die beste Zuordnung mit den Bilddaten bestimmt. Dieser Ansatz ist praktikabel, weil die Anzahl möglicher Gebäudeansichten aufgrund der rekonstruierten 3D-Ecken auf wenige beschränkt ist<sup>6</sup>.

<sup>6</sup>Ein häufig genanntes Gegenargument zur Verwendbarkeit von Aspektgraphen zur Objekterkennung ist, daß sie oftmals unpraktikabel sind, weil es für komplexe Objekten zu viele Aspekte geben kann (siehe [OAD<sup>+</sup>92]). Bei den

Zur Bestimmung der wahrscheinlichsten Hypothese muß einerseits die Güte der jeweiligen Zuordnung zu den Bilddaten und andererseits die Komplexität der Modelle berücksichtigt werden. Die Entwicklung eines Verfahrens, das dazu in der Lage ist, wird schließlich in Kapitel 6 beschrieben.

### 2.5.7 Schätzung der Gebäudeparameter

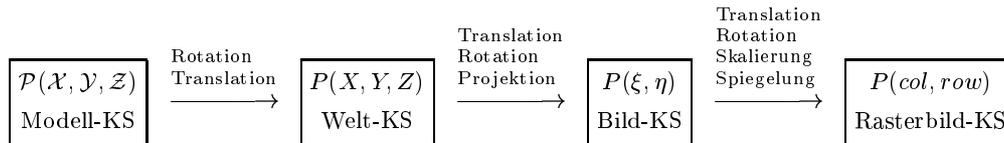
Nach der Identifikation der richtigen Gebäudeansicht können unmittelbar die Gebäudeparameter geschätzt und unter bestimmten Voraussetzungen, die weiter unten erläutert werden, die Gebäudehypothese vollständig dreidimensional rekonstruiert werden.

Ein erfolgreiches Matching ordnet den Modellmerkmalen extrahierte Bildmerkmale zu. Damit wird die Geometrie der Merkmale der entsprechenden Gebäudeansicht fixiert. Da sich die Geometrie der 2D-Ansichtsmerkmale funktional aus der Geometrie der entsprechenden 3D-Gebäudehypothese herleitet, und diese wiederum von den Gebäudeparametern abhängt, können die Gebäudeparameter durch die Anwendung eines Standard-Parameterschätzverfahrens aus den Beobachtungen näherungsweise bestimmt werden. Dabei machen allerdings einige Besonderheiten, die mit der in Abschnitt 2.4 vorgestellten Modellierung zusammenhängen, die praktische Umsetzung recht aufwendig. Dies soll in diesem Abschnitt näher erläutert werden.

Die zu bestimmenden Gebäudeparameter sind die Lokation, der Drehwinkel  $\alpha$  um die Vertikale und die vom jeweiligen Haustyp abhängigen Formparameter  $fp$ . Zur Bestimmung der Lokation müssen die Parameter  $t_x, t_y, t_z$  der Translation vom Modell- in das Weltkoordinatensystem berechnet werden.

#### 2.5.7.1 Abbildungsgleichungen

Als Voraussetzung zur Parameterschätzung müssen die funktionalen Abhängigkeiten der Beobachtungen von den Parametern explizit modelliert sein. Die Parametrisierung des Gebäudemodells beschreibt die Koordinaten der Gebäudeeckpunkte in Abhängigkeit der Gebäudeparameter (vgl. Abbildung 2.5 auf Seite 21). Bei der Abbildung eines 3D-Gebäudeeckpunktes auf einen 2D-Punkt im Rasterbild werden insgesamt die folgenden Transformationen durchgeführt (vgl. dazu [Kra94, HS93a, FvDFH90]):



Als erstes werden die Modellkoordinaten in ein übergeordnetes geodätisches „Weltkoordinatensystem“ (Welt-KS) überführt. Das Modell, welches sich im Ursprung des Modellkoordinatensystems befindet, wird zunächst so um die vertikale Achse gedreht, daß die Koordinatenachsen deckungsgleich mit denen des Weltkoordinatensystems werden. Dazu werden die Modellkoordinaten mit der Rotationsmatrix  $R_\alpha$  multipliziert. Danach wird das Modell mit Hilfe der Translationsmatrix  $T_{t_x, t_y, t_z}$  an seine Position im Weltkoordinatensystem verschoben:

$$\underbrace{\begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}}_{P^{\text{Welt}}} = \underbrace{\begin{bmatrix} 1 & 0 & 0 & t_x \\ 0 & 1 & 0 & t_y \\ 0 & 0 & 1 & t_z \\ 0 & 0 & 0 & 1 \end{bmatrix}}_{T_{t_x, t_y, t_z}} \underbrace{\begin{bmatrix} \cos \alpha & -\sin \alpha & 0 & 0 \\ \sin \alpha & \cos \alpha & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}}_{R_\alpha} \underbrace{\begin{bmatrix} \mathcal{X}(fp) \\ \mathcal{Y}(fp) \\ \mathcal{Z}(fp) \\ 1 \end{bmatrix}}_{P_{fp}^{\text{Modell}}} \quad (2.1)$$

Ansichtenhierarchien wird die Anzahl der Aspekte durch die Eckenbeobachtungen jedoch auf wenige reduziert.

Die Transformation und Projektion von Welt- in Bildkoordinaten wird ausschließlich durch die Parameter der Luftbildaufnahme bestimmt. Die Parameter der *äußeren Orientierung* setzen sich zusammen aus dem Aufnahmeort (Koordinaten  $X_0, Y_0, Z_0$  des Projektionszentrums bzgl. des Weltkoordinatensystems) und dem Aufnahmewinkel (Drehung des Kamerakoordinatensystems um die drei Achsen bzgl. des Weltkoordinatensystems um die Winkel  $\omega, \phi, \kappa$ ). Aus diesen Drehwinkeln ergibt sich die Rotationsmatrix  $R_{\omega\phi\kappa}$  zu

$$R_{\omega\phi\kappa} = \begin{bmatrix} \cos\phi\cos\kappa & -\cos\phi\sin\kappa & \sin\phi \\ \cos\omega\sin\kappa + \sin\omega\sin\phi\cos\kappa & \cos\omega\cos\kappa - \sin\omega\sin\phi\sin\kappa & -\sin\omega\cos\phi \\ \sin\omega\sin\kappa - \cos\omega\sin\phi\cos\kappa & \sin\omega\cos\kappa + \cos\omega\sin\phi\sin\kappa & \cos\omega\cos\phi \end{bmatrix} \quad (2.2)$$

Die Parameter der *inneren Orientierung* sind durch die Kamerakonstante  $c$  (Abstand des Projektionszentrums zur Projektionsebene) und die Verschiebung  $\xi_0, \eta_0$  des Bildhauptpunktes zum Bildmittelpunkt gegeben<sup>7</sup>.

Für die Projektion der dreidimensionalen Weltkoordinaten in zweidimensionale Bildkoordinaten wird zunächst mit Hilfe der Translationsmatrix  $T_{X_0, Y_0, Z_0}$  der Aufnahmeort in den Ursprung verschoben. Danach wird das Koordinatensystem durch eine Drehung an das der Kamera angepaßt. Dazu werden die Koordinaten mit der invertierten Rotationsmatrix  $R_{\omega\phi\kappa}^{-1} = R_{\omega\phi\kappa}^T$  multipliziert. Anschließend wird das Projektionszentrum durch Multiplikation mit der Translationsmatrix  $T_c$  auf die Bildebene verschoben und die Koordinaten dann durch die Multiplikation mit der Zentralprojektionsmatrix  $\Pi$  in den projektiven Raum abgebildet:

$$\begin{bmatrix} u^* \\ v^* \\ w^* \end{bmatrix} = \underbrace{\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \frac{1}{c} & 1 \end{bmatrix}}_{\Pi} \underbrace{\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & -c \\ 0 & 0 & 0 & 1 \end{bmatrix}}_{T_c} \underbrace{\begin{bmatrix} R_{\omega\phi\kappa} & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}^T}_{R_{\omega\phi\kappa}^{-1}} \underbrace{\begin{bmatrix} 1 & 0 & 0 & -X_0 \\ 0 & 1 & 0 & -Y_0 \\ 0 & 0 & 1 & -Z_0 \\ 0 & 0 & 0 & 1 \end{bmatrix}}_{T_{X_0, Y_0, Z_0}} \underbrace{\begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}}_{P^{\text{Welt}}} \quad (2.3)$$

Die Bildkoordinaten  $\xi, \eta$  ergeben sich aus der Division der beiden Koordinaten  $u^*$  und  $v^*$  durch  $w^*$  und der Ausgleichung der Verschiebung des Bildhauptpunktes zum Bildmittelpunkt:

$$\xi = \xi_0 - \frac{u^*}{w^*} \quad \eta = \eta_0 - \frac{v^*}{w^*} \quad (2.4)$$

Zuletzt werden die Bildkoordinaten in Koordinaten des gescannten Rasterbildes überführt. Dazu wird der Punktvektor des Bildes  $P^{\text{Bild}}$  mit einer weiteren Transformationsmatrix,  $A$ , multipliziert:

$$\underbrace{\begin{bmatrix} col \\ row \end{bmatrix}}_{P^{\text{Pixel}}} = \underbrace{\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix}}_A \underbrace{\begin{bmatrix} 1 \\ \xi \\ \eta \end{bmatrix}}_{P^{\text{Bild}}} \quad (2.5)$$

Matrix  $A$  reflektiert die Parameter des Scan-Vorganges und wird dem digitalisierten Rasterbild beigelegt. Sie berücksichtigt die Pixelauflösung durch eine Skalierung, die Verschiebung des Rasterbildursprungs zum Bildursprung, die Spiegelung an der Horizontalachse<sup>8</sup> sowie eine evtl. notwendige Rotation, falls die Bildachsen beim Scannen nicht deckungsgleich mit den Rasterbildachsen waren.

<sup>7</sup>Die Parameter der inneren Orientierung hängen vom jeweiligen Kamertyp ab und werden vom Hersteller bestimmt. Der Bildhauptpunkt sollte normalerweise genau im Bildmittelpunkt liegen, was aus fertigungstechnischen Gründen jedoch selten erreicht wird.  $\xi_0, \eta_0$  bezeichnen diese (sehr kleinen) Abweichungen.

<sup>8</sup>Der Ursprung liegt bei digitalisierten Bildern meistens in der linken oberen Bildecke, wobei die vertikalen Koordinatenwerte größer werden, desto weiter unten ein Punkt im Bild liegt.

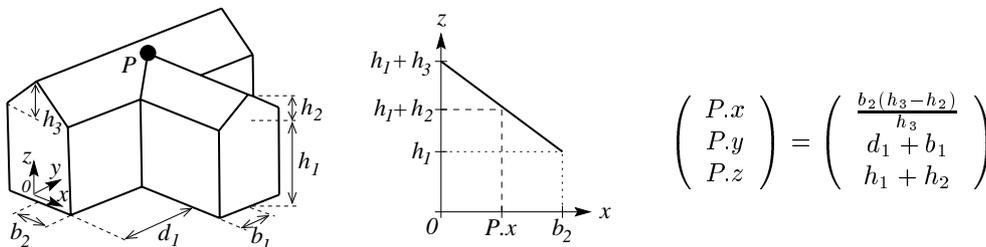
### 2.5.7.2 Rekonstruktion der Gebäudeparameter

Die Gleichungen 2.1 bis 2.5 beschreiben die Abbildung eines Eckpunktes der Gebäudehypothese in ein digitalisiertes Rasterbild in Abhängigkeit von den Gebäude- und den Aufnahmeparametern. Da die Aufnahmeparameter für ein gegebenes Luftbild bekannt sind, müssen zur Bestimmung der Gebäudeparameter aus zugeordneten Bildmerkmalen die Formparameter  $fp$ , die Lokation  $t_x, t_y, t_z$  und der Drehwinkel  $\alpha$  geschätzt werden. Dabei treten drei wesentliche Schwierigkeiten in Erscheinung:

1. In den beiden Gleichungen 2.4 müssen aus zwei Bildkoordinaten  $\xi, \eta$  drei Parameter  $u^*, v^*, w^*$  bestimmt werden, d. h. diese Gleichungen sind unterbestimmt. Die Parameter können nur bis auf einen Skalierungsfaktor  $s$  berechnet werden, wobei es aber mehrere Möglichkeiten gibt, den Skalierungsfaktor zu bestimmen:
  - (a) Durch die gleichzeitige Einbeziehung der Gleichungen einer zweiten Luftbildaufnahme des selben Objekts von einem anderen Aufnahmeort kann  $s$  für beide Bilder wegen der Epipolargeometrie eindeutig bestimmt werden (siehe [Kra94] und [HS93a]).
  - (b) Liegt nur ein Bild vor, ist es möglich den Skalierungsfaktor aus dem Schattenwurf eines Gebäudes zu ermitteln. Sofern das Aufnahmedatum und die Aufnahmezeit bekannt sind, kann die Position der Sonne bzgl. der abgebildeten Szene rekonstruiert werden. Da die Länge der Schattenkante einer vertikalen Wand proportional zu der Höhe der Wand ist, und sich dieser Faktor mit Hilfe des gegebenen Sonnenvektors berechnen läßt, kann direkt die Höhe der vertikalen Wand und damit  $s$  bestimmt werden. Dieses Prinzip wird u. a. von Huertas und Nevatia sowie Shufelt zur 3D-Rekonstruktion von Gebäuden aus Einzelbildern eingesetzt (siehe [HN88] und [Shu96]).
  - (c) Sofern zusätzlich zu dem Bild ein digitales Geländemodell vorliegt, kann aus der Differenz der Flughöhe zur Bodenhöhe an der Position des Gebäudes der Skalierungsfaktor  $s$  ungefähr geschätzt werden (vgl. [Kra94]).

Zur vollständigen, parameterfreien 3D-Rekonstruktion des Gebäudes muß also entweder ein erfolgreiches Matching in mindestens zwei verschiedenen Bildern vorliegen, ein DGM der beobachteten Szene vorhanden sein, oder es müssen der Sonnenstand bekannt und geeignete Schattenkanten beobachtbar sein.

2. Die Gleichungen 2.1 und 2.4 sind an den hellgrau unterlegten Stellen nichtlinear. Das bedeutet, daß zur Bestimmung der gesuchten Gebäudeparameter auf ein Verfahren zur nichtlinearen Parameterschätzung zurückgegriffen werden muß. Abbildung 2.13 zeigt ein Beispiel für eine Gebäudeparametrisierung, bei der eine Eckpunktcoordinate nichtlinear von den zu schätzenden Gebäudeparametern abhängt.



**Abbildung 2.13:** Beispiel für eine nichtlineare Parametrisierung einer Gebäudeeckpunktcoordinate. Die  $x$ -Koordinate des Punktes  $P$  berechnet sich nach dem zweiten Strahlensatz aus dem Gleichsetzen der Verhältnisse von  $b_2/h_3$  und  $P.x/(h_3 - h_2)$ .

3. Durch das Matching mit unterschiedlichen Merkmalstypen wird die Geometrie der Aspektmerkmale an vielen Stellen redundant bestimmt, z.B. können die 2D-Koordinaten einer Gebäudeecke unmittelbar einem zugeordneten extrahierten Punktmerkmal entnommen, aber auch aus dem Geradenschnitt zweier in diesem Punkt zusammenlaufender extrahierter Linienmerkmale berechnet werden. Das bedeutet, daß das Gleichungssystem für die Parameterschätzung überbestimmt ist, und zur Lösung ein Ausgleichungsverfahren verwendet werden muß.

Zur Lösung des nichtlinearen Ausgleichungsproblems wird die Methode der kleinsten Fehlerquadrate verwendet (siehe [Kra94, LH95, Koc97]). Es soll an dieser Stelle daher nur gezeigt werden, welche Gestalt die in die Ausgleichung eingehenden Gleichungen annehmen. Die Abbildung für jeden Punkt  $P_i$  der Gebäudeansicht stellt sich als Funktion  $f_i$  in Abhängigkeit von den gesuchten Parametern  $fp, \alpha, t_x, t_y, t_z$  wie folgt dar:

$$P_i^{\text{Pixel}} = f_i(fp, \alpha, t_x, t_y, t_z) \quad (2.6)$$

Dem Matching lassen sich nun mehrere Beobachtungen zu jeder Gebäudeecke entnehmen. Zum einen sind das die Koordinaten eines etwaig zugeordneten Punktmerkmals, und zum anderen alle Schnittpunktkoordinaten der paarweise geschnittenen, in dieser Gebäudeecke zusammentreffenden Linien. Wenn  $L_i$  die Anzahl der von der Gebäudeecke  $i$  ausgehenden Linien in der Gebäudeansicht ist, ist die Anzahl der Beobachtungen  $\hat{n}_i$  zu dieser Gebäudeecke somit beschränkt durch

$$\hat{n}_i = 1 + \frac{(L_i - 1)L_i}{2} \quad (2.7)$$

Da aber — wie später in Kapitel 4 näher erläutert wird — u. U. bestimmte Merkmale im Bild nicht beobachtbar sind, kann die tatsächliche Anzahl auch kleiner sein ( $n_i \leq \hat{n}_i$ ); im Extremfall liegt für eine Ecke keine Beobachtung vor ( $n_i = 0$ ). Für alle Gebäudeecken und Beobachtungen bezeichnet nun  $col_{i,j}$  die Spaltenkoordinate der  $j$ -ten Beobachtung für die  $i$ -te Gebäudeecke im Rasterbild. Analog bezeichnet  $row_{i,j}$  die Zeilenkoordinate der  $j$ -ten Beobachtung für die  $i$ -te Gebäudeecke. Damit ergeben sich für jede Gebäudeecke  $i$  mit ihren  $n_i$  Beobachtungen die folgenden Gleichungen:

$$\begin{aligned} d\beta_{col_{i,1}} &= col_{i,1} - f_i.col(fp, \alpha, t_x, t_y, t_z) \\ d\beta_{row_{i,1}} &= row_{i,1} - f_i.row(fp, \alpha, t_x, t_y, t_z) \\ &\vdots \\ d\beta_{col_{i,n_i}} &= col_{i,n_i} - f_i.col(fp, \alpha, t_x, t_y, t_z) \\ d\beta_{row_{i,n_i}} &= row_{i,n_i} - f_i.row(fp, \alpha, t_x, t_y, t_z) \end{aligned} \quad (2.8)$$

Sollten mehrere erfolgreiche Zuordnungen desselben Gebäudes für verschiedene Bilder vorliegen, werden die obigen Gleichungen für jedes Bild generiert und nur eine Ausgleichung über sämtliche Gleichungen durchgeführt. Dies erlaubt zum einen die maßstabsgerechte 3D-Rekonstruktion der Parameter (siehe oben, Anmerkung 1.) und zum anderen eine Erhöhung der Parametergenauigkeit, da alle Beobachtungen gleichzeitig in die Parameterbestimmung einfließen.

### 2.5.7.3 Realisierung und Durchführung des Ausgleichsverfahrens

Die Gebäudegeometrien sind aufgrund der Generik der Modellierung nicht von vornherein festgelegt. Darüberhinaus ist auf jeder Ansicht einer Gebäudehypothese nur ein Teil des Gebäudes sichtbar. Das bedeutet, daß die Menge der dargestellten Punktgleichungen sowie deren Parametrisierungen durch die Formparameter variabel sind, und von der konkreten Gebäudehypothese und Ansicht abhängen. Das bedeutet aber auch, daß die oben gezeigten Gleichungen nicht statisch sind und nicht etwa für alle in Frage kommenden Gebäudemodelle einmal durchgerechnet werden könnten. Vielmehr entstehen die Gleichungen dynamisch bei der Generierung von Gebäudehypothesen und deren Ansichten; sie müssen also symbolisch verarbeitet werden.

Für die Realisierung wird das iterative Standardverfahren von Levenberg und Marquardt eingesetzt (siehe [PTVF92]), welches sowohl erlaubt, die Genauigkeit der einzelnen Beobachtungen in Form individueller Varianzen einzubeziehen, als auch zu den geschätzten Parametern die Genauigkeit in Form einer Kovarianzmatrix liefert. Darüberhinaus werden bei der Schätzung auch die an die Formparameter geknüpften Bedingungen einbezogen (vgl. Abschnitt 2.4).

Für die Ausgleichung müssen Anfangswerte für die zu schätzenden Parameter vorgegeben werden. Zur Vermeidung des unerwünschten Verhaltens, daß das verwendete Ausgleichsverfahren divergiert oder zu einem lokalen Minimum konvergiert, sollten die Anfangswerte möglichst nahe an den zu erwartenden Werten liegen. Da die Gebäudegeometrie gleichzeitig von den Gebäudeparametern und der Lokation abhängt (insbesondere die Gebäudegröße von der geographischen Höhe), wird der Ausgleichungsprozeß in zwei Schritten durchgeführt:

1. Die Formparameter werden auf Durchschnittswerte, die empirisch ermittelt werden können, gesetzt und festgehalten. Es werden dann zunächst nur die Orientierungsparameter bestimmt.
2. Mit den ermittelten Näherungen für die Orientierungsparameter und den Durchschnittswerten für die Formparameter als Startwerte wird eine zweite Ausgleichung durchgeführt, bei der nun die Werte aller Parameter geschätzt werden.

Sollte trotz einer erfolgreichen Zuordnung die Genauigkeit der rekonstruierten Parameter z.B. aufgrund vereinzelter Falschzuordnungen<sup>9</sup> beeinträchtigt sein, kann die Ausgleichung noch verbessert werden. In [LG98] wird ein Verfahren zur robusten Schätzung von Gebäudeparametern vorgestellt, welches Ausreißer in den dem Gebäudemodell zugeordneten Bildmerkmalen während der Ausgleichung identifiziert und eliminiert. Das Verfahren wird von den Autoren zur robusten Schätzung von Gebäudeparametern im Kontext einer halbautomatischen Gebäudeextraktion verwendet, ließe sich jedoch auf die automatische Extraktion übertragen.

---

<sup>9</sup>Bei der Zuordnung von Bild- zu Modellmerkmalen kann es u.U. vorkommen, daß z.B. Störmerkmale im Bild die geforderten Modellbedingungen noch knapp erfüllen und ebenfalls dem Modell zugeordnet werden. Diese Merkmale können bei der 3D-Rekonstruktion die Parameterwerte leicht verfälschen.

## Kapitel 3

# Effiziente Objektidentifikation mit Constraint-Techniken

Die verwendete Strategie zur Gebäudeerkennung basiert auf der Generierung und Verifikation von Hypothesen. Die Gebäudehypothesen werden auf der Basis rekonstruierter 3D-Eckenstrukturen erzeugt. Es werden dreidimensionale Gebäudemodelle konstruiert, die die Beobachtung dieser Eckenstrukturen erklären können. Da im allgemeinen nur ein Teil der Ecken eines Gebäudes rekonstruiert werden, können die Gebäudemodelle noch Freiheitsgrade bzgl. der Form, Ausdehnung, Lage und Orientierung besitzen. Zur Verifikation der Gebäudehypothesen werden die möglichen zweidimensionalen Ansichten berechnet. Wegen der freien Parameter kann sowohl die Lage und die Orientierung als auch die Ausdehnung der Gebäudeansichten im Bild unbestimmt sein (vgl. Abschnitt 2.5.6). Zur Identifikation der richtigen Hypothese müssen die Gebäudeansichten deshalb im Bild gesucht werden.

Die Bilddaten sind in Form extrahierter Punkt-, Linien- und Flächenmerkmale gegeben. Zur Suche der Gebäudeansichten werden diese ebenfalls in die einzelnen Merkmale zerlegt, und es wird die passende Zuordnung zwischen den Modellmerkmalen und den zu dem abgebildeten Gebäude gehörenden extrahierten Bildmerkmalen gesucht. Aufgrund der freien Parameter stehen viele Merkmalseigenschaften wie z.B. die Linienlängen oder Flächengrößen nicht fest. Zudem sind die absoluten Merkmalseigenschaften verhältnismäßig störanfällig gegenüber Fehlern bei der Merkmalsextraktion, wie z.B. bei der Fragmentierung oder dem Verschmelzen von Merkmalen. Da für die Suche eine gegenüber der verschiedenen Freiheitsgrade der Gebäudeansichten invariante Repräsentation benötigt wird, kann die Zuordnung nicht auf Basis der absoluten Merkmalseigenschaften geschehen.

Relationen zwischen den Merkmalen, wie beispielsweise die Linienparallelität oder die Merkmalsnachbarschaft, sind jedoch invariant gegenüber den Freiheitsgraden. Außerdem sind sie relativ robust gegenüber Bildstörungen und Fehlern der Merkmalsextraktion. Die Zuordnung erfolgt deshalb anhand der Relationen: Es wird eine Zuordnung von Bild- und Modellmerkmalen gesucht, wobei zwischen den zugeordneten Bildmerkmalen die gleichen Relationen erfüllt sein müssen, wie zwischen den Merkmalen des Modells. Da (binäre) Relationen durch Graphen repräsentiert werden können, entspricht die Suche nach der Gebäudeansicht der Bestimmung eines Subgraph-Isomorphismus des (kleinen) Modellgraphen in dem (großen) Bildgraphen.

Bei der Bestimmung eines Subgraph-Isomorphismus handelt es sich um ein kombinatorisches Suchproblem hoher Komplexität. Aus diesem Grund wird ein möglichst effizientes Verfahren zur Bestimmung der Zuordnung von Modell- und Bildmerkmalen benötigt. Constraint-Techniken nutzen die netzwerkartige Struktur der relationalen Modell- und Bildrepräsentation zur Propagierung von Einschränkungen, und ermöglichen damit die frühzeitige Erkennung von „Sackgassen“ während der Suche. Es wird oftmals eine deutliche Beschränkung des Suchraumes und damit eine schnelle

Bestimmung der Zuordnung erreicht. Die Repräsentation von Modellen durch die Konjunktion von Constraints legt es nahe, die Modelle als logische Regeln aufzufassen, die beschreiben, wann eine Menge von Bildmerkmalen eine Gebäudeansicht darstellt. Die Realisierung des Konzepts erfolgt deshalb auf Basis des Constraint Logic Programming, einer Erweiterung der logischen Programmierung um spezielle Constraint-Techniken. Da zur Erfüllung der Regeln kein Constraint verletzt sein darf, werden in diesem Kapitel zunächst die idealisierenden Annahmen gemacht, daß alle Modellmerkmale der Gebäudeansicht im Bild beobachtbar sind und daß es eine Zuordnung gibt, die alle Constraints erfüllt.

Im weiteren Verlauf des Kapitels werden zunächst die Grundlagen der Constraint-Techniken erläutert (Abschnitt 3.1). Es wird gezeigt, wie sich das Problem der Bestimmung des Subgraph-Isomorphismus von Modell- zu Bilddaten auf ein Constraint-Problem abbilden läßt, wobei die Berechnungskomplexität verdeutlicht wird. Im anschließenden Abschnitt 3.2 werden die benötigten Grundbegriffe des Constraint Logic Programming erklärt. In Abschnitt 3.3 werden dann die zur relationalen Repräsentation von Gebäudeansichten verwendeten geometrischen und topologischen Constraints vorgestellt, und die relationalen Modell- und Bildstrukturen auf Planarität hin untersucht, die eine Anwendung besonders effizienter Subgraph-Isomorphismus-Algorithmen ermöglichen würde. In Abschnitt 3.4 wird die Realisierung des Suchverfahrens in CLP erläutert, und die spezifischen Vorteile dieser Realisierung u.a. an einem praktischen Beispiel mit realen Luftbilddaten verdeutlicht. Schließlich wird in Abschnitt 3.5 gezeigt, wie sich das Problem der Mehrfachzuordnungen, das bei relationalen Zuordnungsverfahren in Verbindung mit der Fragmentierung entsteht, mit den Mitteln von CLP auf elegante Weise lösen läßt, indem während des Suchvorgangs eine modellbasierte Gruppierung vorgenommen wird.

### 3.1 Constraint Satisfaction Probleme

In diesem Abschnitt werden die erforderlichen Grundlagen und Begriffe zum Thema Constraint Satisfaction Probleme, zukünftig abgekürzt durch CSP, erläutert. Eine generelle Einführung wird in den Übersichtsartikeln [Mes89] und [Kum92] gegeben.

Der CSP-Formalismus bezeichnet eine Klasse von Problemen, die sich auf die im folgenden definierten, deklarativen Weise formulieren lassen:

**Definition 3.1 (Constraint Satisfaction Problem, CSP)**

Ein *Constraint Satisfaction Problem* wird durch ein Tripel  $CSP(V, D, C)$  beschrieben, wobei  $V = \{v_1, \dots, v_n\}$  eine endliche Menge von Variablen mit assoziierten Wertebereichen  $D = \{D_1, \dots, D_n\}$  mit  $\{v_1 : D_1, \dots, v_n : D_n\}$  ist.  $C$  ist eine endliche Menge von Constraints, wobei jeder Constraint  $c_i(V_i)$  eine Teilmenge  $V_i = \{v_{i_1}, \dots, v_{i_l}\} \subseteq V$  der Variablen zueinander in Relation setzt und deren gültige Wertekombinationen auf eine Teilmenge von  $\{D_{i_1} \times \dots \times D_{i_l}\}$  beschränkt.  $\square$

Constraints über eine Variable heißen unär, über zwei Variablen binär und über drei Variablen ternär. CSP, die nur binäre Constraints enthalten, lassen sich als Graph darstellen, indem Variablen durch Knoten und Constraints als Kanten zwischen den entsprechenden Knoten repräsentiert werden.

**Definition 3.2 (Lösung eines CSP)**

Eine *Lösung eines CSP* besteht aus einer eindeutigen Belegung der Variablen mit genau einem Wert aus dem jeweiligen Wertebereich, so daß alle Constraints erfüllt sind. Der *Lösungsraum eines CSP* wird durch alle unterschiedlichen Lösungen aufgespannt.  $\square$

Da auch unendliche Wertebereiche wie z.B. die Menge der natürlichen Zahlen  $\mathbb{N}$  oder Intervalle reeller Zahlen zugelassen sind, ist der Lösungsraum in diesen Fällen im allgemeinen ebenfalls unendlich. Eine wichtige spezielle Klasse von CSP, bei der alle Wertebereiche diskret und endlich

sind, wurde von Mackworth bereits in den siebziger Jahren untersucht [Mac77] und später von Haralick und Shapiro wie folgt formal definiert [HS79]:

**Definition 3.3 (Consistent Labeling Problem)**

Ein *Consistent Labeling Problem* ist ein CSP, bei dem alle Wertebereiche  $D_1, \dots, D_n$  diskret und endlich sind. □

**Satz 3.4 (Berechnungskomplexität des Consistent Labeling Problems)**

Das Consistent Labeling Problem ist NP-vollständig<sup>1</sup>. □

Da bei einem Consistent Labeling Problem alle Wertebereiche endlich sind und alle Lösungen voneinander verschieden sein müssen, ist auch der Lösungsraum endlich. Die maximale Zahl unterschiedlicher Lösungen ergibt sich aus dem kartesischen Produkt aller Wertebereiche  $D_1 \times \dots \times D_n$ . Die Kardinalität dieser Menge und somit auch der Berechnungsaufwand wachsen exponentiell bzgl. der Anzahl der Variablen. In [HS79] wird durch Reduktion des Consistent Labeling Problems auf das SAT-Problem [HU90, EC93] die NP-Vollständigkeit formal bewiesen.

Die Formulierung eines Problems als Menge von Randbedingungen über eine Menge von Variablen abstrahiert von einem konkreten Lösungsverfahren. Insofern stellt das CSP-Schema zunächst ein Konzept zur formalen, deklarativen Repräsentation einer Klasse von Problemen dar. Sachverhalte werden aber in der Regel als CSP formuliert, weil sie die Beantwortung bestimmter Fragestellungen durch spezifische Lösungsverfahren auf effiziente Weise unterstützen. Diese Fragestellungen sind zum einen die Fragen nach der Existenz einer Lösung bzw. bei mehreren Lösungen nach ihrer Anzahl, und zum anderen die Frage nach den konkreten Ausprägungen der Lösungen. Bei Optimierungsproblemen stellt sich darüberhinaus die Frage nach der besten Lösung.

Die Trennung von Problemstellung und Lösungsverfahren ermöglicht die Verwendung spezieller Algorithmen für CSPs unterschiedlicher Wertebereiche. Beim Consistent Labeling Problem ist die Aufgabe der Entscheidung, ob eine Lösung existiert, äquivalent zur Bestimmung einer konkreten Lösung [Mes89]. Deswegen bestehen die zwei wesentlichen Aufgaben in der Suche nach einer (der erstbesten) Lösung und in der Aufzählung aller Lösungen.

### 3.1.1 Das Subgraph-Isomorphismus-Problem als CSP

In diesem Abschnitt soll gezeigt werden, wie sich die Bestimmung von Subgraph-Isomorphismen auf ein Constraint Satisfaction Problem abbilden läßt (vgl. [HS79]). Dazu wird die von Haralick und Shapiro in [SH87, HS93a] vorgestellte Notation verwendet, bei der eine Relationalstruktur  $D = (P, R)$  durch eine Menge von Primitiven  $P$  und den darauf bestehenden Relationen  $R$  beschrieben wird. Wenn  $R$  nur binäre Relationen enthält, entspricht  $D$  der klassischen Definition eines Graphen. Die folgenden Definitionen beziehen sich auf binäre Relationen, sind aber auf Relationen mit beliebiger Stelligkeit übertragbar.

**Definition 3.5 (Komposition)**

Die Komposition der Relationen  $R_1$  einer Relationalstruktur  $D_1 = (P_1, R_1)$  mit einer Abbildungsfunktion  $h : P_1 \rightarrow P_2$  beschreibt eine Relationenmenge auf den Primitiven  $P_2$  einer zweiten Relationalstruktur  $D_2 = (P_2, R_2)$ :

$$R_1 \circ h = \{(q_i, q_j) \in P_2 \times P_2 \mid (p_i, p_j) \in R_1 \text{ mit } h(p_i) = q_i \wedge h(p_j) = q_j\}. \quad (3.1)$$

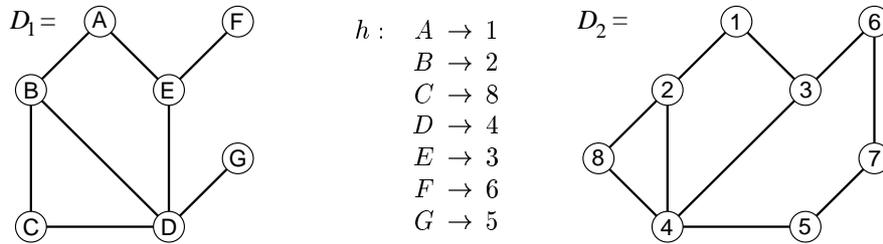
□

---

<sup>1</sup>NP-vollständig bedeutet, daß das Problem nur mit einem nichtdeterministischen Algorithmus in polynomieller Zeit gelöst werden kann. Es besteht die starke Vermutung, daß es keinen deterministischen Algorithmus gibt, der das Problem in polynomiell beschränkter Zeit lösen kann. NP-vollständige Probleme gehören insgesamt zu den schwierigsten Problemen der Informatik (siehe [EC93]).

**Definition 3.6 (Subgraph-Isomorphismus)**

Eine Relationalstruktur  $D_1 = (P_1, R_1)$  besitzt einen Subgraph-Isomorphismus mit  $D_2 = (P_2, R_2)$ , gdw. es eine injektive Abbildung  $h : P_1 \mapsto P_2$  gibt, so daß gilt:  $R_1 \circ h \subseteq R_2$ .  $\square$



**Abbildung 3.1:** Beispiel für einen Subgraph-Isomorphismus. Zuordnung  $h$  bildet alle Knoten des (kleineren) linken Graphen  $D_1$  auf Knoten des (größeren) rechten Graphen  $D_2$  ab, wobei die zugeordneten Knoten alle Relationen erfüllen, die auch zwischen den ursprünglichen Knoten bestehen. Das Problem besteht darin, die in der Mitte dargestellte Zuordnungsfunktion  $h$  effizient zu bestimmen.

Abbildung 3.1 zeigt ein Beispiel für die Bestimmung eines Subgraph-Isomorphismus. Zur Repräsentation des Problems als CSP werden alle Knoten  $\{A, B, C, D, E, F, G\}$  des kleineren Graphen als Variablen interpretiert, denen jeweils ein Knoten  $\{1, 2, 3, 4, 5, 6, 7, 8\}$  des größeren Graphen zugewiesen werden kann. Die ersten Constraints ergeben sich aus der Forderung nach der Injektivität der Zuordnungsfunktion  $h$ : alle Werte der Variablen müssen paarweise verschieden sein. Für jede Kante  $(v, w) \in R_1$  wird ein Constraint der Form  $(V, W) \in R_2$  generiert, wobei  $V$  und  $W$  die Namen der zu  $v$  bzw.  $w$  korrespondierenden Variablen sind. Für das obige Beispiel würden die Constraints  $(A, B) \in R_2$ ,  $(A, E) \in R_2$ ,  $(E, F) \in R_2$  usw. generiert. Zur Bestimmung der Zuordnungsfunktion  $h$  müssen jeder Variablen ein Knoten zugeordnet werden, so daß alle Constraints erfüllt sind.

### 3.1.2 Globale Lösungsverfahren für das Consistent Labeling Problem

Lösungsverfahren für CSP werden nach zwei Kriterien unterschieden: systematisch bzw. stochastisch und global bzw. lokal. Die in diesem Abschnitt behandelten systematischen Lösungsverfahren heißen global, weil sie zur Bestimmung von Lösungen das CSP in seiner gesamten Struktur betrachten. Sie sind vollständig in dem Sinne, daß sie immer entscheiden können, ob das CSP Lösungen besitzt, und welche das sind.

Im Gegensatz zu stochastischen Verfahren wie z.B. Simulated Annealing [KGV83] oder genetischen Algorithmen [EC93] garantieren systematische Verfahren, daß alle existierenden Lösungen gefunden werden. Aus diesem Grund und weil, wie im folgenden dargelegt wird, systematische Verfahren die Eigenschaften von Constraint-Netzwerken auf spezifische Weise ausnutzen, wird in dieser Arbeit auf stochastische Verfahren nicht weiter eingegangen.

#### 3.1.2.1 Generate & Test

Bei diesem Verfahren werden zunächst alle Variablen mit einem (deterministisch oder nichtdeterministisch gewählten) Wert aus ihrem jeweiligen Wertebereich belegt. Anschließend wird überprüft, ob alle Constraints erfüllt sind. Ist dies nicht der Fall, so wird eine neue Belegung bestimmt und das ganze wiederholt.

Da dieses Verfahren beim Vorschlagen einer Variablenbelegung keinerlei Nutzen aus den gegebenen Constraints zieht, hat es unabhängig von spezifischen Problemen generell ein extrem schlechtes

Laufzeitverhalten. Wenn dieses Verfahren zur Bestimmung aller Lösungen verwendet werden soll, muß immer der vollständige Suchraum durchlaufen werden.

### 3.1.2.2 Baumtiefensuche mit Backtracking

Bei der Tiefensuche wird systematisch eine Variable ausgewählt und ihr ein Wert aus ihrem Wertebereich zugewiesen. Wenn der Wert mit allen bislang zugewiesenen Variablenwerten kompatibel ist, wird die nächste Variable ausgewählt usw. Falls der selektierte Wert zu einer Verletzung eines Constraints führt, wird der nächste Wert aus dem Wertebereich probiert, bis ein kompatibler Wert gefunden wurde, oder alle Werte des Wertebereichs probiert wurden. Im letzten Fall wird eine Ebene zurückgegangen (Backtracking) und für die davor ausgewählte Variable ein neuer Wert probiert. Der Ablauf der Suche läßt sich als *Suchbaum* repräsentieren, wobei jeder Knoten und jedes Blatt die Festlegung einer Variablen auf einen Wert darstellt (vgl. [Nil82]). Der Baum besitzt genauso viele Ebenen wie das CSP Variablen.

Der Vorteil gegenüber dem naiven Generate & Test besteht darin, daß nach der Festlegung einer Variablen auf einen Wert zunächst überprüft wird, ob dieser Wert mit den bisherigen Zuweisungen kompatibel ist, bevor mit der Zuweisung der nächsten Variablen weitergemacht wird. Da jedoch nicht überprüft wird, ob der ausgewählte Wert zu einem Konflikt mit den noch festzulegenden Variablen führt, werden Fehlschläge oftmals erst zu einem relativ späten Zeitpunkt der Suche festgestellt (vgl. [HE80]).

### 3.1.2.3 Baumtiefensuche mit Forward-Checking

Forward-Checking ergänzt die Baumtiefensuche um eine Komponente zur vorausschauenden Auswertung (siehe [HS79]). Sobald eine Variable ausgewählt und ihr ein Wert zugewiesen wurde, werden anhand der zu dieser Variablen inzidenten Constraints alle Werte aus den Wertebereichen der noch freien Variablen eliminiert, die inkompatibel mit dem aktuell gewählten Wert sind. Dies führt dazu, daß die Wertebereiche der Variablen bereits vor ihrer Festlegung reduziert werden, und damit der Suchraum dramatisch verkleinert wird. Darüberhinaus kann es vorkommen, daß durch das Forward-Checking der Wertebereich einer noch uninstantierten Variablen leer wird. Das bedeutet, daß bereits früh festgestellt werden kann, daß zu einem späteren Zeitpunkt der Suche keine Lösung mehr gefunden werden kann, weil für eine Variable kein passender Wert mehr existiert. In dieser Situation kann sofort ein Backtracking ausgelöst werden.

### 3.1.2.4 Baumtiefensuche mit Look-Ahead

Look-Ahead baut auf dem Prinzip des Forward-Checking auf, und führt weitere Konsistenzüberprüfungen zwischen den noch nicht festgelegten Variablen durch (siehe [HS79]). Es wird für jeden Wert aller bislang uninstantierten Variablen anhand der inzidenten Constraints überprüft, ob ein dazu kompatibler Wert in den Wertebereichen der anderen (feststehenden und freien) Variablen existiert. Falls dies nicht der Fall sein sollte, wird der Wert aus dem Wertebereich der entsprechenden Variablen eliminiert.

Look-Ahead erreicht zwar im Gegensatz zu Forward-Checking eine stärkere Beschränkung des Suchraumes, führt jedoch eine deutlich größere Anzahl an Konsistenztests für die einzelnen Constraints durch. Haralick und Elliott haben in [HE80] die Performanz der verschiedenen Baumsuchverfahren zur Lösung von CSP untersucht, wobei sich herausgestellt hat, daß die Anwendung des Forward-Checking in den meisten Fällen bzw. für die meisten Anwendungen aufgrund der im Vergleich zu Look-Ahead kleineren Zahl an Konsistenztests zu der kürzesten Laufzeit geführt hat.

### 3.1.3 Lokale Lösungsverfahren: Konsistenztechniken

Globale Lösungsverfahren besitzen eine zentrale Kontrollstrategie; die Constraints nehmen dabei eine passive Rolle ein. Lokale Lösungsverfahren sehen die Constraints als aktive Komponenten eines Netzwerkes, in dem jeder Constraint eigene Wertebereichseinschränkungen auf seinen inzidenten Variablen vornehmen kann. Durch die Verbindung der Constraints über die Variablen werden Bereichseinschränkungen durch das Netz propagiert. Diese Idee wurde von Waltz zur Interpretation von Linienzeichnungen dreidimensionaler Objekte verwendet (siehe [Wal75]). Mackworth griff die Idee auf, und entwickelte das allgemeine Konzept der Konsistenztechniken (vgl. [Mac77]), die sich im wesentlichen durch das Prinzip „Einschränken und Propagieren“ beschreiben lassen. Zwei wichtige Begriffe sind die Knoten- und die Kantenkonsistenz, die wie folgt definiert sind:

**Definition 3.7 (Knotenkonsistenz (Node Consistency))**

Eine Variable  $v$  mit assoziiertem Wertebereich  $d$  ist knotenkonsistent, gdw. für jeden Wert  $x \in d$  die unären Constraints  $c(v)$  erfüllt sind. □

**Definition 3.8 (Kantenkonsistenz (Arc Consistency))**

Ein Constraint  $c(v_1, v_2)$  auf den Variablen  $v_1$  und  $v_2$  mit den assoziierten Wertebereichen  $d_1$  und  $d_2$  ist kantenkonsistent, gdw. für jeden Wert  $x_1 \in d_1$  ein Wert  $x_2 \in d_2$  existiert, so daß  $c(x_1, x_2)$  erfüllt ist. Bei dem allgemeinen Fall der ungerichteten Kantenkonsistenz muß dies auch andersherum gelten, d.h. für jeden Wert  $x_2 \in d_2$  muß ein Wert  $x_1 \in d_1$  existieren, der den Constraint erfüllt. Ein Constraint-Netzwerk ist kantenkonsistent, gdw. alle Constraints kantenkonsistent sind. □

Die Erzwingung der Knotenkonsistenz beschränkt die Wertebereiche der Variablen vorab auf bestimmte, durch unäre Constraints erlaubte Werte. Die Kantenkonsistenz sorgt dafür, daß die Wertebereiche der Variablen keine Werte mehr beinhalten, die mit den unmittelbar über einen Constraint benachbarten Variablen inkompatibel sind. Bei der Anwendung der Linieninterpretation können bereits viele Zeichnungen alleine durch Erzwingung der Kantenkonsistenz und Propagierung eindeutig interpretiert werden (siehe [Wal75]). Es können dabei verschiedene Algorithmen zur Herstellung der Kantenkonsistenz eines CSP eingesetzt werden (siehe [Mac77, MM88, DvH91]). Die Kantenkonsistenz reicht im allgemeinen jedoch nicht aus, um eine eindeutige Lösung des CSP zu bestimmen, oder festzustellen, daß es keine Lösung gibt. So existiert z.B. für das CSP mit den drei Variablen  $A : \{1, 2\}, B : \{1, 2\}, C : \{1, 2\}$  und den Constraints  $A \neq B, B \neq C, C \neq A$  keine Lösung, obwohl es kantenkonsistent ist.

#### 3.1.3.1 Generieren von Werten

Wenn nach dem Abschluß der Propagierung von Einschränkungen keine global eindeutige Lösung vorliegt, wird — analog zu den Baumsuchverfahren — eine Variable ausgewählt, und ihr ein fester Wert zugewiesen. Die Zuweisung stellt eine Wertebereichseinschränkung dar, die wieder durch das Constraint-Netz propagiert wird. Diese Schritte werden wiederholt, bis alle Variablen einen eindeutigen Wert besitzen. Bei der Generierung von Werten gibt es zwei verschiedene Freiheitsgrade:

1. Die Auswahl der Variablen, die auf einen Wert festgelegt wird, und
2. die Auswahl eines Wertes aus dem Wertebereich dieser Variablen.

Die Auswahl kann anhand von Heuristiken erfolgen, wobei man diese danach unterscheidet, ob die Bestimmung der Reihenfolge statisch oder dynamisch erfolgt (vgl. [HE80, Nud83, DP88, FSB89]). Bei statischen Heuristiken wird vor dem Suchprozeß eine feste Reihenfolge für die Auswahl der Variablen und der Werte bestimmt. Dynamische Heuristiken wählen in jedem Zustand der Suche die Reihenfolge nach bestimmten Kriterien dynamisch aus. Der Vorteil der dynamischen gegenüber den statischen Heuristiken ist, daß sie die die aktuelle Beschränkung des Suchraums aktiv zur Steuerung während der Suche verwenden.

Eine der effizientesten Heuristiken zur Variablenauswahl ist die sogenannte *First-Fail-Heuristik*. Sie wählt als nächste zu instantiierende Variable jeweils diejenige mit dem kleinsten Wertebereich aus. Diese Heuristik wurde von Haralick und Elliott in [HE80] vorgestellt und empirisch evaluiert. Aufgrund der guten Performanz im Vergleich zu anderen Heuristiken (vgl. [Nud83]) wird in dieser Arbeit die First-Fail-Heuristik zur Suche nach den Gebäudeansichten verwendet.

## 3.2 Constraint Logic Programming

Die Formulierung eines CSP als konjunktive Verknüpfung von Constraints legt es nahe, das CSP als logische Regel zu interpretieren. Constraint Logic Programming (CLP) erlaubt die effiziente Auswertung dieser Regeln unter Verwendung der Constraint-Techniken, und bietet überdies alle Möglichkeiten der logischen Programmierung in Form einer universellen Programmiersprache. Fakten, Regeln, Terme und Listen können zur strukturierten Wissensrepräsentation verwendet werden. Darüberhinaus erlaubt CLP die Realisierung hybrider Verfahren, bei denen die Constraint-Techniken mit anderen Suchverfahren kombiniert werden können. Dies spielt weiter unten in Abschnitt 3.5 eine Rolle bei der Gruppierung fragmentierter Linienmerkmale. Einen informellen Überblick über CLP gibt [FHK<sup>+</sup>92]. Die allgemeine Theorie wurde von Jaffar, Lassez und Maher entwickelt und wird in [JM94] beschrieben. Die Beziehung von CLP zu CSP wird in [vHSD92, Mac92] näher untersucht. Einen Überblick über bestehende Anwendungen von CLP geben Beierle und Plümer in [BP95] und Frühwirth in [FA97a].

### 3.2.1 Logische Programmierung und PROLOG

Die logische Programmierung beruht auf dem Paradigma der deklarativen Programmierung, bei dem es darum geht, die Lösung und nicht den Lösungsweg für ein Problem zu beschreiben (siehe [Llo87]). Einen Ansatz zur Realisierung dieses Konzepts stellt die logische Programmiersprache PROLOG dar. PROLOG-Programme bestehen aus Fakten und Regeln, und werden nach einer festen Strategie ausgewertet (vgl. [CM84, SS94]). Eine Regel (Klausel)  $A \leftarrow B_1, \dots, B_n$  besteht aus dem Regelkopf  $A$  und dem Regelrumpf  $B_1, \dots, B_n$ , wobei  $B_1, \dots, B_n$  eine Konjunktion der einzelnen Ziele  $B_i$  darstellt. Die Ziele  $B_i$  können einfache logische Formeln sowie Aufrufe von Prädikaten (andere Regeln) sein. Der Regelrumpf ist die Prämisse und der Kopf die Konklusion, d.h. wenn alle Bedingungen  $B_i$  wahr sind, wird daraus geschlossen, daß  $A$  wahr ist. Fakten sind einfache Regeln der Form  $A \leftarrow \text{wahr}$ . Es können mehrere gleichnamige Regeln existieren, wobei diese dann disjunktiv verknüpft sind.

Bei der Auswertung einer Anfrage  $\leftarrow G_1, \dots, G_k$ , die aus der Konjunktion der einzelnen Teilziele  $G_i$  besteht, werden die Teilziele von links nach rechts abgearbeitet. Es wird zuerst eine zu  $G_1$  passende Klausel  $A \leftarrow B_1, \dots, B_n$  gesucht. Die Variablen von  $A$  und  $G_1$  werden unifiziert, und der Klauselrumpf  $B_1, \dots, B_n$  ersetzt dann  $G_1$ , so daß der nächste Berechnungszustand durch  $\leftarrow B_1, \dots, B_n, G_2, \dots, G_k$  gegeben ist. Anschließend wird die Auswertung mit  $B_1$  fortgesetzt. Wenn es bei der Auswahl der passenden Klausel verschiedene Möglichkeiten gibt, merkt sich PROLOG die entsprechenden Stelle. Falls im weiteren Verlauf der Auswertung eine Regel fehlschlägt, wird ein Backtracking an die letzte Stelle erzwungen, an der noch eine andere Wahl möglich ist, und mit dieser die Auswertung fortgesetzt.

### 3.2.2 Von LP zu CLP

Die Auswertestrategie von PROLOG realisiert eine Tiefensuche mit Backtracking. Die gestellten Ziele einer Anfrage werden dabei in strenger Reihenfolge abgearbeitet. Diese Bedingung wird in CLP gelockert. Constraints  $c(v_1, \dots, v_n)$  stellen spezielle Prädikate dar, deren Auswertung verzögert stattfinden kann. Die Bedingungen für eine Verzögerung werden von dem Zustand der Variablen  $v_1, \dots, v_n$  abhängig gemacht. Die Auswertestrategie läuft bei einer Anfrage  $\leftarrow c_1, \dots, c_k$  zwar

weiterhin „von links nach rechts“, wenn jedoch beispielsweise für den Constraint  $c_1$  noch nicht eindeutig entschieden werden kann, ob er erfüllt oder verletzt ist, wird er suspendiert und in den sogenannten *Constraint-Speicher*  $\sigma$  verlagert. Die Auswertung wird dann mit  $c_2$  fortgesetzt usw. Der aktuelle Berechnungszustand wird nicht mehr ausschließlich durch die Liste der noch zu beweisenden Ziele  $G$  gebildet, sondern durch das Paar  $\langle G \sqcap \sigma \rangle^2$ . Um eine Antwort auf eine Anfrage zu erhalten, müssen nicht nur alle Ziele bewiesen sein, es muß auch für die Constraints im Constraint-Speicher entschieden werden können, ob bzw. durch welche Variablenbelegung sie erfüllbar sind.

Die Gestalt der Entscheidungsprozedur hängt zwar vom jeweils verwendeten Wertebereich von CLP ab, jedoch basieren alle Verfahren auf der Erkennung der folgenden Situationen (die Definitionen sind nach [Jam96] formuliert, die deutsche Übersetzung der Begriffe erfolgt nach [FA97b]):

**Definition 3.9 (Wertebelegung, Erfüllung eines Constraints, Lösung)**

Eine *Wertebelegung* ist die Belegung der Variablen des CSP mit genau einem Wert aus ihrem jeweils assoziierten Wertebereich. Eine Wertebelegung *erfüllt* einen Constraint, wenn der Constraint wahr ist unter der gegebenen Wertebelegung. Eine *Lösung* zu einem CSP ist eine Wertebelegung, die alle Constraints des Problems erfüllt. □

**Definition 3.10 (Folgerung (Entailment))**

Ein Constraint  $c$  *folgt* aus einer Menge von Constraints  $\sigma$  (engl.: *is entailed*), wenn bei jedem Modell (Wertebelegung), bei dem alle Constraints in  $\sigma$  wahr sind,  $c$  ebenfalls wahr ist. Notation:  $\sigma \models c$ . □

**Definition 3.11**

Ein Constraint  $c$  *folgt nicht* aus einer Menge von Constraints  $\sigma$ , wenn es mindestens ein Modell gibt, bei dem alle Constraints in  $\sigma$  wahr sind, aber  $c$  unwahr ist. Notation:  $\sigma \not\models c$ . □

**Definition 3.12 (Widerspruch (Contradiction))**

Ein Constraint  $c$  *widerspricht* einer Menge von Constraints  $\sigma$ , wenn es kein Modell gibt, bei dem alle Constraints in  $\sigma$  wahr sind und  $c$  ebenfalls wahr ist. Notation:  $\sigma \wedge c \models \perp$ . □

**Definition 3.13**

Ein Constraint  $c$  ist *konsistent* mit einer Menge von Constraints  $\sigma$ , wenn es mindestens ein Modell gibt, bei dem alle Constraints in  $\sigma$  wahr sind und  $c$  ebenfalls wahr ist. Notation:  $\sigma \wedge c \not\models \perp$ . □

Die beiden Situationen der Folgerung und des Widerspruchs eines Constraints spielen eine zentrale Rolle bzgl. der Effizienz von CLP. Die Folgerung (Entailment) eines Constraints aus dem Constraint-Speicher beschreibt eine Situation, in der Vereinfachungen des Constraint-Speichers während der Suche möglich sind. Ein Constraint, der aus den anderen Constraints in  $\sigma$  folgt, erlaubt die gleichen Einschränkungen wie der Rest der Constraints. Er ist demnach redundant und wird aus dem Constraint-Speicher entfernt, was zur Steigerung der Performanz beiträgt. Die Erkennung eines widersprüchlichen Constraints während der Suche ist ebenfalls von großem Interesse, weil ein solcher Constraint niemals mehr erfüllt werden kann, und damit die Suche an der aktuellen Stelle im Suchbaum sofort abgebrochen werden kann. Es wird ein Backtracking zum letzten Punkt im Suchbaum durchgeführt, an dem eine Instantiierung vorgenommen wurde und dabei noch weitere Werte möglich waren. Bei einem Backtracking wird neben dem Zustand der noch auszuwertenden Ziele immer auch der entsprechende Zustand des Constraint-Speichers rekonstruiert.

---

<sup>2</sup>Die Notation  $\langle G \sqcap \sigma \rangle$  erfolgt nach [vHSD92]. '□' dient dabei als Trennzeichen.

### 3.2.3 CLP auf endlichen Wertebereichen

Constraint Logic Programming auf endlichen Wertebereichen, CLP(FD), stellt eine Umsetzung der in Abschnitt 3.1.3 vorgestellten Konsistenztechniken dar (siehe [vHD86, vHD87, vH87, vH89]). Jede Bereichsvariable kann dabei Werte aus einem ihr zugeordneten endlichen Wertebereich annehmen. Die Realisierung der beiden Konsistenztechniken Forward-Checking und Look-Ahead erfolgt mittels spezieller Inferenzregeln. Van Hentenryck hat in [vH89] gezeigt, daß die durch die Inferenzregeln vorgenommenen Suchraumeinschränkungen gleichmächtig zu den oben beschriebenen Realisierungen sind.

Die Forward-Checking-Inferenzregel (FCIR) besagt, daß ein Constraint suspendiert und erst geweckt wird, wenn alle bis auf eine Bereichsvariable an einen Wert gebunden sind. Der Vorteil dieser Vorgehensweise besteht darin, daß nach dem Wecken des Constraints immer entschieden werden kann, ob er definitiv erfüllt oder widersprüchlich ist. Im ersten Fall wird der Wertebereich der noch freien Bereichsvariable anhand des Constraints eingeschränkt, und der Constraint aus dem Constraint-Speicher entfernt. Im zweiten Fall wird unmittelbar ein Backtracking ausgelöst.

Die Look-Ahead-Inferenzregel (LAIR) besagt, daß ein Constraint solange suspendiert wird, bis er entweder als definitiv erfüllt oder widersprüchlich erkannt wird. Um dies festzustellen, wird im Gegensatz zur FCIR der Constraint immer geweckt, wenn der Wertebereich einer seiner Variablen verkleinert wurde. Es wird dann anhand der Definition des Constraints überprüft, ob diese Wertebereichseinschränkung auf andere Variablen propagiert werden kann. Wenn alle Constraints mittels LAIR ausgewertet werden, sind die im Constraint-Speicher befindlichen Constraints nach jedem Abschluß der Propagierungen kantenkonsistent (vgl. [vH89]).

Da in der Regel mit den beiden Inferenzregeln keine global konsistente Variablenbelegung erreicht wird, müssen Variablen auf Werte aus ihrem Wertebereich festgelegt werden. Zur Auswahl der nächsten zu instantiierenden Variablen steht in CLP(FD) die First-Fail-Heuristik zur Verfügung. Die Verwendung der Look-Ahead-Inferenzregel kommt zwar insgesamt mit weniger Suchbaumknoten aus, jedoch führt das häufige Wecken zu einer deutlich größeren Anzahl von Konsistenztests, wodurch der Zeitaufwand der Suche in der Regel größer wird, als bei Verwendung der Forward-Checking-Inferenzregel (vgl. [HE80, vH89]).

## 3.3 Modellierung von Gebäudeansichten durch Constraint-Mengen

Im Bereich der Gebäudeerkennung wurden relationale Gebäuderepräsentationen in Form von Constraint-Mengen bislang nur von Mohan und Nevatia (vgl. [MN88] und Abschnitt 2.3.1) sowie Schutte et. al. (vgl. [Sch96] und Abschnitt 2.3.3) eingesetzt. Mohan und Nevatia leiten zur Verifikation generierter Gebäudehypothesen ein Constraint-Netzwerk her, dessen Konsistenz mittels eines Neuronalen Netzes bewertet wird. Die Constraint-Repräsentation dient hier nicht dem Auffinden sondern ausschließlich der Bewertung der Gebäudehypothesen. Bei dem Ansatz von Schutte et. al. werden Constraint-Techniken zur Identifikation möglicher Gebäudeaspekte aus dem Aspektgraphen der beiden Hausmodelle Flachdach- und Satteldachhaus eingesetzt. Die verwendeten Relationen beziehen sich jedoch ausschließlich auf Flächenmerkmale und ihre morphologischen Eigenschaften, und sind deshalb besonders anfällig gegenüber Bildstörungen und Fehlern der Merkmalsextraktion.

Relationale Modellrepräsentationen werden häufiger zur Bilderkennung im Bereich der industriellen Fertigungsüberwachung eingesetzt (vgl. [SFH92, Rad93]), wobei die Modelle eine feste Ausdehnung und Form besitzen (CAD-Modelle). Flynn und Jain beschreiben in [FJ91] einen Ansatz, der auf der Verwendung von Aspektgraphen basiert, und bei dem die relationalen Modellrepräsentationen einmal für alle möglichen Aspekte der Modelle hergeleitet und in einer Datenbank gespeichert werden. Diese Vorgehensweise ist bei der in dieser Arbeit vorgestellten Strategie aufgrund der gene-

rischen Modellierung nicht möglich; die Constraint-Mengen müssen dynamisch aus den generierten Gebäudeansichten erzeugt werden.

Die Anwendung von Constraint-Techniken setzt voraus, daß alle Relationen kategorisch zwischen „erfüllt“ und „nicht erfüllt“ unterschieden werden können. Aus diesem Grund werden die geometrischen Constraints über Schwellwerte definiert. Da einzelne Constraints keine abgestufte Bewertung ermöglichen, definieren sie ein verhältnismäßig grobes Raster. Um eine ausreichende Spezifität der Modelle zu erreichen, müssen diese durch ausdrucksstarke Constraints repräsentiert werden, und die Modelle gleichzeitig möglichst groß sein. Aus diesem Grund werden für die Suche von Gebäuden im Bild die Constraint-Repräsentationen der vollständigen Gebäudeansichten verwendet.

Im folgenden Unterabschnitt werden zunächst die zur relationalen Repräsentation von Gebäudeansichten verwendeten geometrischen und topologischen Constraints vorgestellt. Radiometrische Constraints werden z.Zt. nicht verwendet, weil bislang noch kein Beleuchtungsmodell berücksichtigt wurde<sup>3</sup>. Im darauffolgenden Unterabschnitt wird gezeigt, wie die Schwellwerte der Constraints empirisch anhand von Trainingsdaten ermittelt werden.

### 3.3.1 Geometrische und topologische Constraints

Zur Modellierung der Gebäudeansichten werden insgesamt sechs verschiedene Arten von Constraints verwendet. Der einzige topologische Constraint ist die Merkmalsnachbarschaft, die besagt, daß ein Merkmal räumlich unmittelbar an ein anderes grenzen muß. Merkmalsnachbarschafts-Constraints wurden (mit z.T. leicht abweichender Definition) bereits in vielen Verfahren zur Gebäude- bzw. Objekterkennung verwendet (vgl. [NP82, Bro83, HK87, FJ91, Vos92, Sch96, Wal96, DH97]). Der erste geometrische Constraint ist die Parallelität zweier Linienmerkmale. Auch dieser Constraint wurde bereits in einigen Verfahren eingesetzt (vgl. [MN88, WHK90]). Daneben wird die Kollinearität zwischen einer Linie und einem Punktmerkmal als Constraint definiert, sowie der absolute Winkel von Linienmerkmalen bzgl. der horizontalen Bildachse (vgl. [Wal96]). Einen besonderen geometrischen Constraint stellt der sogenannte `Same_Side_Lines`-Constraint dar, der drei Linienmerkmale zueinander in Beziehung setzt, und fordert, daß zwei Linienmerkmale auf derselben Seite der dritten Linie liegen müssen. Daneben gibt es schließlich noch den Verschiedenheits-Constraint, der erzwingt, daß zwei Modellmerkmalen unterschiedliche Bildmerkmale zugeordnet werden müssen.

Die Constraints werden in den sechs folgenden Unterabschnitten im Detail erläutert, wobei jeweils zunächst die formale Definition angegeben wird. Anschließend werden die Eigenschaften der Constraints und ihrer zugrundeliegenden Relationen diskutiert. Es wird gezeigt, wie für eine Gebäudeansicht die Constraints der jeweiligen Art anhand des Gebäudemodells hergeleitet werden. Schließlich wird noch für jeden Constraint gezeigt, daß er invariant bzgl. der verschiedenen geometrischen Ausprägungen der Gebäudeansichten ist.

#### 3.3.1.1 Nachbarschaft zweier Merkmale

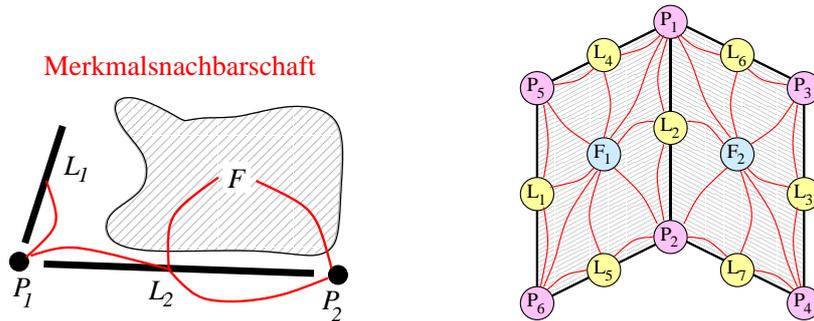
Bei der Nachbarschaft zweier Merkmale handelt es sich um einen binären topologischen Constraint, der wie folgt definiert ist:

**Definition 3.14 (Merkmalsnachbarschafts-Constraint)**

Zwei Merkmale gelten als benachbart, wenn sie räumlich unmittelbar aneinandergrenzen. □

Die Merkmalsnachbarschaftsrelation ist eine binäre, symmetrische und irreflexive Relation. Ihre Berechnung bzgl. der Bilddaten findet im Zuge der Merkmalsextraktion statt (eine detaillierte Beschreibung wird in [För94] und [Fuc97] gegeben), und wird in Form des Merkmalsnachbarschaftsgraphen MNG extensional repräsentiert. Zwei Bildmerkmale sind demnach benachbart,

<sup>3</sup>Einen möglichen Ansatz zur Integration eines Beleuchtungsmodells stellt Steinhage in [Ste97] vor.



**Abbildung 3.2:** Links: Nachbarschaft zwischen Merkmalen unterschiedlicher Klassen. Rechts: Merkmalsnachbarschafts-Constraints eines Aspekts eines Satteldachs. Linienmerkmale sind durch gelbe, Flächenmerkmale durch hellblaue und Punktmerkmale durch lilafarbene Knoten dargestellt.

wenn zwischen ihnen eine Kante im MNG besteht. Abbildung 3.2 zeigt links ein Beispiel für die zwischen extrahierten Bildmerkmalen bestehenden Nachbarschaften.

Die Merkmalsnachbarschafts-Constraints werden aus den Inzidenzbeziehungen von Punkten, Linien und Flächen der Randflächendarstellung (B-Rep) der Gebäudemodelle hergeleitet. Diese Inzidenzbeziehungen besagen, daß Linien durch zwei Endpunkte aufgespannt, und Flächen von einem einfachen Linienzyklus umschlossen werden, wobei sich die Linien nur an ihren Endpunkten treffen. Laut idealem Bildmodell der Merkmalsextraktion (siehe [Fuc97]) können nur Merkmale unterschiedlicher Klassen unmittelbar zueinander benachbart sein (vgl. Abschnitt 2.5.1), weshalb auch jeweils nur Constraints zwischen den Paaren (Linie, Punkt), (Fläche, Linie) und (Fläche, Punkt) generiert werden. Abbildung 3.2 zeigt rechts die für die Ansicht eines Satteldachs generierten Constraints.

Um zu zeigen, daß die Merkmalsnachbarschafts-Constraints invariant bzgl. aller durch einen Aspekt repräsentierten Sichten sind, muß überprüft werden, ob sich die Inzidenzbeziehungen innerhalb eines Aspekts ändern. Eine Änderung der Inzidenzbeziehungen tritt immer dann auf, wenn neue Merkmale des Modells sichtbar werden, oder bislang sichtbare Merkmale verdeckt und somit unsichtbar werden. Dabei würde sich aber auch die Topologie der Ansicht verändern. Da die Aspekte jedoch die jeweils topologisch isomorphen Sichten zusammenfassen, bedeutet dies für die Inzidenzbeziehungen und damit die Merkmalsnachbarschaftsrelationen, daß sie invariant innerhalb eines Aspektes sind.

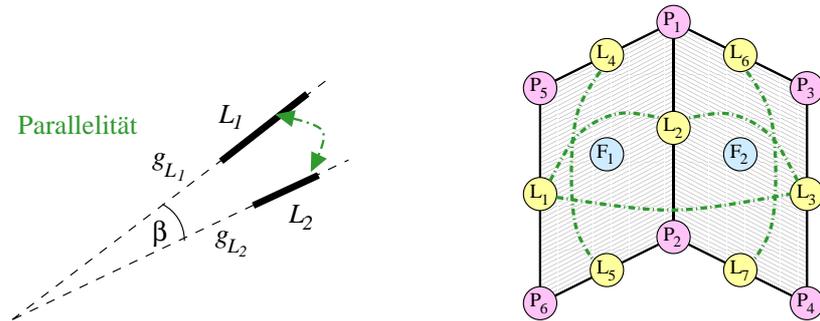
### 3.3.1.2 Parallelität zweier Linien

Bei der Linienparallelität handelt es sich um einen binären geometrischen Constraint, der wie folgt definiert ist:

#### Definition 3.15 (Parallelitäts-Constraint)

Zwei Liniensegmente  $L_1$  und  $L_2$  werden als parallel angesehen, wenn die durch sie verlaufenden Geraden  $g_{L_1}$  und  $g_{L_2}$  echt parallel sind, oder anderenfalls ihr kleinerer Schnittwinkel  $\beta$  kleiner oder gleich einem definierten Schwellwert  $\varepsilon_{\text{parallel}}$  ist.  $\square$

Da nicht zu erwarten ist, daß zwei extrahierte Bildmerkmale exakt parallel zueinander sind, werden sie als parallel angesehen, wenn die absolute Differenz der Linienwinkel von  $L_1$  und  $L_2$  bzgl. der horizontalen Bildachse kleiner dem Schwellwert  $\varepsilon_{\text{parallel}}$  ist (siehe Abbildung 3.3 links). Dieser Schwellwert relativiert dabei die folgenden drei Phänomene:



**Abbildung 3.3:** Links: Zwei Linienmerkmale  $L_1$  und  $L_2$  werden als parallel angesehen, wenn der Schnittwinkel  $\beta$  der durch sie verlaufenden Geraden  $g_{L_1}$  und  $g_{L_2}$  kleiner oder gleich einem Schwellwert  $\varepsilon_{\text{parallel}}$  ist. Rechts: Linienparallelitäts-Constraints einer Ansicht eines Satteldachs. Linienmerkmale sind durch gelbe Knoten und die Constraints als grüne Kanten dargestellt.

1. Die extrahierten Merkmale besitzen eine inhärente Unschärfe, die bei der Digitalisierung durch die Quantisierung des Bildes in Bildpunkte entsteht. Diese Unschärfe hängt im wesentlichen von der verwendeten Auflösung ab.
2. Die Lokalisierung der Linienmerkmale erfolgt bei der Merkmalsextraktion durch die Schätzung der wahrscheinlichsten Position, die jedoch mit einer bestimmten Unsicherheit behaftet ist (vgl. [Fuc97]).
3. Linien, die im 3D-Objektraum parallel sind, werden bei der durch die Kamera realisierten Zentralprojektion perspektivisch verzerrt, und laufen — sofern sie nicht parallel zur Bildebene verlaufen — in Richtung des Fluchtpunktes zusammen. Aufgrund des großen Abstands der Gebäude zur Projektionsebene (in der Kamera) und der gleichzeitig verhältnismäßig kleinen Objektgrößen, wird bei der Generierung von Constraints näherungsweise eine Parallelprojektion angenommen.

Wegen der noch freien Form- und Orientierungsparameter der 3D-Gebäudehypothesen weisen auch die 2D-Gebäudeansichten noch Freiheitsgrade auf. So kann z.B. die Winkelabweichung der Wandlinien, die im 3D-Objektmodell vertikal und parallel sind, je nach Lage des Gebäudes im Bild stark variieren, denn in der Nähe des Nadirpunktes ist die perspektivische Verzerrung größer als an den Bildrändern. Da die Lage des Gebäudes im Bild oftmals nicht genau feststeht, können die jeweiligen Abweichungen nicht präzisiert werden. Das gleiche gilt für die von der Linienlänge abhängige Unsicherheit des Linienwinkels. Bei kurzen Linien wirkt sich die aus der Digitalisierung entstehende Unschärfe stärker aus als bei längeren Linien (siehe [Fuc97]). Deshalb sollte der Schwellwert für die Linienparallelität mit wachsender Linienlänge kleiner werden. Das Problem besteht jedoch darin, daß aufgrund der freien Gebäudeparameter oftmals die Längen der Linienmerkmale der Gebäudeansichten zum Zeitpunkt der Suche noch nicht feststehen, und somit auch kein längenabhängiger Schwellwert gewählt werden kann. Aus diesem Grund wird ein einziger Schwellwert definiert, der die maximale Winkelabweichung zweier Linienmerkmale unabhängig von ihrer Länge und den Effekten der projektiven Verzerrung im Bild festlegt.

Die Generierung der Parallelitäts-Constraints für eine Gebäudeansicht erfolgt anhand der bekannten Linienparallelitäten der 3D-Gebäudehypothesen. Da bei einer angenommenen Parallelprojektion die Linien, die im Objektraum parallel sind, auch im Bildraum parallel bleiben, werden für alle Linienparallelitäten der 3D-Gebäudehypothese, deren Linien in der Gebäudeansicht sichtbar sind, entsprechende Constraints erzeugt. Während es sich bei der Parallelitätsrelation um eine Äquivalenzrelation handelt, gilt dies nicht für den entsprechenden Constraint, weil dieser wegen des verwendeten Schwellwertes nicht mehr transitiv ist. Aus diesem Grund werden für die Linien der

Gebäudeansichten alle Relationen der transitiven Hülle der sichtbaren und zueinander parallelen 3D-Linien explizit generiert. Wenn beispielsweise drei Linien  $L_1$ ,  $L_2$  und  $L_3$  zueinander parallel sind, würde es bzgl. des 3D-Gebäudemodells normalerweise reichen, dies durch die zwei Constraints  $L_1 \parallel L_2$  und  $L_2 \parallel L_3$  auszudrücken. Bzgl. der 2D-Gebäudeansicht reicht dies jedoch nicht, da im Bild die Winkelabweichungen der beiden Constraints zwar jeweils kleiner dem Schwellwert sein könnten, die Summe der beiden Winkelabweichungen jedoch größer als  $\varepsilon_{\text{parallel}}$ , so daß  $L_1$  nicht mehr parallel zu  $L_3$  wäre. Deshalb muß für die Bilddaten ebenfalls der Constraint  $L_1 \parallel L_3$  überprüft werden.

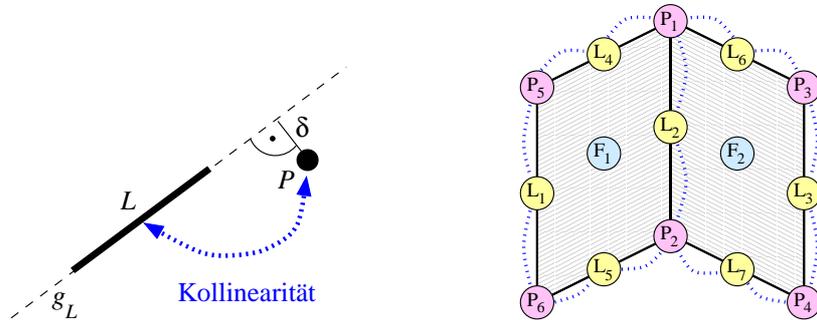
Die Sichtbarkeit von Linienmerkmalen wird durch die Gebäudeansicht definiert, und ist deshalb in bezug auf die möglichen Ausprägungen der Ansicht konstant. Um die Invarianz des Linienparallelitäts-Constraints zu zeigen, muß somit nur gewährleistet werden, daß jeder generierte Constraint invariant bzgl. aller möglichen geometrischen Ausprägungen der Gebäudeansicht ist. Dazu wird der Schwellwert  $\varepsilon_{\text{parallel}}$  so groß gewählt, daß zwei Linien, die im Objektraum parallel sind, in jedem Fall auch im Bild als parallel erkannt werden. Wie ein solcher Schwellwert bestimmt werden kann, wird unten in Abschnitt 3.3.2 erläutert.

### 3.3.1.3 Kollinearität eines Punkts mit einer Linie

Der Kollinearitäts-Constraint ist ein binärer geometrischer Constraint, der wie folgt definiert ist:

#### Definition 3.16 (Kollinearitäts-Constraint)

Ein Liniensegment  $L$  und ein Punktmerkmal  $P$  gelten als kollinear, wenn der Abstand  $\delta$  von  $P$  zu der durch  $L$  verlaufenden Geraden  $g_L$  kleiner oder gleich einem definierten Schwellwert  $\varepsilon_{\text{collin}}$  ist.  $\square$



**Abbildung 3.4:** Links: Ein Liniensegment  $L$  und ein Punkt  $P$  werden als kollinear angesehen, wenn der Abstand  $\delta$  von  $P$  zu der durch  $L$  verlaufenden Geraden  $g_L$  kleiner einem bestimmten Schwellwert ist. Rechts: Alle Kollinearitäts-Constraints einer Ansicht eines Satteldachs.

Linienmerkmale werden bei der Segmentierung häufig zu kurz oder fragmentiert extrahiert. Aus diesem Grund ist die erwartete Merkmalsnachbarschaft zwischen einem Eckpunkt und einer laut Modell inzidenten Linie oftmals nicht gegeben. Die entsprechenden Linienfragmente sind allerdings in Richtung der Eckpunkte ausgerichtet, und deshalb immer kollinear mit diesen. Da nicht zu erwarten ist, daß ein Eckpunkt exakt kollinear mit einer Linie ist, werden ein Punkt- und ein Linienmerkmal als kollinear angesehen, wenn der Abstand des Punktes von der durch das Linienmerkmal verlaufenden Geraden kleiner als ein festgelegter Schwellwert  $\varepsilon_{\text{collin}}$  ist. Dies wird in Abbildung 3.4 in dem linken Bild verdeutlicht.

Der Kollinearitäts-Constraint ist weder reflexiv, symmetrisch noch transitiv, weil er zwei Objekte unterschiedlicher Klassen (ein Punkt und eine Linie) zueinander in Relation setzt. Die Generierung der Kollinearitäts-Constraints für eine Gebäudeansicht erfolgt genau wie bei der Merkmals-

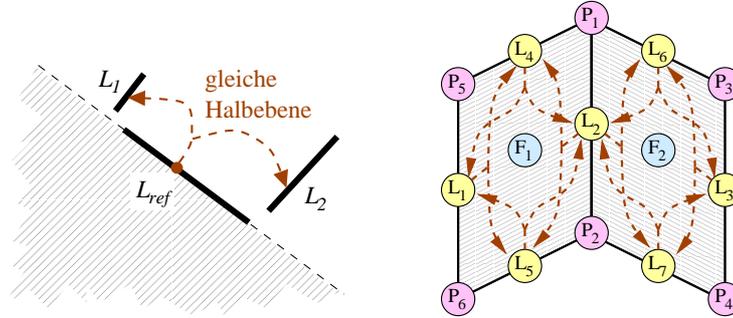
nachbarschaft anhand der Inzidenzbeziehungen zwischen Linie und Punkt der Randflächendarstellung des Modells (siehe Abbildung 3.4 rechts). Der Constraint ist demnach — genau wie der Merkmalsnachbarschafts-Constraint — invariant bzgl. der möglichen geometrischen Ausprägungen der Gebäudeansicht (vgl. Abschnitt 3.3.1.1).

### 3.3.1.4 Zwei Liniensegmente auf der gleichen Seite eines dritten

Die `Same_Side_Lines`-Relation ist eine ternäre geometrische Relation, und ist wie folgt definiert:

#### Definition 3.17 (Same\_Side\_Lines-Constraint)

Drei Liniensegmente  $L_{ref}$ ,  $L_1$  und  $L_2$  erfüllen den `Same_Side_Lines`-Constraint, wenn  $L_1$  und  $L_2$  vollständig in derselben Halbebene liegen, welche von der durch die Referenzlinie  $L_{ref}$  verlaufenden Geraden  $g_{L_{ref}}$  aufgespannt wird.  $\square$



**Abbildung 3.5:** Links: Drei Liniensegmente  $L_1$ ,  $L_2$  und  $L_{ref}$  erfüllen die `Same_Side_Lines`-Relation, wenn sich  $L_1$  und  $L_2$  in derselben Halbebene befinden, welche von der durch  $L_{ref}$  verlaufenden Geraden  $g_{L_{ref}}$  gebildet wird. Rechts: Alle `Same_Side_Lines`-Constraints einer Ansicht eines Satteldachs.

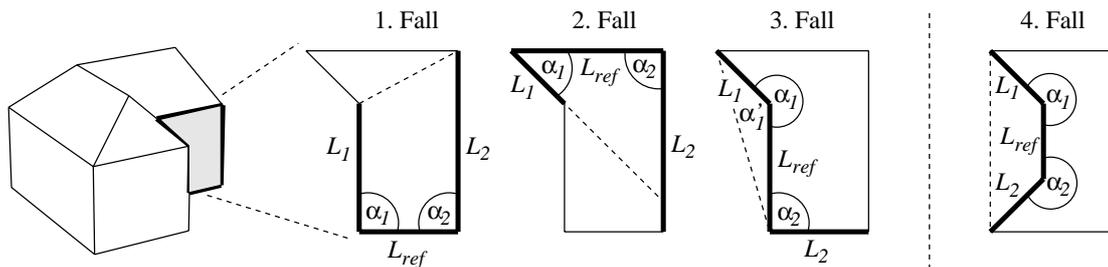
Es läßt sich beobachten, daß die meisten Flächen der Gebäudeansichten konvex sind. Der `Same_Side_Lines`-Constraint (kurz: `SSL`-Constraint) dient zunächst der Überprüfung der Konvexität von jeweils drei aufeinanderfolgenden Linienmerkmalen eines Flächenumringpolygons. Da die Orientierung der Linienmerkmale im Bild unbekannt ist, ist es nicht möglich, nur einen binären Constraint zu definieren, der besagt, daß eine Linie auf der rechten oder linken Seite einer anderen liegen muß. Die Aussagen „rechts von“ und „links von“ sind von der Durchlaufrichtung einer Linie eines Flächenumringpolygons abhängig. Diese ist zwar für die Linienmerkmale der Ansichten konsistent definierbar, jedoch nicht für die extrahierten Liniensegmente, weil für die Linien im Bild nicht a priori bekannt ist, ob sie zu einer Fläche gehören, und falls ja, zu welcher. Deshalb wird durch den `SSL`-Constraint mittels einer Referenzlinie die Bildebene in zwei Halbebenen geteilt, und durch ein zweites Liniensegment die Halbebene bestimmt, in der sich auch das dritte Liniensegment befinden muß (siehe Abbildung 3.5 links). Der `SSL`-Constraint ist aus diesem Grund auch für die drei Linienmerkmale zweier nebeneinanderliegender konkaver Ecken erfüllt. Dies muß bei der Generierung der `SSL`-Constraints entsprechend berücksichtigt werden. Der `SSL`-Constraint ist irreflexiv und symmetrisch bzgl. der Vertauschung von  $L_1$  und  $L_2$ .

Die `SSL`-Constraints werden aus der Inzidenzbeziehung von Linie und Fläche sowie Linie und Punkt der Randflächendarstellung der Gebäudeansicht hergeleitet. Es wird für jede zu einer Fläche inzidenten Linie  $L_{ref}$  betrachtet, ob die an den beiden Enden anschließenden Linien desselben Flächenumringpolygons auf derselben Seite von  $L_{ref}$  liegen. Um dies festzustellen, reicht es zu überprüfen, ob die Innenwinkel der beiden Ecken, die durch die drei Linien gebildet werden, entweder beide kleiner als  $180^\circ$  oder beide größer als  $180^\circ$  sind. Es wird für eine Linie  $L_{ref}$  somit kein `SSL`-Constraint generiert, wenn eine der anschließenden Linien  $L_1$  oder  $L_2$  kollinear zu  $L_{ref}$

ist (Innenwinkel gleich  $180^\circ$ ), oder einer der Innenwinkel größer als  $180^\circ$  und der andere kleiner als  $180^\circ$  ist. Da jede Modelllinie zu maximal zwei Flächen inzident sein kann, kann sie auch nur höchstens in zwei SSL-Constraints als  $L_{ref}$  auftreten.

Für konvexe Flächen ist der Beweis, daß die SSL-Constraints invariant bzgl. der verschiedenen geometrischen Ausprägungen einer Gebäudeansicht sind, recht einfach. In [HS93a] zeigen Haralick und Shapiro den Beweis des Satzes, daß die Abbildungen einer konvexen 3D-Fläche bei Parallel- oder Zentralprojektionen konvex bleiben. Für alle sichtbaren konvexen 3D-Flächen einer 3D-Gebäudehypothese ist somit klar, daß die entsprechenden Flächen der 2D-Gebäudeansichten konvex sind. Nun sind bei konvexen Polygonen alle Innenwinkel jeweils kleiner als  $180^\circ$ . Das bedeutet, daß für drei aufeinanderfolgende Linien eines konvexen Flächenumringpolygons die erste und die dritte Linie immer auf derselben Seite der durch die mittlere Linie verlaufenden Gerade liegen, und damit die Same\_Side\_Lines-Relation unabhängig von den Freiheitsgraden der Gebäudeansicht immer erfüllt ist.

Da der obige Satz nur für konvexe Flächen gilt, müssen konkave Flächen differenziert betrachtet werden. Beschränkte, d.h. geschlossene und endliche konkave Flächen besitzen Ecken, deren Innenwinkel kleiner als  $180^\circ$  ist. Eine Same\_Side\_Lines-Relation für drei aufeinanderfolgende Linien besteht nur dann, wenn die eingeschlossenen Innenwinkel entweder beide kleiner als  $180^\circ$  oder beide größer als  $180^\circ$  sind. Da die drei Linien als offenes Polygon interpretiert werden können, wird dieses nun so geschlossen, daß sich entweder ein (konvexes) Dreieck oder ein konvexes Viereck ergibt. Diese Konvexität bleibt laut obigem Satz in allen Projektionen erhalten, so daß daraus auf die Invarianz der Same\_Side\_Lines-Relation für die beiden Fälle geschlossen werden kann.



**Abbildung 3.6:** Verschiedene Situationen bei der Generierung von Same\_Side\_Lines-Constraints für konkave Flächen. Die ersten drei Fälle stammen aus der links gezeigten Gebäudeansicht. Der vierte Fall zeigt die Situation, in der zwei nebeneinanderliegende Ecken einer Fläche konkav sind. Bis auf den dritten Fall wird jeweils ein SSL-Constraint für  $L_{ref}$ ,  $L_1$  und  $L_2$  erzeugt.

Abbildung 3.6 illustriert die vier verschiedenen Fälle, die bei einer konvexen Fläche einer Gebäudehypothese im Objektraum auftreten können:

1. Beide Innenwinkel  $\alpha_1$  und  $\alpha_2$  sind kleiner als  $180^\circ$ , und  $\alpha_1 + \alpha_2$  ist größer oder gleich  $180^\circ$ . In diesem Fall werden die freien Enden von  $L_1$  und  $L_2$  durch eine Linie miteinander verbunden, wodurch ein Viereck entsteht. Da die Summe der Innenwinkel eines Vierecks  $360^\circ$  beträgt, und bereits  $\alpha_1 + \alpha_2 \geq 180^\circ$  ist, kann keine konkave Ecke mehr entstehen. Das Viereck ist somit konvex und bleibt es auch bei der Projektion in den Bildraum. Damit ist sichergestellt, daß der SSL-Constraint in allen geometrischen Ausprägungen der Gebäudeansicht erfüllt bleibt.
2. Beide Innenwinkel  $\alpha_1$  und  $\alpha_2$  sind kleiner als  $180^\circ$ , und  $\alpha_1 + \alpha_2$  ist kleiner als  $180^\circ$ . In diesem Fall werden die Linien  $L_1$  und  $L_2$  an ihren offenen Enden solange verlängert, bis sie sich treffen und ein (konvexes) Dreieck bilden. Wie im ersten Fall bleibt die Konvexität bei der Projektion und damit auch die Erfüllung des SSL-Constraints für alle geometrischen Ausprägungen der Gebäudeansicht erhalten.

3. Innenwinkel  $\alpha_1$  ist größer als  $180^\circ$  und  $\alpha_2$  ist kleiner als  $180^\circ$ ; beide Linien liegen auf unterschiedlichen Seiten von  $L_{ref}$ , und es wird kein SSL-Constraint generiert. Es muß hier gezeigt werden, daß dieser Zustand für alle möglichen geometrischen Ausprägungen der Gebäudeansicht erhalten bleibt. Dazu wird das offene Ende von  $L_1$  durch eine Linie mit dem Eckpunkt von  $L_{ref}$  und  $L_2$  verbunden, wodurch aufgrund von  $\alpha_1 > 180^\circ$  und damit  $\alpha'_1 < 180^\circ$  ein Dreieck entsteht. Da das Dreieck bei allen Projektionen erhalten bleibt, bleibt auch die Bedingung  $\alpha'_1 < 180^\circ$  erhalten. Das bedeutet, daß  $\alpha_1$  in allen Projektionen der Gebäudehypothese größer als  $180^\circ$  bleibt. Analog läßt sich zeigen, daß  $\alpha_2$  immer kleiner als  $180^\circ$  bleibt, und somit die SSL-Relation in keiner Projektion erfüllt wird.
4. Beide Innenwinkel  $\alpha_1$  und  $\alpha_2$  sind größer als  $180^\circ$ . Je nach Winkelsumme von  $\alpha_1$  und  $\alpha_2$  entspricht diese Situation dem ersten oder dem zweiten Fall. Der Unterschied besteht darin, daß das konvexe Viereck bzw. Dreieck nicht innerhalb der ursprünglichen Fläche gebildet wird, sondern auf der anderen Seite von  $L_{ref}$  (siehe Abbildung 3.6 ganz rechts), und damit sichergestellt wird, daß beide Ecken konkav bleiben und der SSL-Constraint in allen Projektionen erfüllt ist.

### 3.3.1.5 Verschiedenheitsrelation

Die Verschiedenheitsrelation ist eine binäre, symmetrische und irreflexive Relation, die wie folgt definiert ist:

**Definition 3.18 (Verschiedenheits-Constraint)**

Zwei Modellmerkmale  $M_1$  und  $M_2$  sind verschieden, wenn ihnen nicht dasselbe Bildmerkmal zugeordnet wird. □

Die Verschiedenheitsrelation sichert die Injektivität der Abbildungsfunktion  $h$ , so daß allen Modellmerkmalen paarweise verschiedene Bildmerkmale zugeordnet werden (vgl. Abschnitt 3.1.1). Die Verschiedenheits-Constraints auf den Merkmalen einer Gebäudeansicht müssen dafür sorgen, daß alle Variablen, die die Modellmerkmale repräsentieren, paarweise verschiedene Werte beinhalten. CLP(FD) bietet dazu den eingebauten Constraint `alldistinct` an, der erzwingt, daß die ihm in einer Liste angegebenen Variablen paarweise verschiedene Werte besitzen (siehe [DvHS<sup>+</sup>88, vH89]). Durch die Verwendung des `alldistinct`-Constraints läßt sich vermeiden, daß zwischen den  $n$  Variablen  $n(n-1)/2$  einzelne Verschiedenheits-Constraints generiert werden müßten.

Da den Modellmerkmalen generell paarweise verschiedene Bildmerkmale zugeordnet werden müssen, ist diese Relation unabhängig von (und damit invariant bzgl.) der konkreten Ausprägung einer Gebäudeansicht.

### 3.3.1.6 Bekannter Linienwinkel von Modelllinien

Der Linienwinkel-Constraint ist ein unärer geometrischer Constraint, der wie folgt definiert ist:

**Definition 3.19 (Linienwinkel-Constraint)**

Ein Liniensegment  $L$  erfüllt den Linienwinkel-Constraint bzgl. eines vorgegebenen Winkels  $\gamma$ , wenn die absolute Differenz des Linienwinkels von  $L$  bzgl. der horizontalen Bildachse und  $\gamma$  kleiner oder gleich einem bestimmten Schwellwert  $\varepsilon_{\text{line\_angle}}$  ist. □

Dieser Constraint dient dazu, für Linienmerkmale einer Gebäudeansicht nur die Zuweisung von extrahierten Liniensegmenten zu erlauben, die einen bestimmten Linienwinkel besitzen. Für welche Modelllinien der Winkel bekannt ist, hängt von den rekonstruierten 3D-Ecken ab, die bei der Generierung der Gebäudehypothesen berücksichtigt wurden. Das bedeutet, daß nur für die Modelllinien Linienwinkel-Constraints generiert werden, deren Orientierungen in der Gebäudeansicht feststehen und damit auch invariant sind.

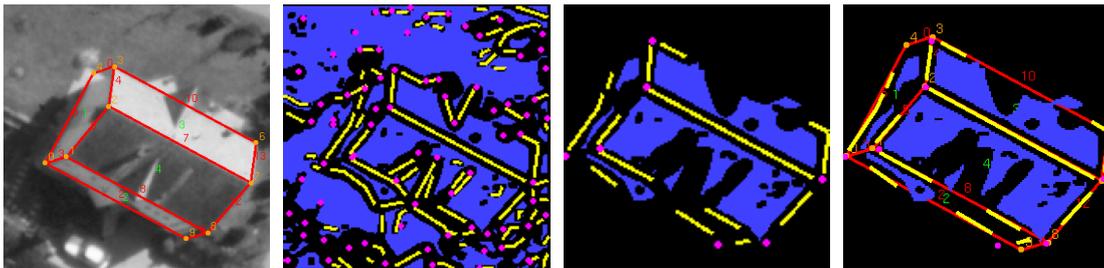
### 3.3.2 Bestimmung der Schwellwerte

Die Erfüllung der drei geometrischen Constraints Linienparallelität, Kollinearität und Linienwinkel wird jeweils über einen Schwellwert abgeprüft. Diese Schwellwerte relativieren verschiedene Effekte der Unschärfe und Unsicherheit der Luftbildaufnahme, Digitalisierung und Merkmalsextraktion. Darüberhinaus relativiert der Schwellwert des Linienparallelitäts-Constraints auch die perspektivische Verzerrung der durch die Kamera realisierten Zentralprojektion, so daß für die Generierung und Überprüfung der Linienparallelitäts-Constraints vereinfachend eine Parallelprojektion angenommen werden kann.

Ferner egalisieren die Schwellwerte kleinere Abweichungen der Daten vom Modell. Wenn z.B. bei einem Satteldachhaus das Dach ein wenig übersteht, sind die vertikalen Kanten der Stirnseiten des Hauses nicht mehr kollinear mit den Dachecken. Dies kann durch einen entsprechend großen Schwellwert des Kollinearitäts-Constraints berücksichtigt werden.

Es stellt sich nun die Frage nach den Größen dieser Schwellwerte, und danach, wie sie bestimmt werden können. Je größer ein Schwellwert gewählt wird, desto höher wird die Toleranz des entsprechenden Constraints gegenüber Abweichungen. Andererseits sinkt mit wachsendem Schwellwert die Spezifität des Constraints, weil immer mehr Merkmale als zueinander kompatibel angesehen werden (vgl. [Rad93]). Konkret bedeutet dies, daß ein zu kleiner Schwellwert dazu führen kann, daß ein Modell im Bild nicht erkannt wird, weil einer der Constraints des Modells nicht erfüllbar ist. Ein zu großer Schwellwert kann hingegen zu Fehlzuordnungen führen, weil die Trennschärfe des Modells nicht mehr ausreichend ist.

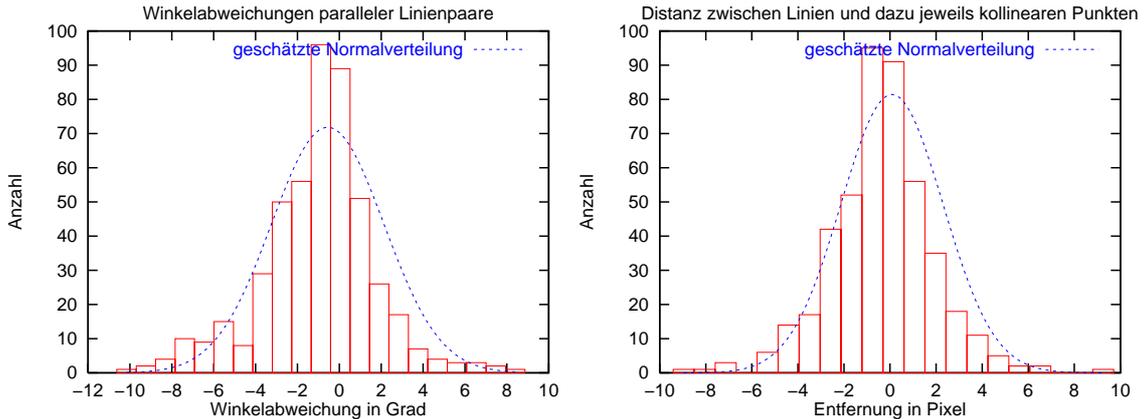
Da eine umfassende Modellierung der oben aufgezählten Effekte der Unschärfe, perspektivischen Verzerrung und tolerierbaren Modellabweichungen nicht mit vertretbarem Aufwand realisierbar ist, werden die Schwellwerte anhand von Trainingsdaten bestimmt. Dazu werden zu den abgebildeten Häusern verschiedener Luftbildausschnitte die entsprechenden Gebäudehypothesen generiert, und die jeweils zutreffende Gebäudeansicht berechnet. Diese Gebäudeansichten werden manuell in das Bild eingepaßt, so daß das Modell bestmöglich mit den Bilddaten übereinstimmt. Anschließend wird eine Merkmalsextraktion auf den Bildern durchgeführt, und den Merkmalen der Gebäudeansichten die korrespondierenden extrahierten Bildmerkmale (ebenfalls manuell) zugeordnet. Abbildung 3.7 demonstriert diese Teilschritte an einem Beispiel.



**Abbildung 3.7:** Bestimmung der Schwellwerte anhand von Trainingsdaten. Von links nach rechts: Luftbildausschnitt mit manuell eingepaßter Gebäudeansicht, extrahierte Bildmerkmale, manuell zugeordnete Bildmerkmale, Überlagerung von idealem Modell und bester manueller Zuordnung. Die drei Schwellwerte für Linienparallelität, Linienwinkel und Kollinearität ergeben sich aus den diesbzgl. maximalen Differenzen zwischen Modellmerkmalswerten und zugeordneten Bildmerkmalswerten über alle Trainingsbilder.

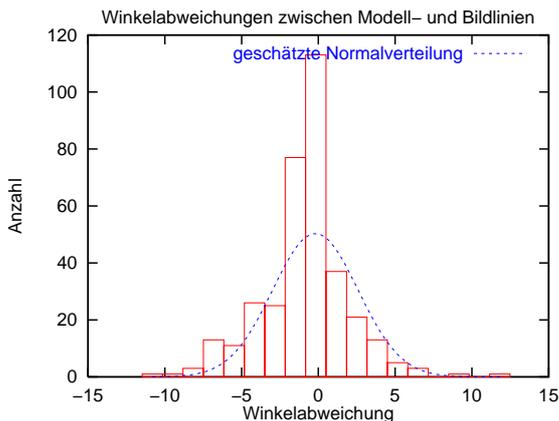
Nach der Zuordnung der korrespondierenden Bild- zu den Modellmerkmalen, wird für jeden Linienparallelitäts-Constraint des Modells die Winkeldifferenz der beiden zugeordneten Bildmerkmale bestimmt. Das gleiche wird für jeden Kollinearitäts-Constraint durchgeführt, mit dem Unterschied, daß hier der Abstand des Punktmerkmals zu der durch das entsprechende Linienmerkmal verlaufenden Geraden bestimmt wird. Abbildung 3.8 zeigt die Häufigkeitsverteilungen der Differenzen

für die Linienparallelitäts- und Kollinearitäts-Constraints, die anhand von 28 manuell zugeordneten Gebäudeansichten unterschiedlicher Art (bzgl. des Dachtyps und der Grundform) bestimmt wurden. Anhand der Differenzen wurde jeweils eine Normalverteilung geschätzt, und in die entsprechende Häufigkeitsverteilung eingetragen. Es läßt sich erkennen, daß die Häufigkeitsverteilungen zwar nicht normalverteilt sind, jedoch in ihrer Ausprägung einer Normalverteilung entsprechen<sup>4</sup>. Der Grund, warum in dem rechten Diagramm sowohl positive als auch negative Distanzen auftreten, liegt darin begründet, daß bei den Kollinearitäts-Constraints die Punkte auf verschiedenen Seiten der durch die Linien verlaufenden Geraden liegen können.



**Abbildung 3.8:** Häufigkeitsverteilungen der Winkelabweichungen paralleler Linien (links) und Distanzen von Linien zu kollinearen Punkten (rechts), jeweils überlagert durch eine geschätzte Normalverteilung.

Die Winkeldifferenzen der Linienwinkel-Constraints werden analog zu den Linienparallelitäts- und Kollinearitäts-Constraints bestimmt. Dabei werden die Differenzen zwischen den Linienwinkeln der Modelllinien und der ihnen zugeordneten extrahierten Liniensegmente ermittelt. Abbildung 3.9 zeigt die Häufigkeitsverteilung der Differenzen, überlagert mit einer geschätzten Normalverteilung.



**Abbildung 3.9:** Häufigkeitsverteilung der Abweichungen von Linienwinkeln des Modells zu den Winkeln der jeweils zugeordneten extrahierten Linienmerkmale, überlagert durch eine geschätzte Normalverteilung.

Zur Erkennung der Gebäudeansichten im Bild müssen alle Constraints erfüllt werden. Aus diesem Grund müssen die Schwellwerte mindestens so groß gewählt werden, daß alle in den Trainingsdaten enthaltenen Zuordnungen die Constraints der entsprechenden Gebäudeansichten erfüllen. Wenn

<sup>4</sup>Zur Überprüfung wurden  $\chi^2$ -Tests durchgeführt (siehe [Pap97]), die ergeben haben, daß die ermittelten Häufigkeitsverteilungen nicht normalverteilt sind. Der Grund dafür liegt in den Ausreißern bei den zugeordneten Bildmerkmalen.

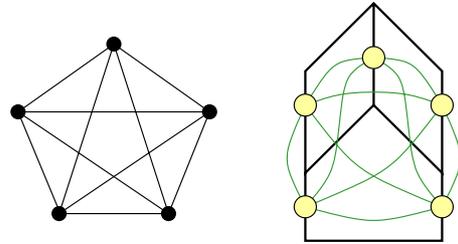
man davon ausgeht, daß eine ausreichend große und repräsentative Anzahl von Trainingszuordnungen bestimmt wurde, können die drei Schwellwerte auf den jeweils größten Differenzbetrag für die drei Constraint-Typen festgesetzt werden. Bei den verwendeten Trainingsdaten, die anhand von Luftbildausschnitten mit einer Bodenauflösung von 15 cm bestimmt wurden, betragen die Schwellwerte für den Linienparallelitäts-Constraint  $9,7^\circ$ , für den Kollinearitäts-Constraint 6,74 Pixel und für den Linienwinkel-Constraint  $8,37^\circ$ . Da sowohl die Unschärfe der extrahierten Bildmerkmale als auch die Abweichungen der im Bild befindlichen Häuser bzgl. der Gebäudemodelle von der Auflösung abhängen, gelten die Schwellwerte im allgemeinen nur für Szenen mit der gleichen Bildauflösung; es müssen demnach für verschiedene Bildauflösungen jeweils eigene Schwellwerte trainiert werden.

### 3.3.3 Eigenschaften der Constraint-Graphen

Sowohl die extrahierten Bildmerkmale als auch die Modelle (Gebäudeansichten) werden mittels der vorgestellten Constraints als relationale Strukturen repräsentiert. Zur Identifikation einer Gebäudeansicht in einem Luftbild wird nach einem Subgraph-Isomorphismus des Modells zu den extrahierten Bilddaten gesucht.

In Abschnitt 3.1.1 wurde erläutert, daß das Problem der Bestimmung eines Subgraph-Isomorphismus im allgemeinen NP-vollständig ist. Für planare Graphen gibt es jedoch spezielle Algorithmen, die die Bestimmung von Subgraph-Isomorphismen effizient in linearer Zeit ( $O(n)$ ) berechnen können (siehe [Epp95]). Es gilt deshalb festzustellen, ob die relationalen Modell- und Bildstrukturen diese Eigenschaft aufweisen, und damit das aufwendige Zuordnungsproblem entscheidend vereinfacht werden könnte, was den Einsatz von Constraint-Techniken überflüssig machen würde.

Im folgenden wird gezeigt, daß weder die Graphstruktur der extrahierten Bildmerkmale noch die der Gebäudeansichten plättbar sind. Nach dem Satz von Kuratowski ist ein Graph nicht plättbar, wenn er einen zu  $K_{3,3}$  oder  $K_5$  isomorphen Subgraphen enthält (siehe [RS94, BSMM97])<sup>5</sup>. Letzterer Fall tritt bereits auf, wenn ein Modell oder Bild mindestens fünf zueinander parallele Linien enthält, weil die Parallelitätsrelation transitiv ist und damit jede der fünf Linien mit den anderen vier in Relation steht. Abbildung 3.10 zeigt links einen  $K_5$ -Graph und rechts eine Gebäudeansicht, deren Linien eine zum  $K_5$  isomorphen Graph bilden. Sowohl bei den Daten als auch den Modellen handelt es sich im allgemeinen also nicht um plättbare Graphen; die speziellen Algorithmen zur Bestimmung von Subgraph-Isomorphismen können aus diesem Grund nicht angewendet werden.



**Abbildung 3.10:** Bereits fünf zueinander parallele Linien einer Gebäudeansicht bilden einen zum  $K_5$ -Graph (links) isomorphen Graphen.

<sup>5</sup>Der  $K_{3,3}$ -Graph ist ein bipartiter Graph mit drei Knoten in jeder der beiden Partitionen, wobei jeder Knoten der einen Partition mit allen anderen Knoten der anderen Partition verbunden ist. Der  $K_5$ -Graph ist ein Graph mit fünf Knoten, bei dem jeder Knoten mit allen anderen verbunden ist.

### 3.4 Realisierung des Konzepts in CLP

In diesem Abschnitt wird die praktische Realisierung des entwickelten Konzepts unter Verwendung der Möglichkeiten von Constraint Logic Programming über endlichen Wertebereichen, CLP(FD), vorgestellt. Es soll gezeigt werden, daß die Wissensrepräsentation von Gebäudeansichten durch Fakten, Regeln und Constraints in CLP unmittelbar als Programm zur Suche nach den entsprechenden Gebäudeansichten verwendet werden kann. Dazu wird im folgenden zunächst erläutert, wie die extrahierten Bildmerkmale durch Fakten extensional repräsentiert werden. Anschließend wird beispielhaft gezeigt, wie die Überprüfung eines Constraints durch eine Regel realisiert wird, und diese Regel durch die Benennung der zu verwendenden Inferenzregel als Constraint deklariert wird. Danach wird erläutert, wie die verschiedenen Constraints, die eine Gebäudeansicht repräsentieren, zu einer CLP-Klausel aggregiert werden, die das zur Erkennung der Ansicht zu lösende Constraint Satisfaction Problem vollständig beinhaltet. Abschließend wird an einem Beispiel mit realen Luftbilddaten demonstriert, daß die verwendeten Constraints eine ausreichende Ausdrucksstärke besitzen, um Gebäudeansichten in einem Luftbild zu identifizieren. Ferner werden die verschiedenen Auswirkungen von Vorwissen, Heuristiken und Inferenzregeln auf die benötigte Laufzeit des Suchvorgangs diskutiert.

#### 3.4.1 Fakten, Regeln und Constraints

Bei den extrahierten Bildmerkmalen handelt es sich um komplexe Objekte mit einer Vielzahl von Attributen. Sie werden als einfache Fakten mit atomaren Argumenten dargestellt:

```
point(id, x, y, typ)
line(id, x1, y1, x2, y2, winkel, länge, typ)
blob(id, bbox_x1, bbox_y1, bbox_x2, bbox_y2, streifen, fläche, umfang,
      formfaktor, lochanzahl, schwerpunkt_x, schwerpunkt_y,
      hauptachsenwinkel, grauwert, grauwertvarianz)
```

Punkte werden durch `point`-Fakten, Linien durch `line`-Fakten und Flächen durch `blob`-Fakten repräsentiert. Um die Bildmerkmale benennen und voneinander unterscheiden zu können, werden sie mit einem global eindeutigen Objektbezeichner (Attribut `id`) versehen. Auf die Bedeutung der anderen Attribute wird hier nicht weiter eingegangen; nähere Erläuterungen befinden sich in [Fuc97] und [KPC96]. Der von dem verwendeten Segmentierungsverfahren extrahierte Merkmals-nachbarschaftsgraph (vgl. Abschnitt 2.5.1) wird ebenfalls extensional durch Fakten der Form

```
adjacent(object-id1, object-id2)
```

repräsentiert. Zur Überprüfung, ob zwei Bildmerkmale benachbart sind, muß nur nachgeschaut werden, ob für die beiden Merkmale ein entsprechendes Fakt existiert. Die anderen Relationen werden durch Regeln in Form von Prolog-Prädikaten beschrieben und besitzen somit eine intensionale Repräsentation (vgl. [Rei91]). Beispielsweise erfolgt die Überprüfung, ob zwei Linien zueinander parallel sind, anhand der folgenden Regel<sup>6</sup>:

```
line_parallel(L1,L2) :-
    line(L1,_,_,_,_,Alpha,_,_),
    line(L2,_,_,_,_,Beta,_,_),
    Diff is abs(Alpha-Beta),
    Diff =< SCHWELLWERT.
```

<sup>6</sup>Die Überprüfung der Winkeldifferenz ist hier vereinfacht dargestellt. Der Fall  $\alpha = 179^\circ$ ,  $\beta = 1^\circ$  und  $\varepsilon_{\text{parallel}} = 3$ , in dem die Linien  $L_1$  und  $L_2$  als parallel angesehen werden müssen, wird von dieser vereinfachten Überprüfung nicht abgedeckt.

In den ersten beiden Zeilen des Regelrumpfes werden zunächst die Winkel der beiden durch L1 und L2 bezeichneten Linien bzgl. der horizontalen Bildachse ermittelt. Anschließend wird der Betrag der Winkeldifferenz berechnet und überprüft, ob dieser kleiner oder gleich dem Schwellwert ist, der aus den Trainingsdaten ermittelt wurde.

Der implementierte Constraint-Interpreter setzt auf dem CLP-System ECLIPSE auf (siehe [MS92, WNS97]), und ermöglicht die Deklaration von Regeln und Fakten als Constraints. Die Regeln werden dann nicht mehr mittels der Standard-Prolog-Auswertestrategie abgearbeitet, sondern anhand der in Abschnitt 3.2.3 erläuterten Konsistenztechniken ausgewertet. Die Deklaration eines Constraints erfolgt in Anlehnung an das CLP-System CHIP (vgl. [DvHS<sup>+</sup>88]), indem die zu verwendende Inferenzregel und die betroffenen Variablen benannt werden:

```
:- look_ahead adjacent(+,+).
:- forward_checking line_parallel(+,+).
```

Durch diese Deklaration wird dem System mitgeteilt, daß die Merkmalsnachbarschafts-Constraints mittels der Look-Ahead-Inferenzregel ausgewertet werden sollen, und beide Argumente Bereichsvariablen darstellen (durch das '+'-Zeichen markiert). Ferner sind die Linienparallelitäts-Constraints mittels der Forward-Checking-Inferenzregel auszuwerten, und auch hier beinhalten beide Argumente Bereichsvariablen. Der Vorteil dieses Schemas besteht darin, daß beliebige Prolog-Regeln mittels des Constraint-Interpreters ausgewertet werden können und damit eine einfache Erweiterung um weitere Constraints möglich ist. Die Bereitstellung der beiden Inferenzregeln Look-Ahead und Forward-Checking ermöglicht darüberhinaus eine differenzierte Auswertestrategie für verschiedene Constraint-Typen. Wie bei der praktischen Evaluation noch gezeigt wird, kann dies einen erheblichen Einfluß auf die benötigte Laufzeit zur Identifikation einer Gebäudeansicht im Bild haben.

#### 3.4.1.1 Caching von Wertebereichsreduktionen

Die Auswertung der geometrischen Relationen erfordert z.T. verhältnismäßig aufwendige numerische Berechnungen. Für jeden Konsistenztest, der feststellen soll, ob zwei bzw. drei Merkmale eine Relation erfüllen, müssen diese Berechnungen durchgeführt werden. Dabei kann es aufgrund des Backtrackings dazu kommen, daß derselbe Konsistenztest für eine bestimmte Merkmalskonfiguration während des Suchvorgangs mehrfach wiederholt werden muß.

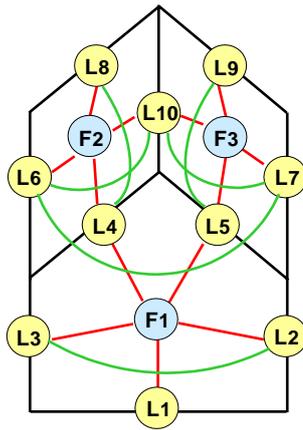
Je nach verwendeter Inferenzregel wird ein Constraint geweckt, sobald entweder alle bis auf eine Variable einen eindeutigen Wert besitzen (Forward-Checking) oder der Wertebereich einer seiner Variablen verkleinert wurde (Look-Ahead). Nach dem Wecken wird überprüft, ob anhand des Constraints Reduktionen der Wertebereiche der anderen Variablen durchgeführt werden können. Dazu werden für alle noch freien Variablen die Werte herausgefiltert und aus dem Wertebereich eliminiert, die in keiner der noch möglichen Variablenbelegung enthalten sind, die den Constraint erfüllen können.

Diese Reduktionen werden nun im Verlauf der Suche in einem Cache registriert. Dabei werden die kompletten Wertebereiche der Variablen vor und nach der Konsistenzüberprüfung gespeichert. Wird nun ein Constraint geweckt, so wird vor der Durchführung der Konsistenzüberprüfung zuerst im Cache nachgeschaut, ob diese Situation schon einmal aufgetreten ist, und die entsprechenden Reduktionen der Wertebereiche bereits bekannt sind. Wird ein passender Cache-Eintrag gefunden, kann auf die Konsistenzüberprüfung für die einzelnen Bildmerkmale verzichtet werden, und die im Cache gespeicherten reduzierten Wertebereiche ersetzen unmittelbar die bisherigen Wertebereiche der Variablen. Nähere Details der technischen Umsetzung sind in [KPC96] beschrieben.

#### 3.4.2 Transformation von Gebäudeansichten in CLP-Klauseln

Zur Suche der Gebäudehypothesen im Bild wird für jede Gebäudeansicht eine CLP-Klausel erzeugt, die nur aus der Festlegung der Wertebereiche der Variablen und den verschiedenen geometrischen

und topologischen Constraints besteht. Dazu wird für jedes Modellmerkmal eine Bereichsvariable generiert, deren Wertebereich vom jeweiligen Merkmalstyp abhängt. Der Wertebereich für Linienvariablen wird aus der Menge der extrahierten Linienmerkmale gebildet, der Bereich für Punktvariablen aus den extrahierten Punktmerkmalen und der Wertebereich der Flächenvariablen aus den extrahierten Flächenmerkmalen. Die Constraints werden anhand der in den Abschnitten 3.3.1.1 bis 3.3.1.6 erläuterten Regeln aus der Randflächendarstellung einer Gebäudeansicht hergeleitet, und konjunktiv miteinander verknüpft. Abbildung 3.11 zeigt einen Auszug der Klausel, die für die Vorderansicht eines Satteldachhauses generiert wurde.



```

house_aspect(F1,F2,F3,L1,...,L10) :-
    extracted_blobs(BlobDom), [F1,F2,F3] ∈ BlobDom,
    extracted_lines(LineDom), [L1,...,L10] ∈ LineDom,
    all_distinct([F1,F2,F3]),
    all_distinct([L1,...,L10]),
    adjacent(F1,L1), adjacent(F1,L2), adjacent(F1,L3),
    adjacent(F1,L4), adjacent(F1,L5), adjacent(F2,L4),
    adjacent(F2,L6), adjacent(F2,L8), adjacent(F2,L10),
    adjacent(F3,L5), adjacent(F3,L7), adjacent(F3,L9),
    adjacent(F3,L10),
    line_parallel(L2,L3), line_parallel(L6,L7),
    line_parallel(L6,L10), line_parallel(L7,L10),
    line_parallel(L4,L8), line_parallel(L5,L9).

```

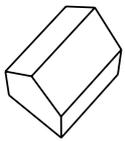
**Abbildung 3.11:** CLP-Klausel zur Repräsentation der Vorderansicht eines Satteldachhauses. Aus Gründen der Übersichtlichkeit werden nur einige der Relationen dargestellt und die Punktmerkmale weggelassen. Diese Klausel besagt, daß wenn die den Linienvariablen  $L_1, \dots, L_{10}$  und den Flächenmerkmalen  $F_1, F_2, F_3$  zugeordneten extrahierten Bildmerkmale alle im Klauselrumpf enthaltenen Constraints erfüllen, ihre Aggregation eine Instanz der betreffende Gebäudeansicht beschreibt.

Die bei der Generierung von Gebäudehypothesen berücksichtigten rekonstruierten 3D-Ecken beinhalten bereits Zuordnungen von Bild- zu Modellmerkmalen (vgl. Abschnitt 2.5.3). Dieses Vorwissen wird genutzt, um den Suchraum a priori einzuschränken, indem am Anfang der CLP-Klausel den betreffenden Modellvariablen die korrespondierenden Bildmerkmale direkt zugewiesen werden. Damit werden die Wertebereiche für diese Modellvariablen bereits auf einen eindeutigen Wert festgelegt. Sofern für die Luftbildszene ein digitales Oberflächenmodell (DOM) zur Verfügung steht, können darüberhinaus die Flächenvariablen, die Dachflächen repräsentieren, auf diejenigen extrahierten Flächenmerkmale eingeschränkt werden, die in dem Bildbereich liegen, für den eine signifikante Erhebung im DOM registriert wurde (vgl. Abbildung 2.9 auf Seite 25).

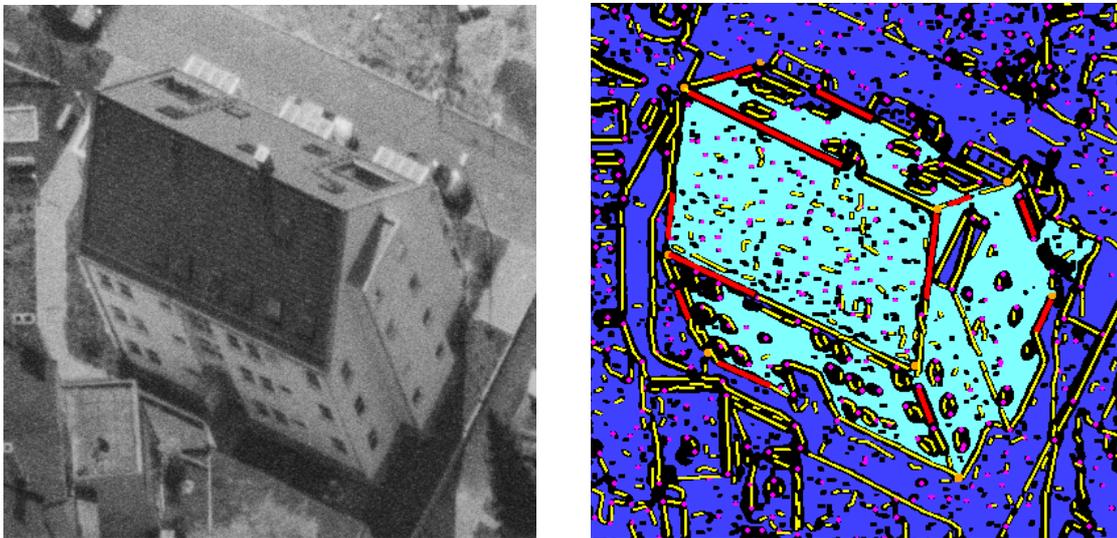
### 3.4.3 Praktische Evaluation

In diesem Abschnitt wird anhand eines praktischen Beispiels an einem realen Luftbildausschnitt gezeigt, daß mit der vorgestellten Constraint-Modellierung Gebäude korrekt und effizient identifiziert werden können. Es sollen ferner die Auswirkungen von Vorwissen durch rekonstruierte 3D-Ecken und mögliche Dachflächen auf die Dauer der Suche illustriert werden. Darüberhinaus soll gezeigt werden, welche Vorteile das in Abschnitt 3.4.1.1 beschriebene Caching sowie die differenzierte Verwendung der Forward-Checking- und Look-Ahead-Inferenzregeln für die verschiedenen Constraints erbringen. Schließlich soll noch der Unterschied der Verwendung der First-Fail-Heuristik, die eine dynamische Instantiierungsreihenfolge der Variablen realisiert, im Gegensatz zu einer festen Instantiierungsreihenfolge verdeutlicht werden.

Für das Beispiel werden die idealisierenden Annahmen gemacht, daß zu allen Modellmerkmalen entsprechende Bildmerkmale beobachtet werden können, und daß diese Bildmerkmale alle Constraints erfüllen. Der verwendete Luftbildausschnitt wurde dementsprechend ausgewählt (siehe Abbildung 3.12 links). Er zeigt ein Satteldachhaus in Schrägansicht, wobei von der Merkmalsextraktion alle wesentlichen Gebäudemerkmale extrahiert werden konnten. Es wurde die entsprechende Gebäudeansicht generiert und in eine Constraint-Klausel transformiert. Dabei wurden nur Merkmalsnachbarschafts-Constraints zwischen Flächen und Linien generiert, da diese relativ robust gegenüber Fragmentierungen sind. Die Rolle der Nachbarschafts-Constraints zwischen Punkten und Linien wird von den Kollinearitäts-Constraints übernommen (vgl. Abschnitt 3.3.1.3). Tabelle 3.1 zeigt links die Eigenschaften des Modells und rechts die Eigenschaften des Bildes.

Modellkomplexität			Bildkomplexität	
	Punkte	9	Punkte	360
	Linien	12	Linien	562
	Flächen	4	Flächen	93
	MNG (Fläche, Linie)	17	Nachbarschaften	2266
	Linienparallelität	11	Bildbreite [Pixel]	408
	Kollinearität	24	Bildhöhe [Pixel]	409
	Same_Side_Lines	17	Grauwerttiefe [Bit]	8
	Verschiedenheit	3		

**Tabelle 3.1:** Eigenschaften der relationalen Strukturen der zu identifizierenden Gebäudeansicht (links) und der extrahierten Bildmerkmale (rechts). Das Modell besteht aus insgesamt 25 Merkmalen und 72 Constraints. MNG(Fläche, Linie) steht für die Merkmalsnachbarschafts-Constraints zwischen Flächen und Linien. Für das Bild wurden 1015 Merkmale sowie 2266 Merkmalsnachbarschaften extrahiert.



**Abbildung 3.12:** Ergebnis der Suche nach der in Tabelle 3.1 gezeigten Gebäudeansicht in dem links dargestellten Luftbildausschnitt. Durch die erfolgreiche Zuordnung von Bild- zu Modellmerkmalen konnte die Gebäudeansicht korrekt identifiziert werden. Die den Modellmerkmalen zugeordneten Bildmerkmale sind farbig markiert, wobei zugeordnete Flächen hellblau, Linien rot und Punkte orange hervorgehoben sind.

Abbildung 3.12 zeigt rechts das Ergebnis der Suche, bei der allen Modellmerkmalen die richtigen extrahierten Bildmerkmale zugeordnet wurden. Man beachte, daß weder die Orientierung noch

die Lage sowie die Ausdehnung der Gebäudeansicht im Bild vorher bekannt war. Tabelle 3.2 zeigt die Ergebnisse von sechs verschiedenen Messungen, bei denen ein unterschiedliches Maß an Vorwissen einbezogen wurde (Messungen 1-3) bzw. verschiedene Implementierungsvarianten gewählt wurden (Messungen 4-6). Da die benötigte Laufzeit<sup>7</sup> zur Bestimmung einer Lösung vom jewei-

Suchvorgang	Laufzeit	Suchbaumknoten	Konsistenztests	Cache-Treffer
1 kein Vorwissen	46,55 sec	657	6889750	72,66 %
2 Vorwissen: 1 Ecke	0,15 sec	94	1943122	93,16 %
3 Vorwissen: Dachflächen	15,96 sec	450	9625251	86,58 %
4 kein Caching	60,07 sec	657	6889750	0,00 %
5 nur FCIR	835,29 sec	22095	15498723	33,07 %
6 feste Instantiierungsreihenfolge und Vorwissen: Dachflächen	1191,30 sec	659	13222763	79,73 %

**Tabelle 3.2:** Meßergebnisse von sechs verschiedenen Suchvorgängen. Die ersten drei Suchvorgänge unterscheiden sich nur bzgl. des einbezogenen Vorwissens. Die anderen drei Suchvorgänge zeigen die Auswirkungen unterschiedlicher Implementierungen.

gen Rechnersystem und der Implementierung abhängt, wird für die Bewertung der Effizienz von Constraint-Techniken häufig die Anzahl der besuchten Suchbaumknoten und die Anzahl durchgeführter Konsistenztests betrachtet. Diese Werte sind zum Vergleich ebenfalls in der Tabelle aufgeführt.

Bei dem ersten Suchvorgang wurde kein Vorwissen berücksichtigt. Es läßt sich erkennen, daß trotz der großen Anzahl von Bildmerkmalen nur 657 Suchbaumknoten besucht werden mußten, um die Lösung des CSP zu finden. Die Anwendung der Constraint-Techniken hat zu einer massiven Einschränkung des Suchraumes geführt. Wenn vor der Suche bereits die Zuordnungen der Bildmerkmale für eine Ecke bekannt sind, wird die vollständige Zuordnung aller Merkmale in weniger als einer Sekunde ermittelt (zweiter Suchvorgang). Sind aufgrund eines vorhandenen digitalen Oberflächenmodells bereits die möglichen Dachflächen bekannt, kann die Zuordnung immerhin in einem Drittel der ursprünglichen Suchdauer bestimmt werden (dritte Meßreihe).

Wie an dem vierten Suchvorgang im Verhältnis zum ersten erkennbar ist, spielt das Caching von Wertebereichsreduktion in diesem Beispiel keine entscheidende Rolle; die Beschleunigung der Suche liegt nur bei einem Faktor von ca. 1,3. Dies wird sich später in Kapitel 4 bei der Erweiterung der Constraints um Unbeobachtbarkeit und Unschärfe jedoch ändern.

Alle Constraints bis auf die Merkmalsnachbarschaften werden mittels der Forward-Checking-Inferenzregel (FCIR) ausgewertet; für letztere wird hingegen die Look-Ahead-Inferenzregel (LAIR) verwendet. Obwohl empirische Tests ergeben haben, daß Forward-Checking in den meisten Fällen effizienter arbeitet als Look-Ahead (siehe [HE80, Nud83, vH89]), stellt sich dies für die Merkmalsnachbarschafts-Constraints genau andersherum dar, wie an den unterschiedlichen Ergebnissen der Suchvorgänge 1 und 5 erkennbar ist. Im fünften Suchvorgang wurden auch die Merkmalsnachbarschafts-Constraints mittels der FCIR ausgewertet, was zu einer ca. 18-fachen Erhöhung der Suchdauer geführt hat. Dabei mußten 33-mal so viele Suchbaumknoten besucht werden, als in Suchvorgang 1. Der Grund für die höhere Effizienz bei der Verwendung der LAIR für die Merkmalsnachbarschafts-Constraints gegenüber der FCIR liegt in der Lokalität der Nachbarschaftsrelation, die die Anzahl der zu einem Merkmal in Relation stehenden anderen Merkmale stark begrenzt. Wenn einer Modellvariablen ein Bildmerkmal zugewiesen wird, schränkt die LAIR die Wertebereiche der benachbarten Merkmalsvariablen unmittelbar stark ein.

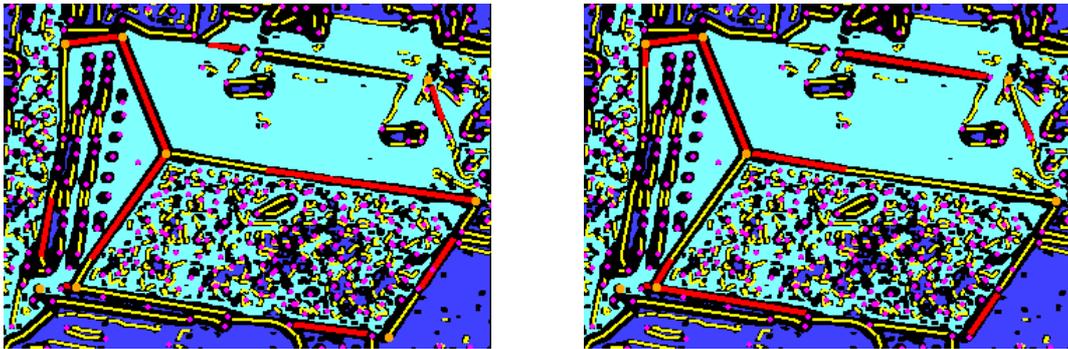
Zur Bestimmung der Reihenfolge, in der den Modellvariablen Bildmerkmale zugeordnet werden, wird in den ersten fünf Suchvorgängen jeweils die First-Fail-Heuristik verwendet, die als nächste zu

<sup>7</sup>Die Laufzeiten wurden auf einem Linux-PC mit 266 MHz AMD-K6-CPU unter ECLIPSE 3.7.1 ermittelt.

instantiierende Variable diejenige mit dem kleinsten Wertebereich dynamisch auswählt. Suchvorgang 6 zeigt in bezug auf Suchvorgang 3, daß die Verwendung einer statischen Instantiierungsreihenfolge, bei der zuerst die Flächen-, dann die Linien- und zuletzt die Punktvariablen instantiiert werden, die Suchdauer erheblich verlängert.

### 3.5 Gruppierung fragmentierter Liniensegmente

Jede im Bild zu suchende Gebäudeansicht wird, wie in den vorangegangenen Abschnitten erläutert, als CSP modelliert, bei dem jedem Modellmerkmal genau ein passendes Bildmerkmal zugeordnet werden soll. Nun kommt es vor, daß z.B. eine Dachkante bei der Merkmalsextraktion in mehrere (kürzere) Liniensegmente zerfällt. Von diesen Fragmenten können u.U. mehrere die geforderten Constraints bzgl. der anderen Modellmerkmale erfüllen. Das bedeutet, daß es mehrere Lösungen für das CSP geben kann, die zwar logisch verschieden, jedoch qualitativ gleich sind, weil sie dasselbe Gebäude mit denselben Parametern beschreiben. Schlimmer noch: für jedes Modellmerkmal, das im Bild fragmentiert erscheint, multipliziert sich die Anzahl qualitativ gleicher CSP-Lösungen mit der Anzahl der passenden Fragmente. Abbildung 3.13 zeigt ein Beispiel mit realen Daten.



**Abbildung 3.13:** Zwei von 972 möglichen Zuordnungen von Bild- zu Modellmerkmalen, die sämtliche Constraints erfüllen. Zugeordnete Linien sind rot, Flächen hellblau und Punkte orange hervorgehoben. Obwohl es sich jeweils um unterschiedliche Lösungen des CSP handelt, bezeichnen alle qualitativ dasselbe Gebäude.

Für die Gebäudeerkennung sind jedoch nur qualitativ verschiedene Matchings von Interesse. Es besteht also offensichtlich eine Differenz zwischen der Aufgabenstellung und dem gewählten Verfahren, die im folgenden untersucht wird. Dazu wird als erstes geklärt, inwieweit die verursachten Schwierigkeiten probleminhärent und damit unvermeidbar sind, und welche Probleme durch die verwendete Modellierung bzw. Repräsentation als CSP bedingt sind. Danach werden drei verschiedene Lösungsansätze anhand von Arbeiten, die sich ebenfalls mit der Fragmentierung im Kontext von Zuordnungsproblemen auseinandersetzen, diskutiert. Zuletzt wird dann ein eigener Ansatz vorgestellt, der eine Berücksichtigung der Fragmentierung im Rahmen der bisherigen Constraint-Modellierung ermöglicht.

Die Fragmentierung ist zunächst eine Folge der Segmentierung, die semantisch (also bzgl. der zu erkennenden Objekte) mit hoher Wahrscheinlichkeit zusammengehörende Bereiche des Rasterbildes zu einem Merkmal zusammenfaßt und diese damit gleichzeitig von anderen Bildbereichen abgrenzt [Rad93]. Da zur Trennung und Klassifizierung der Bildbereiche zum einen fast immer nur lokale Kriterien herangezogen werden (z.B. Grauwert oder Gradienten der Nachbarpixel wie in [Rad93, För94, Fuc97]) und zum anderen nur ein stark eingeschränktes Wissen bzgl. der zu erkennenden Objekte zur Verfügung steht, ist es für die Segmentierung oftmals nicht entscheidbar, ob z.B. durch einen Schattenwurf auseinandergetrennte Bildbereiche zum selben Objektteil gehören oder nicht.

Der wesentliche Grund für die weitgehende Abstraktion der Segmentierungsverfahren von den im Bild befindlichen Objekten ist der, daß sie meistens allgemeingültig im Hinblick auf die verschiedenen zu segmentierenden Bilder (z.B. Bildmaßstab) und die darin vorkommenden Objekte (Gebäude, Straßen, aber auch Objekte anderer Anwendungsgebiete wie z.B. Innenraumszenen) konzipiert sind (vgl. [Fuc97]). Oftmals besteht das einzig verwendete Vorwissen darin, daß Objekte sich durch scharfe, lineare und stetige Ränder von anderen Objekten absetzen und ebene Objektflächen als homogene Flächen im Bild erscheinen. Wird mehr Objektwissen wie z.B. über die Radiometrie und die dreidimensionale Geometrie der gesuchten Objekte in das Segmentierungsverfahren einbezogen, läßt sich die Qualität der extrahierten Merkmale zwar deutlich erhöhen (siehe [Hen95, Hen96]), jedoch treten dabei weiterhin Fragmentierungen auf.

Die Segmentierung bildet die Schnittstelle zwischen der Low-Level-Bildverarbeitung und dem darauf aufsetzenden Prozeß der Bildinterpretation [Rad93]. Die Verknüpfung zwischen diesen beiden Ebenen wird durch ein gemeinsames Bildmodell realisiert, das in unserem Fall aus der überlappungsfreien Darstellung von Objekten durch Punkte, Linien und Flächen besteht. Bislang wurde von einem idealen Bildmodell mit einer 1:1-Beziehung zwischen Modell und Bild ausgegangen, d.h. einer Fläche des Modells entspricht genau eine extrahierte Fläche im Bild; für Linien und Punkte gilt das gleiche. Unter realen Bedingungen handelt es sich aufgrund der Fragmentierung jedoch um eine 1 :  $n$ -Beziehung, bei der einem Modellmerkmal mehrere Bildmerkmale entsprechen können.

Die Annahme einer 1:1-Zuordnung liegt auch der in den vorangegangenen Abschnitten dieses Kapitels dargestellten Formulierung der Gebäudeerkennungsaufgabe als CSP zugrunde. Diese Formulierung läßt sich auf 1:n-Beziehungen erweitern, indem man die einfachen Variablen, denen jeweils nur genau ein Merkmal zugeordnet werden kann, durch mengenwertige Variablen ersetzt. Die Wertebelegung der Variablen einer CSP-Lösung bestünde dann jeweils aus der entsprechenden Menge der bzgl. der Constraints äquivalenten Merkmale. Bei dieser Erweiterung besteht jedoch die große Gefahr einer Explosion des ohnehin bereits großen Suchraumes, denn jedem Objektmerkmal kann dann nicht mehr nur eines von  $m = |M|$  extrahierten Bildmerkmalen  $M$  zugeordnet werden, sondern theoretisch eine beliebige Teilmenge  $M' \subseteq M$ . Die Anzahl der möglichen Teilmengen ist durch die Kardinalität der Potenzmenge  $|\wp(M)| = 2^m$  gegeben. Das würde für die Suche nach einem Objekt mit  $k$  Merkmalen einen Suchraum von  $2^{m \cdot k}$  statt vormals  $m^k$  möglichen Zuordnungen bedeuten. Wenn eine solche Erweiterung vorgenommen werden soll, muß — wie im folgenden Abschnitt erläutert — der Suchraum durch zusätzliche Bedingungen deutlich beschränkt werden.

### 3.5.1 Gruppierungskonzepte im Kontext von Zuordnungsverfahren

Obwohl die Fragmentierung ein ernstes Problem für relationale Zuordnungsverfahren darstellt, wurde bis vor kurzem in den meisten Arbeiten nicht weiter darauf eingegangen, oder es wurde gar nicht erst angesprochen (vgl. [SH87, HS93a]). Oftmals wurde das Problem dadurch ausgeblendet, daß für die zu erkennenden Szenen die idealisierte Annahme einer perfekten Segmentierung gemacht wurde (vgl. [Vos92]).

Erst in den letzten Jahren sind Arbeiten veröffentlicht worden, die sich explizit mit der durch die Fragmentierung verursachten Probleme in diesem Kontext beschäftigen. Dies hängt sicherlich auch damit zusammen, daß relationale Zuordnungsverfahren zum einen aufgrund ihrer kombinatorischen Komplexität und zum anderen aufgrund der durch Bildstörungen verursachten Fehler noch nicht für die Interpretation komplexer realer Bildszenen wie z.B. Luftbilder eingesetzt wurden<sup>8</sup>.

Im folgenden werden drei Ansätze vorgestellt, die das Problem auf verschiedene Weise lösen, wobei es sich bei den ersten beiden um Anwendungen aus dem Bereich der Objekterkennung und bei dem dritten um ein Zuordnungsproblem von Geodaten unterschiedlicher Herkunft handelt.

<sup>8</sup>Bislang wurden relationale Zuordnungsverfahren nur für einfache reale Bilder wie z.B. zur Qualitätsüberprüfung in der industriellen Fertigung eingesetzt. Die dazu zu analysierenden Bilder sind optimal ausgeleuchtet, haben einen definierten Hintergrund und enthalten nur bekannte Objekte (siehe [SFH92, Rad93]).

### CSPs mit mengenwertigen Variablen und Bilevel-Constraints

Deruyver und Hodé beschreiben in ihrem Artikel [DH97] die Anwendung von Constraint-Techniken zur Interpretation von Computertomographiebildern zwecks Klassifikation der einzelnen Hirnregionen. Die Fragmentierung rührt hier zum einen daher, daß die dreidimensional ausgedehnten Hirnregionen bei der Tomographie auf eine Sequenz von Querschnittsbildern des Kopfes abgebildet werden, und somit eine Hirnregion auf mehreren Bildern erscheint. Zum anderen können auch Flächen bestimmter Hirnbereiche innerhalb eines Querschnittsbildes in kleinere Teilflächen fragmentiert sein. Das Ziel ist es, alle extrahierten Flächen der Querschnittsbilder anhand eines relationalen Modells den verschiedenen Hirnregionen zuzuordnen.

Das Zuordnungsproblem wird als CSP formuliert, wobei die Variablen mengenwertig sind. Um dem oben erläuterten Problem der kombinatorischen Explosion entgegenzuwirken, wird neben den Constraints zwischen den Variablen noch eine zweite Sorte von Constraints auf den Werten einer Variablen eingeführt. Diese Constraints beschreiben, welche Werte in einer Variablen zusammen gruppiert werden dürfen. Die für diese Anwendung verwendeten Gruppierungs-Constraints basieren zum einen auf den morphologischen Eigenschaften der Flächenmerkmale wie Form und Größe sowie der geometrischen Nähe und Orientierung. Die auf diese Weise unterschiedenen Constraints werden als *Bilevel-Constraints* bezeichnet.

Zur Berechnung der Klassifikation wird eine Modifikation des AC4-Algorithmus (siehe Abschnitt 3.1.3) vorgeschlagen. Dieser Algorithmus erzwingt zwar nur die Kantenkonsistenz, jedoch konnte für die Anwendung gezeigt werden, daß diese zur eindeutigen Klassifikation der Segmente ausreicht<sup>9</sup>.

### Gruppierung von Linien anhand bekannter Gebäudepositionen

Huertas und Nevatia stellen in [HN98] ein Verfahren zur Erkennung von Änderungen in Luftbildern vor. Die Aufgabe besteht darin, anhand eines Luftbildes für die in einer Datenbank gespeicherten Gebäude festzustellen, ob sie noch vorhanden sind bzw. noch die ursprünglich erfaßte Form besitzen. Zur Verifikation werden die 3D-Gebäudekoordinaten in den Bildraum projiziert und für jede Linie des Gebäudemodells geschaut, ob passende Linienfragmente in dem erwarteten Bereich — ein schmaler rechteckiger Streifen zwischen den zwei Linienendpunkten — im Bild beobachtbar sind. Es werden dabei nur solche Linien gruppiert, die innerhalb bestimmter Schwellwerte kollinear zueinander sowie parallel zur Modelllinie sind. Leider werden keine näheren Angaben zur Größe und Herkunft dieser Schwellwerte gemacht.

Aufgrund der Tatsache, daß für jedes Gebäude der Typ und die Position bereits bekannt ist, muß nicht mehr gesucht, sondern nur noch gruppiert werden. Dabei gestaltet sich die Gruppierung wegen der sehr präzise beschriebenen Bildbereiche, in denen nach passenden extrahierten Bildlinien gesucht wird, sehr einfach.

### n:m-Zuordnung mittels 1:1-Zuordnung a priori gruppierter Objekte

In seiner Dissertation behandelt Walter das Problem der Zuordnung von Geodaten unterschiedlicher Herkunft [Wal96]. Dabei werden Straßendaten aus den beiden Datenquellen ATKIS<sup>10</sup> und GDF<sup>11</sup> desselben geographischen Gebietes einander zugeordnet, um einen automatischen Abgleich der in den jeweiligen Datenbeständen zu den Straßen vorhandenen Informationen zu ermöglichen.

Straßen werden (bis auf wenige Ausnahmen) in beiden Modellen durch die Linienzüge der Straßenmittelachsen repräsentiert. Aufgrund der unterschiedlichen Digitalisierung weicht die Anzahl der Stützpunkte einer Straße in den ATKIS-Daten bzgl. der GDF-Daten derselben

<sup>9</sup> Ähnlich dazu verhält es sich auch mit dem Waltz-Algorithmus [Wal75] zur Interpretation von Linienzeichnungen mit Polyederszenen. Hier reicht fast immer bereits die Erzwingung von Kantenkonsistenz zur vollständigen Interpretation der Szene aus.

<sup>10</sup> ATKIS steht für Amtliches Topographisch-Kartographisches Informationssystem. Es umfaßt das Datenmodell und den Datenbestand des digitalen Landschaftsmodells der Landesvermessungsämter.

<sup>11</sup> GDF steht für Geographic Data File. Es wurde speziell für Anwendungen der Fahrzeugnavigation entworfen.

Straße in der Regel voneinander ab. Es handelt sich also um ein  $n:m$ -Zuordnungsproblem, weil  $n$  ATKIS-Straßenabschnitten  $m$  GDF-Abschnitte entsprechen können. Um dieses Problem der Mehrfachzuordnungen zu lösen, werden nun vor dem eigentlichen Zuordnungsprozeß alle in Frage kommenden Gruppierungen kleinerer Straßenabschnitte sowohl für die ATKIS- als auch die GDF-Daten vorgenommen. Jede dieser Gruppierungen wird als eigenes Objekt der Menge der Straßenabschnitte hinzugefügt. Somit muß das Zuordnungsverfahren nur noch die beste 1:1-Zuordnung von ATKIS- zu GDF-Straßenabschnitten bestimmen.

Diskussion der vorgestellten Ansätze:

Keines der drei Konzepte läßt sich unmittelbar für das Problem der Gebäudeerkennung einsetzen. Bei dem Konzept der Bilevel-Constraints stellt sich die Frage nach den geeigneten Gruppierungs-Constraints, die ausschließlich beschreiben, welche Linien gruppierbar sind. Wie bei dem zweiten Ansatz erläutert, dürfen nur Linien gruppiert werden, die der Lage, Richtung und Ausdehnung der einzelnen Linien eines konkreten Gebäudes entsprechen. Da die Gebäudeparameter jedoch erst nach einem erfolgten Matching bestimmt werden, lassen sich keine von den Constraints des Gebäudemodells entkoppelten Constraints für die Gruppierbarkeit von Linien definieren. Aus diesem Grund läßt sich auch der Ansatz der a priori Gruppierung, die alle möglichen Gruppierungen vor der eigentlichen Gebäudeerkennung erzeugt, nicht anwenden.

Ferner reicht es bzgl. des Konzepts der Bilevel-Constraints bei der Gebäudeerkennung nicht aus, nur die Kantenkonsistenz für das CSP zu erzwingen. Vielmehr müssen bis zur globalen Konsistenz und damit eindeutigen Bestimmung einer Lösung zwischendurch immer wieder Werte für Variablen generiert werden (vgl. Abschnitt 3.1.3.1). Hier gäbe es aber pro Variablenzuweisung aufgrund der Mengenwertigkeit wieder  $2^m$  Möglichkeiten, was das Verfahren darüberhinaus unpraktikabel machen würde.

Beim Ansatz von Huertas und Nevatia werden zwar nur die tatsächlich für eine Gebäudehypothese in Frage kommenden Linien gruppiert, jedoch muß dazu vorher die Position der einzelnen Gebäudemodelllinien im Bild bekannt sein. Im folgenden Abschnitt wird deshalb ein zweistufiges Gruppierungsverfahren vorgestellt, das zunächst die Geometrie des Gebäudemodells und anschließend die zu gruppierenden Linien bestimmt.

### 3.5.2 Geometrische Gruppierung

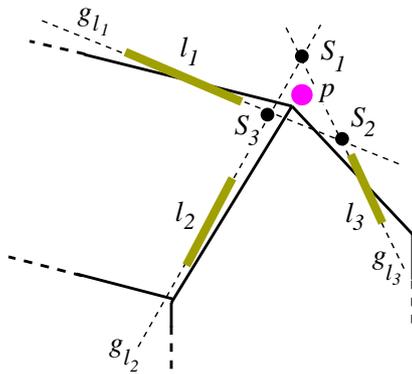
Sobald eine CSP-Lösung bestimmt ist, läßt sich daraus die Geometrie der Gebäudeansicht bestimmen. Die Gebäudegeometrie wird unmittelbar nur durch die Koordinaten der Gebäudeeckpunkte bestimmt, da die Geometrien der Linien als auch der Flächen bei der für Gebäudeansichten verwendeten B-Rep-Repräsentation implizit über die entsprechenden Inzidenzrelationen zwischen Linie und Punkt bzw. Fläche und Linie definiert sind. Das bedeutet, daß im ersten Schritt die Gebäudeeckpunktkoordinaten aus den dem Modell zugeordneten Bildmerkmalen berechnet werden müssen.

Im zweiten Schritt werden dann, ähnlich zu [HN98], um jede Linie der Gebäudeansicht rechteckige Pufferbereiche gelegt, und alle darin liegenden Linienmerkmale ermittelt, die zur berechneten Modelllinie kollinear und parallel sind. Dabei gilt es jedoch die Einführung neuer Schwellwerte zu vermeiden. In den beiden folgenden Unterabschnitten werden nun die beiden Schritte näher erläutert.

#### 3.5.2.1 Berechnung der Eckpunktkoordinaten der Hausansicht

Zur Bestimmung der Eckpunktkoordinaten könnte eine globale Ausgleichung, wie sie in Abschnitt 2.5.7 vorgestellt wurde, durchgeführt werden. Wie dort bereits erläutert wurde, ist diese jedoch sehr aufwendig und sollte deshalb nicht während der Suche sondern nur nach erfolgter Gebäudeerkennung zur Rekonstruktion der Gebäudeparameter aufgerufen werden.

Für die Gruppierung der Linien wird daher vorgeschlagen, die Eckpunktkoordinaten lokal, d.h. für jeden Eckpunkt isoliert zu bestimmen. Dazu muß zunächst überlegt werden, welche Merkmale bei der Berechnung der Eckpunktkoordinaten berücksichtigt werden müssen. Dies sind zum einem das dem Gebäudeeckpunkt zugeordnete Punktmerkmal und zum anderen die Linienmerkmale, die den zu dieser Ecke inzidenten Modelllinien zugeordnet wurden. Abbildung 3.14 zeigt, welche Merkmale beispielsweise in die Berechnung der Giebelpunktkoordinaten eines Satteldachhauses eingehen.



**Abbildung 3.14:** Bestimmung der Koordinaten des Giebelpunktes anhand der den Modellmerkmalen zugeordneten Bildmerkmale. Dem entsprechenden Punktmerkmal des Modells wurde der extrahierte Punkt  $p$  und den drei dazu inzidenten Linien des Modells die extrahierten Liniensegmente  $l_1, l_2, l_3$  zugeordnet. Die Abweichungen sind stark überzeichnet, damit deutlich wird, daß sich nicht alle Geraden in einem Punkt (und zwar  $p$ ) treffen. In die Schätzung der Giebelpunktkoordinaten gehen die Koordinaten des Punktes  $p$  und die der drei Schnittpunkte  $S_1, S_2, S_3$ , der durch die Liniensegmente verlaufenden Geraden  $g_{l_1}, g_{l_2}, g_{l_3}$ , ein.

Bei der Bestimmung der Eckpunktkoordinaten müssen die Genauigkeiten des Punktmerkmals sowie der berechneten Schnittpunkte der durch die Liniensegmente verlaufenden Geraden berücksichtigt werden. Die Genauigkeit der Schnittpunkte wird durch die Länge der Liniensegmente, der Unsicherheit ihrer Endpunktkoordinaten, der Entfernung der Liniensegmente zum Schnittpunkt und maßgeblich vom Schnittwinkel der Geraden beeinflusst.

Bei den Linienparallelitäts- und Kollinearitäts-Constraints blieben bei der konkreten Umsetzung bereits vereinfachend die Liniensegmentlänge und die Unsicherheit der Liniensegmentendpunktkoordinaten sowie beim Kollinearitäts-Constraint die Entfernung des Punktes zum Liniensegment unberücksichtigt. Es würde deshalb an dieser Stelle wenig Sinn machen, diese nun bei der Berechnung der Eckpunktkoordinaten einzubeziehen. In der jetzigen Form wird also nur der Schnittwinkel als gewichtender Faktor berücksichtigt.

Die Unsicherheit der Endpunktkoordinaten einer extrahierten Linie führt dazu, daß der Bereich, in dem die durch die Linie verlaufende Gerade liegen kann, sich an beiden Enden des Liniensegmentes trichterförmig öffnet (siehe [Fuc97]). Der wahre Schnittpunkt zweier solcher Geraden kann irgendwo auf der Schnittfläche der beiden Bereiche liegen. Je kleiner diese Fläche wird, desto größer ist die Wahrscheinlichkeit, daß der berechnete Schnittpunkt nahe des wahren Schnittpunktes liegt. Nun wird die Schnittfläche der Bereiche, durch die die Geraden verlaufen können, größer, je stumpfer bzw. spitzer der Schnittwinkel wird, weil sich die Bereiche dadurch über eine längere Distanz überlappen. Der Bereich ist am kleinsten, wenn die zwei Geraden orthogonal zueinander stehen. Diese Beobachtung geht nun durch die Gewichtung der berechneten Schnittpunktkoordinaten mit dem Kosinus des Schnittwinkels  $\varphi$  in die Berechnung der Eckpunktkoordinaten ein.

Seien  $(x_p, y_p)$  die Koordinaten des der Gebäudeecke zugeordneten Punktmerkmals,  $l$  die Anzahl der zu dem Eckpunkt inzidenten Linien und  $(x_i, y_i)$  die  $n = l(l-1)/2$  Schnittpunkte der durch die Linien verlaufenden Geraden, sowie  $\varphi_i$  deren Schnittwinkel. Dann werden die Koordinaten  $(\tilde{x}, \tilde{y})$  des Eckpunktes im Bild durch die folgende Gleichung geschätzt:

$$\tilde{x} = \frac{x_p + \sum_{i=1}^n (x_i \cos \varphi_i)}{1 + \sum_{i=1}^n \cos \varphi_i} \quad \tilde{y} = \frac{y_p + \sum_{i=1}^n (y_i \cos \varphi_i)}{1 + \sum_{i=1}^n \cos \varphi_i} \quad (3.2)$$

### 3.5.2.2 Bestimmung zu gruppierender Liniensegmente

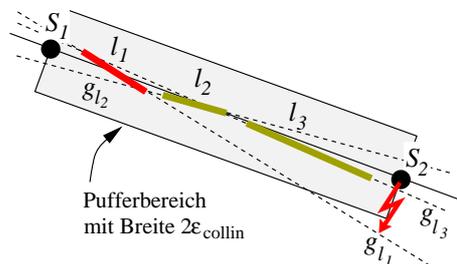
Nachdem die Koordinaten der Gebäudeeckpunkte geschätzt worden sind, können die zu gruppierenden extrahierten Liniensegmente selektiert werden. Dazu wird zunächst um jede Linie

der Gebäudeansicht ein rechteckiger Pufferbereich gelegt. Da laut Constraint-Repräsentation des Gebäudemodells die Eckpunkte kollinear zu den inzidenten Linien sein müssen, wird der Puffer oberhalb sowie unterhalb der Linie auf den Schwellwert  $\varepsilon_{\text{collin}}$  der Kollinearitäts-Constraints beschränkt. Alle extrahierten Linienmerkmale, die in diesem Bereich liegen oder ihn zumindest schneiden, stellen Kandidaten für eine Gruppierung dar.

Nach dieser Selektion sind zwar nur noch Linienmerkmale enthalten, die (partiell) zwischen den Endpunkten der Modelllinie liegen und den Pufferbereich zumindest schneiden, jedoch könnten diese Linien zum einen noch eine zu starke Winkelabweichung bzgl. der Modelllinie aufweisen und zum anderen nur zu einem der beiden Eckpunkte kollinear sein. Um dies auszuschließen, müssen alle Gruppierungskandidaten zusätzlich die folgenden zwei Bedingungen erfüllen:

1. Die Linienmerkmale müssen kollinear zu beiden Endpunkten der Modelllinie sein. Zur Überprüfung wird der Schwellwert des Kollinearitäts-Constraints  $\varepsilon_{\text{collin}}$  herangezogen.
2. Die Linienmerkmale müssen parallel zur Modelllinie sein. Zur Überprüfung wird der Schwellwert des Linienparallelitäts-Constraints  $\varepsilon_{\text{parallel}}$  herangezogen.

Nur eine dieser Bedingungen allein reicht nicht aus, z.B. würde bei Modelllinien, die kürzer als  $2\varepsilon_{\text{collin}}$  sind, eine zwischen den Endpunkten der Modelllinie senkrecht verlaufende extrahierte Linie noch als kollinear gewertet. In diesem Fall würde das Linienmerkmal dann aufgrund der zweiten Bedingung zurückgewiesen. In einem anderen Fall könnte eine extrahierte Linie zwar den Parallelitäts-Constraint erfüllen, jedoch nur zu einem Endpunkt der Modelllinie kollinear sein. Die folgende Abbildung 3.15 illustriert abschließend die Selektion der zu gruppierenden Linienmerkmale.



**Abbildung 3.15:** Bestimmung zu gruppierender Linienfragmente. Es wird eine Modelllinie gezeigt, deren Endpunkte  $S_1, S_2$  mit Hilfe von Gleichung 3.2 aus den dem Modell zugeordneten Bildmerkmalen berechnet wurde. Linien  $l_2$  und  $l_3$  werden gruppiert, weil sie sowohl den Pufferbereich schneiden (in diesem Fall ganz darin liegen), als auch der Abstand der beiden Eckpunkte  $S_1, S_2$  zu den durch die Linienfragmente verlaufenden Geraden  $g_{l_2}$  und  $g_{l_3}$  kleiner  $\varepsilon_{\text{collin}}$  ist. Linie  $l_1$  wird nicht einbezogen, weil der Abstand von  $S_2$  zu  $g_{l_1}$  größer  $\varepsilon_{\text{collin}}$  ist.

### 3.5.3 Einbettung der Gruppierung in das CSP

Das im letzten Abschnitt vorgestellte Gruppierungsverfahren bestimmt die zu einer Lösung des CSPs gruppierbaren Linienmerkmale für jede Linie des Modells. Eine solche Gruppierung stellt auch eine Zusammenfassung qualitativ gleicher Lösungen dar, und zwar aller CSP-Lösungen, die von den gruppierten Linien subsumiert werden.

Um nun zu verhindern, daß nach der ersten Berechnung einer CSP-Lösung weitere, qualitativ gleiche CSP-Lösungen generiert werden, wird unmittelbar nach der Gruppierung die gruppierte Lösung in einer Datenbank vermerkt. Für jede weitere CSP-Lösung wird noch bevor sie ausgegeben bzw. als Ausgangspunkt für eine weitere Gruppierung verwendet wird, überprüft, ob sie bereits durch eine gruppierte Lösung in der Datenbank subsumiert wird. Ist dies der Fall, wird die neue CSP-Lösung verworfen und die nächste bestimmt. Im anderen Fall stellt die neue CSP-Lösung offensichtlich einen Vertreter einer qualitativ neuen Lösung dar, und es wird erneut gruppiert und das Ergebnis der Datenbank hinzugefügt. Nachdem alle CSP-Lösungen bestimmt wurden, können die qualitativ verschiedenen und gruppierten Zuordnungen der Datenbank entnommen werden.

Im folgenden bezeichne  $(l_1, \dots, l_n)$  die den  $n$  Linienvariablen einer Gebäudeansicht zugeordneten Bildmerkmale einer CSP-Lösung,  $(L_1^i, \dots, L_n^i)$  die den  $n$  Linienvariablen zugeordneten Mengen gruppierter Bildmerkmale der  $i$ -ten Gruppierung und  $\mathcal{L} = \{(L_1^1, \dots, L_n^1), \dots, (L_1^k, \dots, L_n^k)\}$  die Gesamtmenge der  $k$  Gruppierungen bzw. qualitativ unterschiedlichen Lösungen. Eine aktuelle CSP-Lösung  $(l_1, \dots, l_n)$  wird dann von einer Gruppierung in  $\mathcal{L}$  subsumiert (Notation:  $(l_1, \dots, l_n) \models \mathcal{L}$ ), wenn gilt

$$\exists (L_1, \dots, L_n) \in \mathcal{L} : \bigwedge_{i=1}^n l_i \in L_i \quad (3.3)$$

Sowohl der Test auf eine neue Lösung als auch die Gruppierung wird durch einen einzigen neuen Constraint `grouping` in CLP realisiert, der der Constraint-Klausel in Abb. 3.11 (auf Seite 56) hinzugefügt wird. Dem Constraint wird als Argument die Liste aller Variablen des CSP übergeben. Da eine Gruppierung erst stattfinden kann, wenn eine vollständige und gültige Zuordnung vorliegt, wird der Constraint so spezifiziert, daß er erst geweckt wird, wenn alle Variablen gebunden sind. Durch die Vergabe der niedrigsten Bearbeitungspriorität wird zudem sichergestellt, daß der Constraint erst bearbeitet wird, wenn alle anderen Constraints bereits (erfolgreich) getestet wurden.

Der Constraint-Rumpf besteht aus zwei Teilen: Zuerst wird überprüft, ob die vorliegende Lösung bereits durch eine vorherige subsumiert wird. In diesem Fall schlägt der Constraint fehl; es wird ein Backtracking ausgelöst und die aktuelle Lösung verworfen. Im anderen Fall wird die Gruppierung gemäß der in Abschnitt 3.5.2 gezeigten Schritte durchgeführt und die gruppierten Lösungen dann in Form eines Fakts `solution(List_of_Lists)` der Prolog-Faktendatenbank hinzugefügt. `List_of_Lists` enthält für jede Linie der Gebäudeansicht eine Liste mit den von der Gruppierung entsprechend zusammengefaßten extrahierten Linienmerkmalen. Für den Test auf Subsumption einer neuen Lösung reicht es, für eine neu berechnete CSP-Lösung nachzuschauen, ob es einen Eintrag in der bisherigen Lösungsmenge (den `solution`-Fakten) gibt, der das einer jeden Linienvariablen aktuell zugeordnete Bildmerkmal in der entsprechenden Liniengruppe enthält. Nach der Suche nach allen CSP-Lösungen können die gruppierten Lösungen schließlich der Faktendatenbank entnommen werden, wobei jedes `solution`-Fakt eine unterschiedliche Lösung darstellt.

Um eine hohe Genauigkeit bei der Bestimmung der Gebäudeeckpunktkoordinaten zu erzielen, sollten möglichst lange Linienmerkmale zur Berechnung der Schnittpunkte verwendet werden. Dies kann durch eine geeignete Reihenfolge bei der Instantiierung von Linienvariablen erreicht werden: Wenn bei der Generierung von Werten die extrahierten Linienmerkmale nach absteigender Länge sortiert zugeordnet werden, ist die erste CSP-Lösung diejenige mit den längsten passenden Linienmerkmalen.

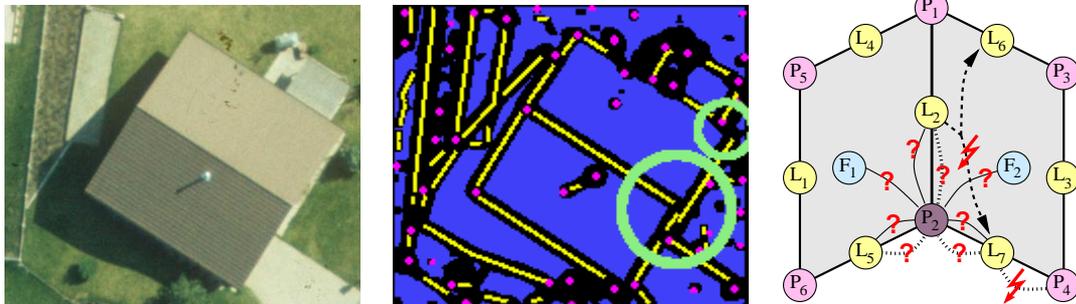
Als Fazit ergibt sich, daß auch das Problem der Fragmentierung mittels Constraint-Techniken, insbesondere unter Verwendung der Möglichkeiten von CLP (Constraint-Prioritäten und Prolog-Faktendatenbank), auf elegante Weise gelöst werden kann. Die Effizienz des bisherigen Verfahrens wird kaum beeinträchtigt, weil der Aggregations-Constraint nur geweckt wird, wenn alle Variablen bereits gebunden und alle anderen Constraints erfüllt sind. Das bedeutet, daß der Test auf Subsumption nur so oft aufgerufen wird, wie ursprüngliche CSP-Lösungen vorliegen. Die Gruppierung wird noch seltener durchgeführt, und zwar einmal für jede qualitativ verschiedene Lösung. Der Berechnungsaufwand der Gruppierung ist linear in der Anzahl der Linienmerkmale des Modells, da die Überprüfung, ob ein Liniensegment ein Rechteck schneidet oder ganz in ihm liegt, in konstanter Zeit durchgeführt werden kann (vgl. [PS85]).



## Kapitel 4

# Erweiterung der Constraint-Techniken um Unbeobachtbarkeit und Unschärfe

Im vorangegangenen Kapitel wurde gezeigt, daß sich Gebäude in segmentierten Luftbildern, die aus einer großen Anzahl extrahierter Bildmerkmale bestehen, unter Verwendung von Constraint-Techniken effizient erkennen lassen. Dabei wurde die idealisierende Annahme gemacht, daß sich alle Constraints erfüllen lassen. Tatsächlich ist dies in realen Bilddaten aber oftmals nicht der Fall, wie Abbildung 4.1 illustriert.



**Abbildung 4.1:** Luftbildauschnitt mit einem Satteldachhaus (links) und extrahierte Bildmerkmale mit typischen Störungen (Mitte). Bei der Segmentierung konnte der rechte Giebelpunkt  $P_2$  nicht beobachtet werden. Die zu ihm inzidenten Relationen sind deshalb ebenfalls nicht beobachtbar, was durch die Fragezeichen im rechten Bild markiert wird. Darüberhinaus sind die Constraints `same_side_line(L2, L6, L7)` und `collinear(L7, P4)` verletzt, was durch die zwei Blitzsymbole im rechten Bild angedeutet wird.

Die Definition der Lösungen eines CSP (Def. 3.2 auf Seite 36) basiert auf der gleichzeitigen Erfüllung aller Constraints, d.h. bereits die Verletzung eines einzigen Constraints führt dazu, daß für das CSP keine Lösung bestimmt und damit auch das entsprechende Gebäudemodell nicht gefunden werden kann. Das in der obigen Abbildung gezeigte Haus würde demnach nicht erkannt werden.

Dies ist offensichtlich inakzeptabel, und wohl auch der maßgebliche Grund dafür, daß nach anfänglich (Anfang der 80er Jahre) vorgeschlagenen Anwendungen von Constraint-Techniken für die Objekterkennung die entwickelten Ansätze oftmals nicht über theoretische Studien hinausgingen, die sich nicht auf reale Daten übertragen ließen.

Stattdessen wurden für die Objekterkennung zunehmend Verfahren eingesetzt, die generell von der Annahme harter Constraints abweichen und Constraint-Netzwerke auf andere Art auswerten. Dies sind z.B. Fuzzy-Logik ([Spi93]), Bayes-Netze ([Pea88]), neuronale Netze ([Spi93]) und kontinuierliche Relaxation ([HZ83, Rad93]). Diese Konzepte sind in der Lage, die bzgl. bestimmter Bewertungskriterien beste Näherungslösung zu ermitteln. Bei den ersten drei Konzepten stellt sich aber die Definition der Kriterien zur Bewertung verschiedener Lösungen als problematisch dar. Insbesondere stellt sich die Frage nach der Semantik der numerischen Gewichte bzw. der Objektivierbarkeit der besten Lösung, denn weder die Fuzzy-Logik noch die neuronalen Netze erlauben eine direkte wahrscheinlichkeitstheoretische Interpretation. Bayes-Netze bestehen aus gerichteten Graphen, die kausale Abhängigkeiten durch Wahrscheinlichkeiten ausdrücken. Die im vorigen Kapitel verwendeten Constraints beschreiben jedoch ungerichtete, beidseitige Beziehungen zwischen den Merkmalen. Kulschewski beschreibt in [Kul97] eine Anwendung von Bayes-Netzen und Mohan und Nevatia in [MN88] eine Anwendung neuronaler Netze zur Gebäudeerkennung. An beiden Arbeiten wird deutlich, daß die Verwendung von Bayes-Netzen und neuronaler Netze eine andere, weniger natürliche Modellrepräsentation erfordert. Anders als bei der Constraint-Darstellung können die Merkmale und Relationen des Modells nicht unmittelbar in entsprechende Variablen und Constraints transformiert werden.

Das größte Problem der oben genannten Verfahren besteht jedoch darin, daß sie im allgemeinen keine den Konsistenztechniken entsprechenden Beschränkungen des Suchraumes erlauben. Aufgrund der Komplexität der Daten und der Modelle ist aber eine aktive Suchraumeinschränkung unverzichtbar. In diesem Kapitel wird deshalb eine Erweiterung der Constraint-Modellierung vorgestellt, die einerseits die auftretenden Bildfehler berücksichtigt und andererseits die Anwendung von Constraint-Techniken ermöglicht.

Dazu werden im folgenden Abschnitt zunächst die möglichen Fehlerfälle und ihre Auswirkungen auf die einzelnen Constraint-Typen systematisch untersucht. Anschließend werden die etablierten Konzepte zur Behandlung überbeschränkter Constraint-Systeme vorgestellt und auf ihre Eignung zur Verwendung der Objekterkennung diskutiert. Da sich zeigen wird, daß keines der bislang vorgestellten Konzepte alle verschiedenen Fehlerauswirkungen gleichzeitig handhaben kann, wird im Anschluß ein eigener Ansatz vorgestellt, der eine Integration zweier etablierter Konzepte zur Lösung überbeschränkter Constraint-Systeme darstellt.

## 4.1 Fehlerfälle

Es gibt viele verschiedene Gründe, warum Störungen in den Luftbildern auftreten. Diese hängen mit dem verwendeten Sensor (Kamera und Film), der Digitalisierung (Scanner), und der abgebildeten Szene (Beleuchtung, Perspektive und Bildinhalt) zusammen. Weitere Störungen können bei der Segmentierung und Merkmalsextraktion auftreten, welche den Ausgangspunkt für die Bildinterpretation darstellt.

Luftbildaufnahmen werden zunächst durch das Sensorrauschen der Kamera und des Films beeinträchtigt. Bei der Digitalisierung führt neben dem Rauschen des Scanners insbesondere das bei der Diskretisierung in einzelne Pixel entstehende Quantisierungsrauschen zu weiteren Störungen. Aufgrund der endlichen Auflösung entsteht ferner eine Unschärfe, und es können nur noch Objekte wahrgenommen werden, die eine bzgl. der Digitalisierungsauflösung ausreichende Größe im Bild besitzen. Die Quantisierung betrifft die Genauigkeit der geometrischen Eigenschaften der Objekte wie z.B. bei Linien die Lage sowie die Orientierung im Bild. Einen wesentlichen Einfluß auf die Beobachtbarkeit von Objekten hat die Aufnahmeperspektive in Verbindung mit der Szenenbeleuchtung. Bei ungünstigen Perspektiven können die Kontraste so klein werden, daß verschiedene Objekte bzw. Objekt und Bildhintergrund nicht mehr voneinander abgegrenzt werden können. Dabei spielt auch die Oberflächenbeschaffenheit der abgebildeten Objekte eine Rolle. Weitere Probleme werden durch teilweise Verdeckungen der zu erkennenden Objekte durch andere Objekte oder Schatten (Fremdverdeckung) hervorgerufen.

Die genannten Effekte wirken sich unmittelbar und mittelbar auf das Ergebnis der Merkmalsextraktion aus (vgl. [FLF94, Fuc97]). Darüberhinaus können bei der Merkmalsextraktion weitere Fehler entstehen. Beispielsweise können Linien, die im Bild zu verschiedenen Objekten gehören und zufällig kollinear sowie benachbart sind, zu einer langen Linie verschmolzen werden. Rauschen kann zur Extraktion von Bildmerkmalen führen, die eigentlich keine Entsprechung in der abgebildeten Szene besitzen. Die Klassifikation der einzelnen Pixel, ihre anschließende Aggregation und die nachfolgende Approximation durch Punkt-, Linien- und Flächenmerkmale bedeutet eine Diskretisierung, mit der ebenfalls weitere Unsicherheiten und Unschärfen verbunden sind.

In [FF95, Fuc97] diskutieren Fuchs und Förstner die verschiedenen Fehlerfälle und teilen ihre Auswirkungen auf das Ergebnis der Merkmalsextraktion in fünf Kategorien ein. Unter die erste Kategorie fallen (zufällige) Störmerkmale durch Rauschen oder Textur, die zweite Kategorie bezeichnet die vollständige Unbeobachtbarkeit und die dritte Kategorie die partielle Unbeobachtbarkeit von Bildmerkmalen. In die vierte Kategorie fällt die Fragmentierung und in die fünfte die Verschmelzung von Merkmalen. Diese Aufteilung wird in dieser Arbeit um die Fälle ergänzt, die bei der Objekterkennung durch den Kontext eines Objektmodells hinzukommen:

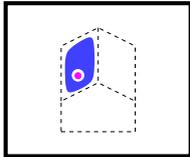
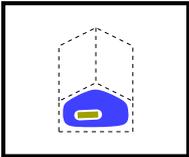
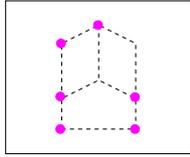
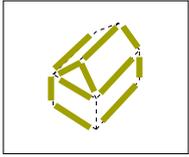
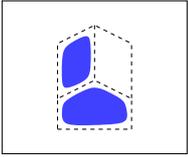
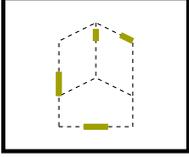
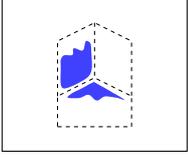
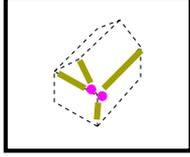
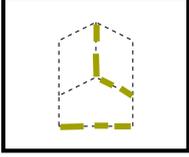
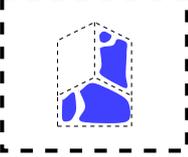
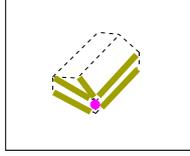
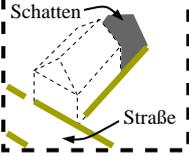
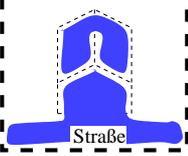
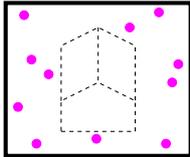
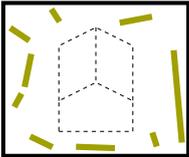
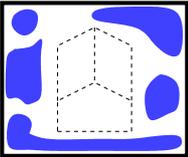
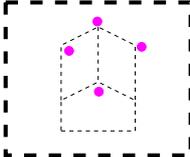
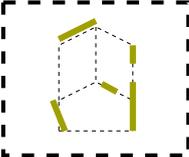
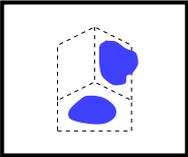
1. Alle Bildmerkmale, die nicht zum gesuchten Objekt gehören, können auch als Störmerkmale interpretiert werden.
2. Bildmerkmale können stärker als erwartet von den Modellmerkmalen durch Verschiebung, Verzerrung und Verformung abweichen.

Abbildung 4.2 zeigt eine Übersicht aller Fehlerfälle, die darstellt, wie sich die Bildstörungen letztlich in den zur Bildinterpretation verwendeten extrahierten Merkmalen ausprägen. Dabei sind die Fehler, die bereits von der im vorigen Kapitel vorgestellten Constraint-Modellierung gehandhabt werden können, durch eine dicke Umrandung hervorgehoben, und die Fälle, die zumindest teilweise davon aufgefangen werden, gestrichelt umrandet. Da mit allen aufgeführten Fehlerfällen in den Bilddaten gerechnet werden muß, wird im folgenden untersucht, wie sich die einzelnen Fehlerarten auf die Erfüllbarkeit der unterschiedlichen Constraints auswirken. Tabelle 4.1 bezieht die Fehlerfälle auf die Constraints, wobei der Eintrag 'X' bedeutet, daß der Constraint durch den entsprechenden Fehler betroffen sein kann.

Constraint	Fehlerfälle																					
	1			2			3			4			5			6			7			
	Störmerkmal			Fehlen von			part. Fehlen			Fragment.			Verschmelzen			Fremdmerkmal.			Verzerrung			
	P	L	F	P	L	F	P	L	F	P	L	F	P	L	F	P	L	F	P	L	F	
MNG(Linie,Punkt)				X	X		X			X	X								X	X		
MNG(Fläche,Linie)					X	X		X	X		X	X								X	X	
MNG(Fläche,Punkt)				X		X			X	X	X								X		X	
Linienparallelität					X																X	
Kollinearität				X	X														X	X		
Same_Side_Lines					X									X							X	
≠(Punkte)				X									X									
≠(Linien)					X									X								
≠(Flächen)						X									X							
Linienwinkel					X																X	

**Tabelle 4.1:** Auswirkungen der sieben Fehlerarten aus Abbildung 4.2 auf die Erfüllbarkeit der einzelnen Constraints. MNG bezeichnet die Merkmalsnachbarschafts- und  $\neq$  die Verschiedenheitsconstraints.

Die erste und sechste Spalte der Tabelle sind besonders signifikant, denn weder Fehler der ersten noch der sechsten Kategorie führen zur Verletzung von Constraints. Dies ist insofern leicht zu erklären, weil die erste Kategorie für Störmerkmale und die sechste für nicht zu dem gesuchten Objekt gehörende Merkmale steht. Vielmehr handelt es sich um Bildobjekte, die gerade nicht den

	(P) Punkte	(L) Linien	(F) Flächen	
Störmerkmale			nicht möglich	(1)
Fehlen von				(2)
partielles Fehlen von	nicht möglich			(3)
Aufspaltung / Fragmentierung				(4)
Verschmelzen von				(5)
unbeteiligte Merkmale				(6)
Verschiebung / Verzerrung				(7)

**Abbildung 4.2:** Beispiele für beobachtbare Störungen in realen Bildern in bezug auf ein zu erkennendes Objekt (gestrichelt hinterlegt). Die Bildfehler lassen sich in insgesamt sieben Kategorien einteilen, von denen die ersten fünf allgemeine Probleme der Bildsegmentierung und Merkmalsextraktion darstellen (siehe [Fuc97]). Die beiden letzten Reihen sind spezifisch für das Problem der Objekterkennung, weil die gezeigten Situationen erst im Kontext eines auf dem Bildmodell aufsetzenden Objektmodells auftreten. Die dick umrandeten Fehlerfälle werden bereits vollständig durch die im vorigen Kapitel vorgestellte Modellierung aufgefangen, und die gestrichelt umrandeten Fälle zumindest teilweise (siehe Text).

Merkmalsvariablen zugewiesen werden sollen, und die somit auch keine Constraints des Modells verletzen können.

Im Gegensatz dazu betrifft die Unbeobachtbarkeit (Spalte 2 in Tabelle 4.1), die sich durch das vollständige Fehlen von Merkmalen ausdrückt, alle Constraints. Wenn ein Merkmal des Modells nicht beobachtbar ist, können offensichtlich auch keine Relationen zu anderen Modellmerkmalen bestehen. Die Constraints werden in diesen Fällen zwar nicht direkt verletzt, jedoch können sie auch nicht erfüllt werden.

Fehler der siebten Kategorie bezeichnen die Verschiebung oder Verzerrung der Bildmerkmale in bezug auf die erwartete Ausprägung der Modellmerkmale. Sie entstehen aufgrund der Unschärfe und Unsicherheit und betreffen alle topologischen und geometrischen Constraints. Die Fehler werden zum Teil durch die Schwellwerte der Constraints aufgefangen, jedoch können die Abweichungen auch so groß werden, daß die Schwellwerte überschritten und die entsprechenden Constraints verletzt werden.

Die Fehler der dritten und vierten Kategorie, die im wesentlichen Effekte der Fragmentierung darstellen, wirken sich nur auf die topologischen Merkmalsnachbarschafts-Constraints aus. Die geometrischen Constraints sind davon nicht betroffen. Die weiteren Auswirkungen und ein entsprechendes Konzept zur Handhabung dieser Auswirkungen wurden bereits in Abschnitt 3.5 vorgestellt.

Das Verschmelzen von Merkmalen (Fehlerkategorie 5) kann sich zunächst auf die Verschiedenheits-Constraints auswirken, weil es vorkommen kann, daß zwei Modellmerkmalen dasselbe Bildmerkmal zugeordnet werden muß. Dies kann beispielsweise eintreten, wenn zwei Dachflächen aufgrund eines niedrigen Kontrasts zu einer Fläche „zusammenlaufen“. Verschmilzt hingegen ein zu dem gesuchten Objekt gehörendes Merkmal mit einem unbeteiligten Merkmal, wie dies beispielsweise in Abbildung 4.2 im Fall 5b bei der Bodenkante und der Schattenkante illustriert wurde, werden die Verschiedenheitsconstraints nicht betroffen. In diesem Fall wird jedoch ein `Same_Side_Lines`-Constraint verletzt.

## 4.2 Überbeschränkte Constraint-Probleme

Wie im vorangegangenen Abschnitt erläutert wurde, führen bestimmte Arten von Bildstörungen letztlich dazu, daß Constraints verletzt werden. Das bedeutet, daß das zu einer Gebäudehypothese hergeleitete Constraint Satisfaction Problem keine Lösung besitzt, und die Suche nach der Gebäudeansicht damit fehlschlägt.

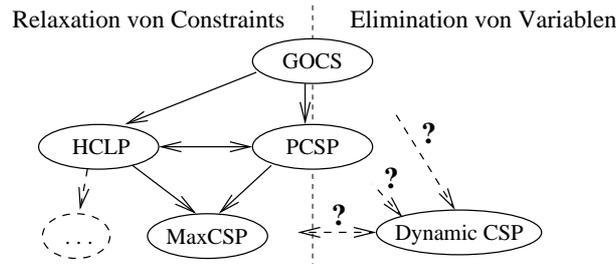
Erst in den letzten Jahren wurde diese Problematik im Forschungsbereich der Constraint-Techniken aufgegriffen und unter dem Oberbegriff *überbeschränkter Systeme* (engl.: *Over-Constrained Systems*, OCS) untersucht. Ein überbeschränktes Constraint-System ist ein CSP, das keine Lösung besitzt, weil einige Constraints in Konflikt mit anderen stehen. Freuder führt in [Fre89] aus, daß es vier Möglichkeiten zur Entschärfung dieses Konflikts gibt, welche die Berechnung einer (suboptimalen) Lösung erlauben:

1. Vergrößerung des Wertebereichs einer Variablen um geeignete Werte.
2. Vergrößerung der Menge der Relationentupel eines Constraints, die angibt welche Wertekombinationen der Variablen zueinander kompatibel sind.
3. Eliminierung von Variablen mit ihren inzidenten Constraints.
4. Eliminierung von Constraints.

Für jedes konkrete CSP gibt es generell verschiedene Möglichkeiten der Relaxation. Wenn beispielsweise ein Constraint  $c_1$  den beiden Constraints  $c_2$  und  $c_3$  widerspricht, könnte der Konflikt

aufgelöst werden, indem entweder  $c_1$  oder  $c_2$  und  $c_3$  entfernt werden. Um die im Sinne der Anwendung beste Lösung auswählen zu können, benötigen OCS-Verfahren eine Bewertungsfunktion, die eine Ordnung auf den verschiedenen möglichen Relaxationen definiert.

Das Buch [JFM96] enthält eine Sammlung der Aufsätze, die die wichtigsten Konzepte zur Behandlung überbeschränkter Systeme erläutert. Keines der bislang vorgestellten Verfahren entstand jedoch im Kontext eines Objekterkennungsproblems; sie wurden für die Anwendungsgebiete Scheduling, Konfiguration, Zeitplanung und graphische Benutzeroberflächen entwickelt.



**Abbildung 4.3:** Taxonomie der Methoden zur Behandlung überbeschränkter Constraint-Probleme. Das einzige Verfahren, das die Eliminierung von Variablen zuläßt, Dynamic CSP, wurde bislang noch nicht in ein übergeordnetes Schema eingegliedert.

Zur Verdeutlichung der Beziehungen dieser Konzepte zueinander werden sie in einer Taxonomie angeordnet, die in Abbildung 4.3 dargestellt ist. Die Verfahren lassen sich dabei hauptsächlich in die zwei Kategorien 1) Relaxation von Variablen und 2) Eliminierung von Variablen einordnen. HCLP kann als generelles Konzept zur Relaxation von Constraints angesehen werden, weil sich sowohl MaxCSP als auch andere (hier nicht weiter diskutierte) Ansätze als Spezialfälle von HCLP darstellen lassen. In den folgenden Unterabschnitten werden die wichtigsten Konzepte vorgestellt und ihre Eignung zur Verwendung bei der Gebäudeerkennung diskutiert.

#### 4.2.1 MaxCSP

*Maximum Constraint Satisfaction* (MaxCSP) wurde von Freuder und Wallace in [FW92] vorgestellt und stellt die einfachste und bislang bekannteste Methode zur Handhabung überbeschränkter Constraint-Probleme dar. Wenn es für ein CSP keine Lösung gibt, werden solange Constraints relaxiert, bis eine Lösung berechnet werden kann. Die zugrundeliegende Metrik zählt die Anzahl erfüllter Constraints eines CSP. Die beste Lösung ist diejenige, die diese Anzahl maximiert. In [FW96] wird die Anwendbarkeit und Erweiterung klassischer Constraint-Solving-Techniken in bezug auf MaxCSP untersucht und empirisch evaluiert. Dabei hat sich gezeigt, daß insbesondere die First-Fail-Heuristik weiterhin zur effizienten Lösungsbestimmung verwendet werden kann.

Praktische Tests zur Verwendbarkeit von MaxCSP zur Gebäudeerkennung haben gezeigt, daß MaxCSP zwar einen Teil der Fehlerfälle abfangen kann, jedoch Probleme im Fall von unbeobachtbaren Merkmalen bekommt (siehe [FKL97]).

#### 4.2.2 HCLP

*Hierarchical Constraint Logic Programming* (HCLP) wurde von Borning und Wilson entwickelt (vgl. [WB89, Wil93]) und löst Konflikte ebenfalls durch die Relaxation von Constraints auf. Dabei wird jeder Constraint mit einer Präferenz gewichtet, die aus einer endlichen Menge von Prioritätsstufen gewählt werden kann. Constraints der höchsten Prioritätsstufe müssen immer erfüllt sein, Constraints auf niedrigeren Stufen dürfen auch verletzt werden. Die lexikographische Ordnung der Hierarchiestufen sorgt dafür, daß die Erfüllung eines höherstufigen Constraints immer höher bewertet wird als die Erfüllung beliebig vieler Constraints der darunterliegenden Ebenen.

Die Anwendbarkeit von HCLP für die Gebäudeerkennung wurde in der Diplomarbeit von Bosch ([Bos97]) untersucht. Die Arbeit setzt auf dem in [KPC96] gezeigten Konzept zum Einsatz von CLP zur Gebäudeerkennung auf, bei dem die Verletzung von Constraints bis dahin nicht berücksichtigt wurde. Es konnte analog zur oben beschriebenen Anwendung von MaxCSP zur Gebäudeerkennung gezeigt werden, daß die Relaxation von Constraints zwar das Problem der Unlösbarkeit des CSP entschärft, jedoch im Falle der Unbeobachtbarkeit von Bildmerkmalen inkorrekte Zuordnungen durchgeführt werden. Ferner stellte sich die Gewichtung der Constraints durch die Zuordnung zu einer Hierarchiestufe als problematisch heraus. Zwar wurden Constraints, die nach empirischer Beobachtung häufiger verletzt sind, in niedrigere Prioritätsstufen eingeteilt, jedoch führt die lexikographische Ordnung auf den Anzahlen erfüllter Constraints der einzelnen Hierarchieebenen generell zu einer Überbewertung der höherstufigen Constraints.

### 4.2.3 Dynamic CSP

*Dynamic CSP* wurde von Mittal und Falkenhainer in dem Aufsatz [MF90] vorgestellt, und ist das einzige Verfahren, daß die Relaxation überbeschränkter Systeme durch die Eliminierung von Variablen realisiert. In Dynamic CSP werden Variablen durch besondere Aktivierungs-Constraints ein- oder ausgeschaltet. Die Umstände, unter denen diese Constraints Variablen aktivieren bzw. deaktivieren, können von der Wertebelegung oder dem Aktivierungszustand anderer Variablen abhängen. Constraints der erstgenannten Sorte haben die Form  $P(v_1, \dots, v_n) \rightarrow active : v_j$ , wobei  $j > n$ , und  $P$  ein beliebiges  $n$ -stelliges Prädikat über den Variablen  $v_1, \dots, v_n$  darstellt. Die zweite Sorte von Constraints hat die Form  $active : v_1 \wedge \dots \wedge active : v_n \rightarrow active : v_j$ , und besagt, daß wenn die Variablen  $v_1, \dots, v_n$  aktiv sind, auch  $v_j$  aktiv sein muß. Die Deaktivierung einer Variablen beinhaltet implizit auch die Deaktivierung aller zu ihr inzidenten Constraints. Die Definition einer Grundmenge von Variablen, die immer aktiv sein müssen und damit Teil jeder Lösung sind, erlaubt ferner die Sicherstellung bestimmter Minimalkonfigurationen.

### 4.2.4 Die generischen Konzepte PCSP und GOCS

*Partial Constraint Satisfaction* (PCSP) ist ein von Freuder und Wallace entwickeltes Rahmenkonzept für die Handhabung überbeschränkter Systeme (siehe [Fre89, FW92, FW96]). Dabei wird jede Modifikation des ursprünglichen CSP als Veränderung des Problemraums  $PS$  betrachtet. Das Konzept ist in dem Sinne generell, daß jede Abstandsfunktion, die die Abweichung von  $PS'$  zum ursprünglichen Problemraum  $PS$  quantifiziert und dabei die Eigenschaften einer Metrik aufweist, zur Konstruktion der Bewertungsfunktion verwendet werden kann. In [FW96] werden verschiedene Metriken vorgeschlagen, wobei nur MaxCSP im Detail untersucht wurde. Leider berücksichtigt keine der Metriken die Eliminierung von Variablen.

In seiner Dissertation stellt Jampel das *General Framework for Over-Constrained Systems* (GOCS) als eine Abstraktion von HCLP und PCSP vor (siehe [Jam96] und vgl. Abbildung 4.3 auf Seite 72). Zunächst wird dabei gezeigt, daß alle in HCLP formulierten Constraint-Probleme nach PCSP transformiert werden können und umgekehrt, wenn die Bewertungsfunktion auf der Basis von Relationentupelergänzungen ausgedrückt wird (siehe oben, Abschnitt 4.2: zweite Methode der Relaxation überbeschränkter Systeme). Der Fokus von GOCS liegt dabei auf dem Aspekt der Kompositionalität, der eine wichtige Rolle für die operationale Umsetzung von OCS-Verfahren darstellt. Die Arbeit konzentriert sich ebenfalls ausschließlich auf die Relaxation von Constraints; die Beziehung zu Dynamic CSP und der Eliminierung von Variablen wurde als offene Frage im Ausblick der Arbeit aufgeworfen.

### 4.2.5 Diskussion der Konzepte

Alle oben vorgestellten Konzepte zur Handhabung überbeschränkter Constraint-Probleme sind nicht in der Lage, die in Verbindung mit der Objekterkennung auftretenden Probleme adäquat

zu berücksichtigen. Das Dilemma besteht darin, daß Constraints aus zwei verschiedenen Gründen nicht erfüllbar sein können, und sie in beiden Fällen zur Auflösung des Konflikts relaxiert werden müssen: Erstens kann die einem Constraint zugrundeliegende Relation verletzt sein. Dies ist beispielsweise der Fall, wenn zwei Linienvariablen des Modells Bildmerkmale zugewiesen bekommen, die zwar laut Modell parallel sein müßten, es aber aufgrund der zu großen Linienwinkelabweichung nicht sind. Zweitens kann ein vom Modell erwartetes Merkmal im Bild unbeobachtbar sein, was bedeutet, daß zu diesem nicht vorhandenen Merkmal auch keine Relationen bestehen können, und deshalb die zu dem entsprechenden Modellmerkmal inzidenten Constraints ebenfalls relaxiert werden müssen.

Nun verbinden die Verfahren, die zur Auflösung der Konflikte nur die Relaxation von Constraints berücksichtigen wie MaxCSP und HCLP, mit jeder Constraint-Relaxation genau einen Kostenwert. Das bedeutet aber wiederum, daß nicht zwischen den beiden oben genannten Gründen für die Relaxation der Constraints unterschieden werden kann. Im Fall der Unbeobachtbarkeit würden aufgrund der Relaxation aller inzidenten Constraints sehr große Kosten entstehen, die nicht unbedingt mit der Wichtigkeit der Variable bzw. des Modellmerkmals korrespondieren. Im Gegenteil: manchmal wäre es besser anzunehmen, daß bestimmte Merkmale nicht beobachtbar sind, um die Anzahl der „echten“ Relationenverletzungen zu minimieren. Ein weiteres Problem, das eng damit zusammenhängt, ist, daß bei den beschriebenen OCS-Verfahren nach wie vor jeder Variablen ein Wert zugewiesen werden muß. Es gibt also gar keine Möglichkeit, die Unbeobachtbarkeit eines Merkmals in der entsprechenden Variable explizit zu machen, weil ihr immer ein Bildmerkmal zugewiesen werden muß. Dieses könnte zudem zufällig einige der Constraints erfüllen und damit den Wert der Bewertungsfunktion weiter verfälschen.

Ähnliches gilt für die Rahmenkonzepte PCSP und GOCS. Obwohl diese eigentlich nicht auf die Relaxation von Constraints beschränkt sind, reduzieren die Autoren jegliche Problemrelaxation eines CSP auf die Ergänzung von kompatiblen Relationentupeln (vgl. Abschnitt 4.2, Relaxationsmethode Nr. 2). Dies ermöglicht zwar die Bestimmung einer Lösung, führt aber zu den gerade genannten Problemen bei der Bewertung von Constraint-Relaxationen. Darüberhinaus wirkt die Ergänzung um neue kompatible Relationentupel während der Suche operationale Probleme auf: Konsistenztechniken basieren auf der Annahme, daß erstens die Wertebereiche von Variablen mit fortschreitender Suche monoton kleiner werden, und zweitens die Menge der den Constraints zugrundeliegenden kompatiblen Relationentupel über die gesamte Suche konstant bleibt.

Dynamic CSP kann zwar die Unbeobachtbarkeit von Merkmalen durch die Deaktivierung bzw. Eliminierung der entsprechenden Variablen und ihrer inzidenten Constraints ausdrücken, jedoch müssen die zwischen aktivierten Variablen bestehenden Constraints alle erfüllt sein. Es können also keine einfachen Verletzungen von Relationen berücksichtigt werden.

### 4.3 Erweiterung der Constraint-Modellierung

Wie die Diskussion im vorigen Abschnitt ergeben hat, eignet sich keines der bestehenden OCS-Verfahren für die Objekterkennung, weil sie entweder nur die Verletzung von Constraints behandeln oder nur die Eliminierung von Variablen ermöglichen. In diesem Abschnitt wird deshalb ein neues Konzept vorgestellt, das eine Integration beider Fälle erlaubt, und damit die Effekte der Unbeobachtbarkeit und der Unschärfe/Unsicherheit explizit unterscheiden kann (siehe auch [Kol98, KPC00, FKL<sup>+</sup>98]).

#### 4.3.1 Relaxation von Constraints

Die wesentliche Idee des Konzepts besteht darin, die Verletzung von Constraints abzufangen, so daß das CSP weiterhin lösbar bleibt. Der Status jedes einzelnen Constraints muß zudem explizit repräsentiert werden, damit er in einer Bewertungsfunktion zur Bestimmung der besten

Lösung berücksichtigt werden kann. Es wird dazu jeder Constraint  $c(v_1, \dots, v_n) \in C$  des gegebenen CSP( $V, D, C$ ) mit den Variablen  $V$  und ihren assoziierten Wertebereichen  $D$  um eine dreiwertige Variable  $b \in \{-1, 0, 1\}$  erweitert zu einem Constraint  $c'(v_1, \dots, v_n, b)$  mit

$$\begin{aligned} c'(v_1, \dots, v_n, b) \Leftrightarrow & (b = 1 \wedge c(v_1, \dots, v_n)) \vee \\ & (b = -1 \wedge \neg c(v_1, \dots, v_n)) \vee \\ & (b = 0) \end{aligned} \quad (4.1)$$

Die Variable  $b$  stellt dabei sowohl eine Statusvariable als auch einen Schalter dar: Wenn auf der einen Seite der ursprüngliche Constraint  $c$  als definitiv erfüllt (entailed) erkannt wird, erhält  $b$  den Wert 1. Wird auf der anderen Seite  $b = 1$  gesetzt, muß  $c$  erfüllt werden, damit  $c'$  erfüllt wird, d.h. der ursprüngliche Constraint  $c$  wird dadurch erzwungen. Analog verhält es sich bei Feststellung des Fehlschlagens von  $c$ : in diesem Fall wird  $b$  auf den Wert  $-1$  gesetzt. Das Setzen von  $b = -1$  wiederum erzwingt die Negation von  $c$ , d.h. der ursprüngliche Constraint  $c$  muß verletzt sein, damit  $c'$  erfüllt wird.

Mit den Werten  $b = -1$  bzw.  $b = 1$  läßt sich also bereits die Verletzung bzw. Erfüllung eines Constraints  $c$  explizit darstellen. Auf den dritten Fall ( $b = 0$ ) wird nun im folgenden eingegangen.

### 4.3.2 Eliminierung von Variablen durch Fehlwerte

Bei der Modellierung der Unbeobachtbarkeit von Bildmerkmalen besteht die Idee darin, die Wertebereiche der Variablen jeweils um einen ausgezeichneten Fehlwert (*Wildcard*, Notation:  $*$ ) zu erweitern. Wenn ein Merkmal im Bild nicht beobachtbar ist, wird der Variablen dann anstelle eines extrahierten Bildmerkmals der Fehlwert zugewiesen. Da ein nicht vorhandenes Objekt auch keine Relationen besitzen kann, müssen im Fall der Fehlwertzuweisung alle Constraints, die zu der entsprechenden Variablen inzident sind, relaxiert werden:

$$\forall v_i \in V : (v_i = * \Leftrightarrow \forall c'_j(\dots, v_i, \dots, b_j) \in C' : b_j = 0) \quad (4.2)$$

Diese Bedingung verbindet die Statusvariablen mit den eigentlichen Variablen des CSP: Wenn einer Variablen  $v_i$  der Fehlwert zugewiesen wird, müssen alle Statusvariablen  $b_j$  der inzidenten Constraints  $c'_j$  den Wert 0 erhalten. Wie der Definition von  $c'$  in Gleichung 4.1 zu entnehmen ist, bedeutet dies die Relaxation der ursprünglichen Constraints  $c_j$ , weil im Fall von  $b_j = 0$  der transformierte Constraint  $c'_j$  unabhängig von  $c_j$  immer erfüllt ist.

Der transformierte Constraint  $c'$  realisiert also eine Kapselung des ursprünglichen Constraints, und ermöglicht damit eine kontrollierte Relaxation des ursprünglichen Constraints  $c$ . Zusammengefaßt haben die Werte der Indikator-/Steuervariablen  $b$  somit die folgende Bedeutung:

$$\begin{aligned} b = -1 & : \text{Constraint wurde verletzt} \\ b = 0 & : \text{Constraint ist unbeobachtbar} \\ b = 1 & : \text{Constraint wurde erfüllt} \end{aligned}$$

Die Wahrheitstabelle für die Erfüllung von  $c'(\dots, b)$  in Abhängigkeit von der Wertebelegung von  $b$  und der Erfüllbarkeit von  $c$  ergibt sich zu

$c'(v_1, \dots, v_n, b)$	$b = -1$	$b = 0$	$b = 1$
$c(v_1, \dots, v_n) = 0$	1	1	0
$c(v_1, \dots, v_n) = 1$	0	1	1

Widersprüche bei der Lösung des transformierten CSP können also nur dann auftreten, wenn die durch  $b = 1$  ausgedrückte Forderung nach der Erfüllung von  $c$  oder die durch  $b = -1$  ausgedrückte Forderung nach dem Fehlschlagen von  $c$  nicht erfüllt werden.

### 4.3.3 Inferenzregeln

Die vorgestellte Kapselung der ursprünglichen Constraints  $c_i$  durch  $c'_i$  sorgt dafür, daß die Verletzung von  $c_i$  nicht notwendigerweise auch ein Fehlschlagen des gesamten CSPs verursacht. Vielmehr wird in einem solchen Fall die Statusvariable  $b_i$  auf einen entsprechenden Wert gesetzt (bei Verletzung von  $c_i$  also auf  $-1$ ), und  $c'_i$  dadurch letztlich erfüllt.

Eine Möglichkeit der Realisierung der erweiterten Constraints  $c'_i$  besteht nun darin, die in Gleichung 4.1 gezeigte logische Formel als Prolog-Klausel zu formulieren, und diese dann unter Verwendung des in Abschnitt 3.4.1 vorgestellten generischen Constraint-Handlers einfach als neuen Constraint zu definieren, der dann mittels der Standard-Forward-Checking- oder Look-Ahead-Inferenzregel ausgewertet wird.

Das Problem bei Verwendung der Look-Ahead-Inferenzregel wäre aber, daß sie viele vergebliche Konsistenzüberprüfungen durchführen würde, weil für die erweiterten Constraints in vielen Fällen zunächst keine Wertebereichseinschränkungen möglich sind. Dies läßt sich an der Bedeutung der Statusvariablen  $b_i$  eines erweiterten Constraints  $c'_i$  verdeutlichen: Sind z.B. noch die beiden Werte  $-1$  und  $1$  im Wertebereich von  $b$  enthalten, bedeutet das, daß der ursprüngliche Constraint  $c_i$  wahr oder auch falsch sein kann. Es können deshalb noch keine Einschränkungen anhand von  $c_i$  gemacht werden. Ein ähnlicher Fall liegt vor, wenn neben einem anderen Wert noch der Wert  $0$  im Wertebereich von  $b_i$  enthalten ist. Auch hier dürfen keine Einschränkungen vorgenommen werden, denn sollte  $b = 0$  gesetzt werden, heißt das, daß der Constraint eigentlich gar nicht existiert und deswegen niemals irgendwelche Einschränkungen hätte vornehmen können.

Im Gegensatz dazu würde die Verwendung der Forward-Checking-Inferenzregel dazu führen, daß der erweiterte Constraint oftmals später als nötig ausgewertet und damit den Suchraum weniger stark beschränken würde, weil Forward-Checking-Constraints erst geweckt werden, wenn alle Variablen bis auf eine an einen eindeutigen Wert gebunden sind. Zur Verdeutlichung: Wenn der Statusvariablen  $b_i$  der Wert  $1$  zugewiesen wird, muß der ursprüngliche Constraint  $c_i$  erfüllt sein, damit auch  $c'_i$  erfüllt ist.  $c'_i$  könnte also zu diesem Zeitpunkt — unabhängig von der Größe der Wertebereiche der anderen Variablen — durch den ursprünglichen Constraint  $c_i$  ersetzt werden, der u.U. sofort Wertebereichseinschränkungen auf seinen Variablen vornehmen kann.

Die Realisierung der erweiterten Constraints mittels Look-Ahead-Inferenzregel versucht zu oft und zu früh, Bereichseinschränkungen herbeizuführen, wohingegen die Forward-Checking-Inferenzregel in vielen Fällen zu spät „feuert“. Im folgenden werden deshalb spezielle Inferenzregeln vorgestellt, die die Bedeutung der Statusvariablen  $b_i$  explizit berücksichtigen und ausnutzen. Zur formalen Darstellung dieser Regeln und ihrer operationalen Semantik wird ein Transitionssystem in Anlehnung an van Hentenryck et. al. ([vHSD92]) verwendet, das wie folgt definiert ist:

**Definition 4.1 (Transitionssystem)**

Ein Transitionssystem ist ein Tripel  $\langle \Gamma, T, \mapsto \rangle$ , wobei  $\Gamma$  eine Menge von Konfigurationen,  $T \subseteq \Gamma$  die Menge der Terminalkonfigurationen und  $\mapsto \subseteq \Gamma \times \Gamma$  die Übergangsrelation bezeichnet. Dabei muß die Übergangsrelation die folgende Eigenschaft erfüllen:

$$\forall \gamma \in T, \quad \forall \gamma' \in \Gamma : \quad \gamma \not\mapsto \gamma'. \quad \square$$

Die Konfigurationen des Transitionssystems sind die Berechnungszustände  $\langle G \square \sigma \rangle$ , wobei  $G$  die noch zu beweisenden Ziele und  $\sigma$  den aktuellen Stand des Constraint-Speichers bezeichnet (vgl. Abschnitt 3.2.2). Terminalkonfigurationen sind erfolgreiche Berechnungszustände, die hier dadurch charakterisiert sind, daß der Zielteil leer ist. Die Lösungen eines CSP werden gemäß Definition 3.2 auf Seite 36 durch die Teilmenge von  $T$  gebildet, bei der der Constraint-Speicher für jede Variable  $V_1, \dots, V_n$  genau einen Constraint der Form  $V_i = \text{Wert}_i$  enthält.

Eine Transition  $\gamma \mapsto \gamma'$  kann interpretiert werden als: Konfiguration  $\gamma$  wird (nichtdeterministisch<sup>1</sup>) reduziert auf  $\gamma'$ . Die möglichen Transitionen werden durch Inferenzregeln beschrieben, die folgendes Format besitzen:

$$\frac{\langle \text{Bedingung } 1 \rangle \quad \vdots \quad \langle \text{Bedingung } m \rangle}{\gamma \mapsto \gamma'}$$

Diese Inferenzregeln besagen, daß die Transition  $\gamma \mapsto \gamma'$  stattfinden kann, wenn alle aufgezählten Vorbedingungen erfüllt sind.

Es werden im folgenden nur die Inferenzregeln vorgestellt, die die erweiterten Constraints  $c'$  auf die ursprünglichen Constraints  $c$  bzw.  $\neg c$  reduzieren, welche danach mittels bekannter Standard-Inferenzregeln weiter ausgewertet werden. Für die formale Definition der allgemeinen CLP-Inferenzregeln (Zielreduktion bzw. Entfalten, Constraint-Lösen bzw. Vereinfachung, und Scheitern) sei auf [vHSD92] und [FA97b] verwiesen. Die Notation wurde bereits in Abschnitt 3.2 beschrieben. Die speziell für Constraints über endlichen Wertebereichen definierten Forward-Checking- und Look-Ahead-Inferenzregeln sind in [vH87, vH89] erläutert.

Erkennung der definitiven Erfüllung (Entailment) eines ursprünglichen Constraints  $c_i$  und Folgerung der Bindung von  $b_i$  an 1:

$$\frac{\sigma \models c_i(v_1, \dots, v_n) \quad c'_i(v_1, \dots, v_n, b_i) \in \sigma}{\langle G \sqcap \sigma \rangle \mapsto \langle b_i = 1, G \sqcap \sigma \setminus \{c'_i\} \rangle} \quad (4.3)$$

Erkennung der Unerfüllbarkeit (Contradiction) eines ursprünglichen Constraints  $c_i$  und Folgerung der Bindung von  $b_i$  an  $-1$ :

$$\frac{\sigma \wedge c_i(v_1, \dots, v_n) \models \perp \quad c'_i(v_1, \dots, v_n, b_i) \in \sigma}{\langle G \sqcap \sigma \rangle \mapsto \langle b_i = -1, G \sqcap \sigma \setminus \{c'_i\} \rangle} \quad (4.4)$$

Zuweisung des Fehlwertes (Wildcard) zu einer Variablen  $v_j$  und Folgerung der Bindung der  $b_i$  aller inzidenten Constraints  $c'_i$  an 0:

$$\frac{\sigma \models v_j = * \quad c'_i(\dots, v_j, \dots, b_i) \in \sigma}{\langle G \sqcap \sigma \rangle \mapsto \langle b_i = 0, G \sqcap \sigma \setminus \{c'_i\} \rangle} \quad (4.5)$$

Erzwingen eines ursprünglichen Constraints  $c_i$  durch die Bindung von  $b_i$  an den Wert 1:

$$\frac{\sigma \models b_i = 1 \quad c'_i(v_1, \dots, v_n, b_i) \in \sigma}{\langle G \sqcap \sigma \rangle \mapsto \langle v_1 \neq *, \dots, v_n \neq *, c_i(v_1, \dots, v_n), G \sqcap \sigma \setminus \{c'_i\} \rangle} \quad (4.6)$$

Erzwingen des Fehlschlagens eines ursprünglichen Constraints  $c_i$  durch die Bindung von  $b_i$  an den Wert  $-1$ :

$$\frac{\sigma \models b_i = -1 \quad c'_i(v_1, \dots, v_n, b_i) \in \sigma}{\langle G \sqcap \sigma \rangle \mapsto \langle v_1 \neq *, \dots, v_n \neq *, \neg c_i(v_1, \dots, v_n), G \sqcap \sigma \setminus \{c'_i\} \rangle} \quad (4.7)$$

<sup>1</sup>Nichtdeterministische Reduktion bedeutet, daß eine der möglichen Reduktionen ausgewählt und durchgeführt wird. Waren mehrere Reduktionen möglich, wird im Kontext von CLP bei einem späteren Fehlschlagen durch das Backtracking an diese Stelle zurückgekehrt und eine andere Wahl getroffen.

Unbedingte Relaxation eines ursprünglichen Constraints  $c_i$  durch die Bindung von  $b_i$  an den Wert 0:

$$\frac{\begin{array}{l} \sigma \models b_i = 0 \\ c'_i(v_1, \dots, v_n, b_i) \in \sigma \end{array}}{\langle G \sqcap \sigma \rangle \mapsto \langle (v_1 = * \vee \dots \vee v_n = *), G \sqcap \sigma \setminus \{c'_i\} \rangle} \quad (4.8)$$

Diese Inferenzregeln können mit den Sprachmitteln leistungsfähiger CLP-Systeme wie ECLIPSE oder SICStus PROLOG unmittelbar realisiert werden. Für die praktische Evaluation (siehe unten) mußte der in Kapitel 3.4.1 beschriebene generische Constraint-Handler nur um wenige Programmzeilen (< 40) erweitert werden.

#### 4.3.4 Optimierung mittels Branch & Bound

Zur Bestimmung der besten Lösung muß zunächst die Bewertungsfunktion definiert werden. Diese Funktion ergibt sich hier aus der Summierung der Statusvariablen aller Constraints, denn ein erfüllter Constraint ( $b = 1$ ) trägt zur Gesamtzuordnung bei, wohingegen ein verletzter Constraint ( $b = -1$ ) dieser widerspricht. Sind Constraints aufgrund der Unbeobachtbarkeit eines Merkmals ebenfalls nicht beobachtbar ( $b = 0$ ), können sie weder zur Gesamtzuordnung beitragen noch ihr widersprechen. Wie im folgenden Kapitel 5 noch erläutert wird, hat diese Bewertung eine informationstheoretische Basis. Wenn  $C'$  die Menge der transformierten Constraints ist, läßt sich die Bewertungsfunktion  $f(C')$  folgendermaßen darstellen:

$$f(C') = \sum_{i=1}^{|C'|} b_i \text{ mit } c'_i(\dots, b_i) \in C' \quad (4.9)$$

Diese Funktion wird nun mit Hilfe eines Branch & Bound-Verfahrens maximiert (vgl. [AHU87]). Dabei gibt eine untere Schranke  $\lambda$  den minimalen Wert an, der von  $f(C')$  erreicht werden muß. Sobald bei der Suche eine Lösung gefunden wurde, die größer als die aktuelle untere Schranke ist, wird diese auf den neuen, größeren Lösungswert von  $f(C')$  gesetzt und die Suche erneut gestartet. Dies wird solange wiederholt, bis sich für den Schrankenwert  $\lambda$  keine Lösung mehr finden läßt mit  $f(C') > \lambda$ .

Die meisten CLP-Systeme unterstützen das Branch & Bound-Verfahren durch eingebaute spezielle Prädikate (vgl. [DvHS<sup>+</sup>88, vH89, PM95, WNS97]). Der Vorteil dieser Einbettung liegt darin, daß von der unteren Schranke und anderen Bedingungen auch Rückschlüsse auf die Belegung der einzelnen Variablen der Bewertungsfunktion getroffen werden können. Während der Suche ist der Wertebereich von  $f(C')$  durch Bindungen verschiedener  $b_i$  an  $-1$ ,  $0$  oder  $1$  auf ein noch mögliches Intervall  $\{min, \dots, max\}$  festgelegt. Die untere Schranke  $\lambda$  wird dem CSP explizit als Constraint  $f(C') > \lambda$  hinzugefügt. Dies führt dazu, daß wenn  $max \leq \lambda$  wird, sofort der Suchzweig abgeschnitten und ein Backtracking ausgelöst wird. Darüberhinaus kann der Constraint-Interpreter aus dem Intervall  $\{min, \dots, max\}$  für  $f(C')$  und dem Constraint für die untere Schranke in bestimmten Situationen eindeutige Belegungen für die noch ungebundenen  $b_i$  schlussfolgern und erzwingen. So kann u.U. festgestellt werden, daß zur Überschreitung der unteren Schranke alle noch ungebundenen  $b$ -Variablen den Wert 1 erhalten müssen. In diesem Fall würden durch das Setzen der betreffenden  $b$ -Variablen aufgrund der oben dargestellten Inferenzregeln damit die korrespondierenden Constraints  $c$  erzwungen.

Das beschriebene Suchverfahren führt einen vollständigen Durchlauf des Suchraumes durch. Das bedeutet, daß auf der einen Seite zwar garantiert das globale Optimum gefunden wird, aber auf der anderen Seite der Suchaufwand erheblich sein kann. Im folgenden Abschnitt wird deshalb näher untersucht, wie der Suchraum bereits vor der Suche beschränkt werden kann.

### 4.3.5 A priori und aktive Beschränkungen des Suchraumes

Die Erweiterung der Constraints berührt zwar die Problemkomplexität der Gebäudeerkennung selber nicht, jedoch wird die Komplexität der Implementierung durch die Einführung der Statusvariablen  $b_1, \dots, b_n$  um den Faktor  $3^n$  vergrößert, weil zusätzlich zu der Zuweisung extrahierter Bildmerkmale zu den Merkmalsvariablen auch jeder Statusvariablen einer von drei möglichen Werten zugewiesen werden muß.

Es können jedoch bereits vor der Suche a priori Einschränkungen des Suchraumes durchgeführt werden. Die Schwellwerte der geometrischen Constraints Linienwinkel, Linienparallelität und Kollinearität wurden so gewählt, daß die Constraints für alle Zuordnungen der Trainingsdaten erfüllt waren. Dies kann auch so gedeutet werden, daß wenn die beteiligten Merkmale beobachtbar sind, in allen akzeptablen Fällen die entsprechenden Constraints erfüllt sind. Das bedeutet, daß wenn ein Constraint der oben aufgezählten Typen verletzt ist, dies nur auf die Unbeobachtbarkeit eines Merkmals zurückzuführen ist. Diese Schlußfolgerung erlaubt es, a priori den Wert  $-1$  aus den Wertebereichen der Statusvariablen  $b_i$  der entsprechenden geometrischen Constraints zu entfernen.

Weitere Suchraumeinschränkungen ergeben sich a priori aus den zuvor rekonstruierten Gebäudeecken, da diese Ecken anhand der Zuordnung extrahierter Bildmerkmale gebildet wurden. Bei der Generierung der Gebäudehypothesen und der Berechnung der Gebäudeansichten wird zu jeder Ecke vermerkt, welche Bildmerkmale den Modellmerkmalen zugeordnet wurden. Die Wertebereiche der Variablen, die den bekannten Eckenmerkmalen der Gebäudeansicht entsprechen, können somit unmittelbar auf die jeweils bei der Eckenrekonstruktion zugeordneten Merkmale beschränkt werden. Wurde ferner bei der Eckenrekonstruktion festgestellt, daß z.B. der Knotenpunkt oder eine inzidente Bildkante in einem Bild unbeobachtbar ist, wird der entsprechenden Variablen von vornherein der Wildcard zugeordnet.

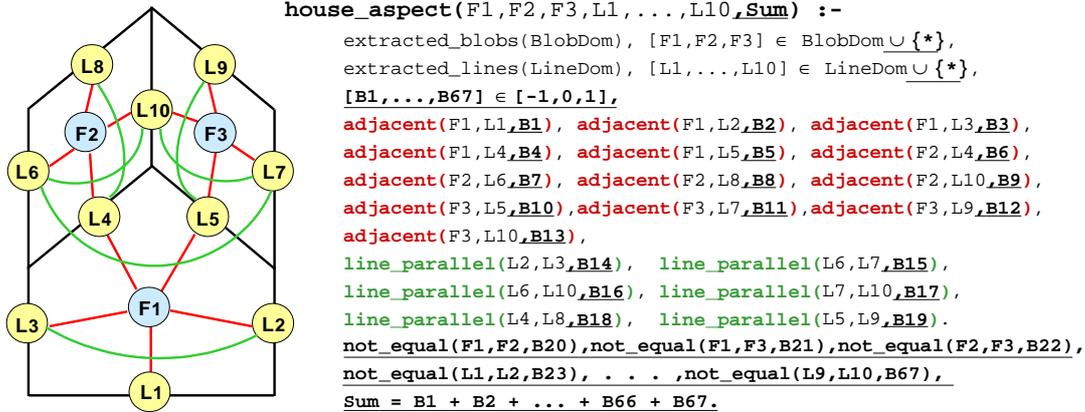
Wenn man davon ausgeht, daß die Bilddaten eine gewisse Mindestqualität aufweisen und für die zuverlässige Rekonstruktion eines Gebäudes eine bestimmte Mindestanzahl von Merkmalen benötigt wird, kann zudem der Suchraum eingeschränkt werden, indem man eine maximale Beschränkung der Anzahl möglicher Wildcard-Zuordnungen vornimmt. Wird während der Suche diese Anzahl erreicht, kann automatisch aus allen Wertebereichen der noch nicht eindeutig festgelegten Variablen der Wildcard sowie aus den Wertebereichen der Statusvariablen der korrespondierenden Constraints der Wert  $0$  entfernt werden. Zur Realisierung kann unmittelbar der in allen CLP(FD)-Systemen eingebaute Standard-Constraint `atmost(N,VarList,Value)` verwendet werden, der festlegt (und auch propagiert), daß der Wert *Value* höchstens  $N$ -mal in der Liste der Variablen *VarList* vorkommen darf. Bei allen folgenden Beispielen wurde  $N$  auf 35% der Anzahl der Merkmalsvariablen gesetzt, d.h. es werden nur Lösungen toleriert, bei denen mindestens 65% der Modellmerkmale im Bild beobachtbar sind. Dieser Wert kann aus den Trainingsdaten ermittelt werden, die bereits zur Bestimmung der Constraint-Schwellwerte verwendet wurden (vgl. Abschnitt 3.3.2).

### 4.3.6 Praktische Evaluation

Zur praktischen Evaluation muß zunächst die Transformation von Gebäudeansichten in eine CLP-Klausel angepaßt werden. Abbildung 4.4 zeigt die bzgl. Abbildung 3.11 auf Seite 56 entsprechend modifizierte Klausel einer Gebäudeansicht eines Satteldachhauses.

Die wesentlichen Änderungen bestehen in der Erweiterung der Wertebereiche um den Fehlwert  $*$ , die Ergänzung der Constraints um die Statusvariablen  $B_1 \dots B_{67}$  und der Summierung dieser Variablen zur Bildung der Bewertungsfunktion. Ein weiterer Unterschied besteht darin, daß zur Forderung nach der Verschiedenheit der zugeordneten Bildmerkmale nicht mehr der eingebaute `alldistinct`-Constraint verwendet werden kann, sondern alle Variablen paarweise durch einen neuen, relaxierbaren `not_equal`-Constraint zueinander in Beziehung gesetzt werden müssen.

Im folgenden wird die Anwendbarkeit der erweiterten Constraint-Modellierung und der vorgestellten Inferenzregeln an zwei Testreihen untersucht. An beiden Reihen soll als erstes demonstriert



**Abbildung 4.4:** Erweiterte CLP-Klausel zur Repräsentation der Vorderansicht eines Satteldachhauses. Aus Gründen der Übersichtlichkeit werden nur einige der Relationen dargestellt und die Punktmerkmale weggelassen. Die sich bzgl. der in Abbildung 3.11 gezeigten Klausel ergebenden Änderungen sind unterstrichen.

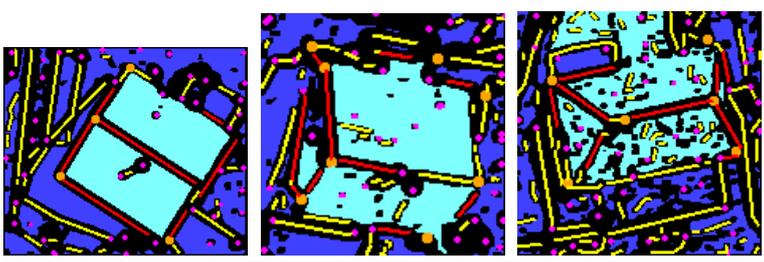
werden, daß das vorgeschlagene Konzept auch bei fehlerbehafteten Daten korrekte Zuordnungen bestimmt, wobei die Phänomene der Unbeobachtbarkeit und Unschärfe entsprechend differenziert werden. Darüberhinaus soll anhand der ersten Testreihe die Effektivität der neuen Inferenzregeln im Gegensatz zur Anwendung der Standard-Forward-Checking-Inferenzregel aufgezeigt werden. Des weiteren geht es darum, die Bedeutung des Vorwissens aus der 3D-Eckenrekonstruktion in bezug auf den Suchaufwand zu illustrieren. Die zweite Testreihe soll schließlich noch einmal verdeutlichen, daß bereits ohne eine differenzierte Gewichtung der Constraints in den meisten Fällen die richtigen Zuordnungen bestimmt werden.

In beiden Testreihen wird weiterhin die First-Fail-Heuristik zur Auswahl der nächsten Variable, der ein Wert zugeordnet werden soll, verwendet. Dabei werden die Variablen, die die Modellmerkmale repräsentieren, sowie die Statusvariablen in einer Menge zusammengefaßt, aus der dann jeweils die Variable mit dem kleinsten Wertebereich ausgewählt wird (vgl. Abschnitt 3.1.3.1).

Die erste Testreihe wird in Tabelle 4.2 dargestellt und besteht aus drei Luftbildausschnitten, wobei die ersten beiden ein Satteldachhaus und der dritte ein Walmdachhaus zeigen. In der linken Spalte wird dabei das Beispiel vom Anfang des Kapitels aufgegriffen (vgl. Abbildung 4.1 auf Seite 67). In der Tabelle sind zu jedem Bild die Anzahl der Modellmerkmale der Gebäudehypothese, die in dem Bild gesucht wird, aufgeführt. Dabei wird angegeben, für wieviele dieser Variablen bereits aufgrund der bei der Generierung der Gebäudehypothesen berücksichtigten rekonstruierten 3D-Ecken bereits das zuzuordnende Bildmerkmal feststeht. Die Anzahl der freien Variablen ergibt sich also aus der Summe der Modellvariablen minus der Anzahl der a priori feststehenden Variablen. Des weiteren werden zu jedem Bild die Anzahlen extrahierter Punkt-, Linien- und Flächenmerkmale angegeben. Die erzielten Ergebnisse<sup>2</sup> sind jeweils unterteilt nach der Art der verwendeten Inferenzregeln. 'FCIR' steht dabei für die Verwendung der Forward-Checking-Inferenzregel und 'neue IR' für die neuen Inferenzregeln (vgl. Abschnitt 4.3.3).

Als erstes läßt sich feststellen, daß in allen drei Bildern die gesuchte Gebäudehypothese korrekt identifiziert wurde, d.h. die ermittelte beste Zuordnung der Bild- zu den Modellmerkmalen entspricht genau der Erwartung. Gebäudemerkmal, die im Bild nicht extrahiert werden konnten, werden richtigerweise als unbeobachtbar erkannt. Das Vorwissen wirkt sich deutlich auf die Laufzeit bzw. die Anzahl besuchter Suchbaumknoten aus. Im linken Bild standen bereits für 53% der Variablen die zuzuordnenden Merkmale fest, was eine sehr kurze Suche zur Folge hatte. Im mittleren Bild waren mit 16% und im rechten Bild mit 21% ungefähr gleich viele Variablen a

<sup>2</sup>Die Laufzeiten wurden auf einem Linux-PC mit 266 MHz AMD-K6-CPU unter ECLIPSE 3.7.1 ermittelt.

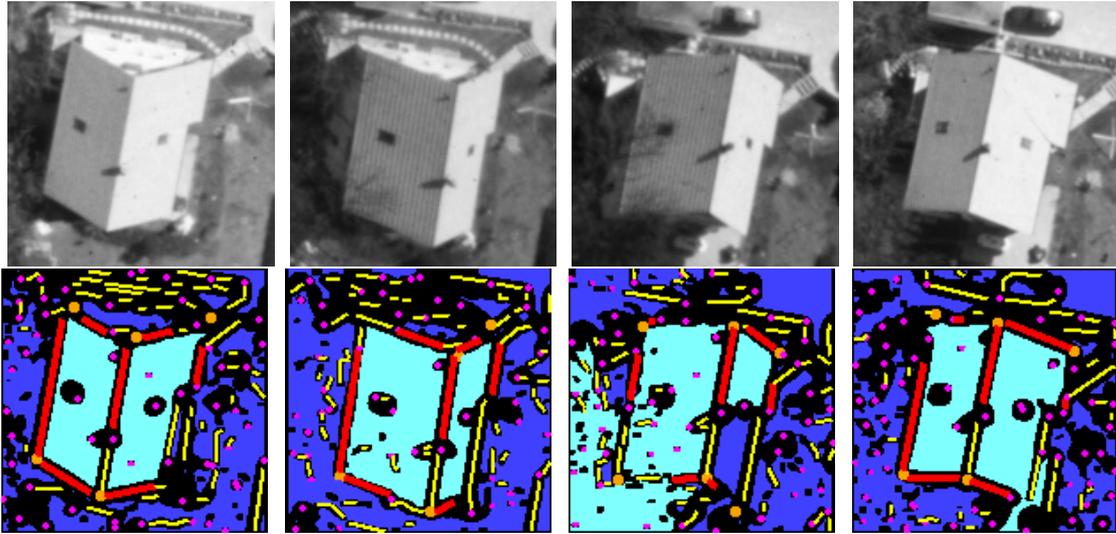
						
<b>Modell (Gebäudeansicht)</b>						
# Punktvariablen	6		9		6	
# Linienvariablen	7		12		9	
# Flächenvariablen	2		4		4	
davon insges. bereits fest	8		4		4	
# Constraints	105		206		139	
<b>Bild</b>						
# Punktmerkmale	54		50		55	
# Linienmerkmale	47		57		109	
# Flächenmerkmale	35		21		90	
<b>Ergebnisse</b>	FCIR	neue IR	FCIR	neue IR	FCIR	neue IR
Bewertung $\sum b_i$	75	75	–	135	–	103
# Wildcard-Zuordnungen	2	2	–	3	–	0
# verletztter Constraints	4	4	–	11	–	18
besuchte Suchbaumknoten	127088	8112	–	297863	–	929857
# Konsistenztests	9612513	47796	–	2180912	–	8327051
davon aus dem Cache	94,19 %	78,67 %	–	77,99 %	–	59,66 %
Laufzeit in Sek.	168,35	5,53	–	351,53	–	999,06

**Tabelle 4.2:** Drei Beispiele für die Bestimmung der besten Zuordnung von Bild- zu Modellmerkmalen unter expliziter Berücksichtigung von Unbeobachtbarkeit und Unschärfe. Zu-geordnete Flächenmerkmale sind hellblau, Linienmerkmale rot und Punktmerkmale orange hervorgehoben. In allen drei Fällen wurde trotz schlechter Bilddaten eine korrekte Zuordnung vorgenommen. Das linke Bild zeigt das Beispiel vom Anfang des Kapitels (siehe Abbildung 4.1 auf Seite 67). Der rechte Giebelpunkt wurde als unbeobachtbar erkannt, und auch der rechte obere Dacheckpunkt wurde als nicht vorhanden eingestuft, weil der an dieser Stelle im Bild extrahierte Punkt nicht kollinear zu der rechten Dachkante ist. Im mittleren Bild wird die rechte obere Dachkante als unbeobachtbar angenommen, weil die dort im Bild sichtbaren Linien eine zu große Winkelabweichung von der Modelllinie aufweisen. Im rechten Bild wurde aufgrund der geringen Bildauflösung nur nach der Dachstruktur des Walmdachhauses gesucht, die ebenfalls korrekt identifiziert wurde. Die Meßergebnisse werden im Text näher erläutert.

priori festgelegt. Dennoch war der Suchaufwand im rechten Bild deutlich größer, was sich durch die größere Anzahl extrahierter Bildmerkmale erklärt.

Wie sich an den Meßergebnissen zum linken Bild erkennen läßt, ist die Verwendung der in Abschnitt 4.3.3 vorgestellten Inferenzregeln deutlich effektiver, als die Auswertung der erweiterten Constraints mittels der Standard-Forward-Checking-Inferenzregel. Anstelle von 168 sec. mit der FCIR werden nur knapp über 5 sec. für die Bestimmung der besten Lösung mit den neuen Inferenzregeln benötigt. Bei den anderen beiden Beispielen mußte die Suche mit der FCIR sogar jeweils nach 24 Stunden abgebrochen werden.

Die zweite Testreihe wird in Abbildung 4.5 dargestellt. Sie zeigt vier Aufnahmen desselben Hauses aus unterschiedlichen Perspektiven und die dazu bestimmten besten Zuordnungen, die alle innerhalb weniger Sekunden bestimmt wurden. Wegen der geringen Bildauflösung wurde nur nach den



**Abbildung 4.5:** Die Aufnahmen zeigen dasselbe Gebäude in vier unterschiedlichen Perspektiven. In allen Bildern wurde nach der Dachstruktur der jeweiligen Gebäudehypothese gesucht, und es wurde in allen Fällen die korrekte Zuordnung erkannt, obwohl die verschiedenen Constraints bislang nicht weiter gewichtet wurden.

Dachstrukturen der Gebäudehypothesen gesucht. Als Ergebnis dieser Testreihe soll festgehalten werden, daß die um Unbeobachtbarkeit und Unschärfe erweiterte Constraint-Modellierung bereits ohne eine ausdifferenzierte Gewichtung der einzelnen Constraints in vielen Fällen eine korrekte Zuordnung von Bild und Modell erreicht.

### 4.3.7 Übertragung auf andere überbeschränkte Probleme

Zum Abschluß dieses Kapitels soll noch gezeigt werden, daß die erweiterte Constraint-Modellierung eine natürliche Verbindung zwischen dem MaxCSP- und dem Dynamic CSP-Formalismus herstellt, die, wie sich Abbildung 4.3 auf Seite 72 entnehmen läßt, bislang noch nicht zueinander in Beziehung gesetzt wurden. Diese Verbindung läßt sich auch für ganz andere Anwendungsgebiete nutzen, was abschließend an einem Beispiel für ein (einfaches) Konfigurationsproblem demonstriert wird.

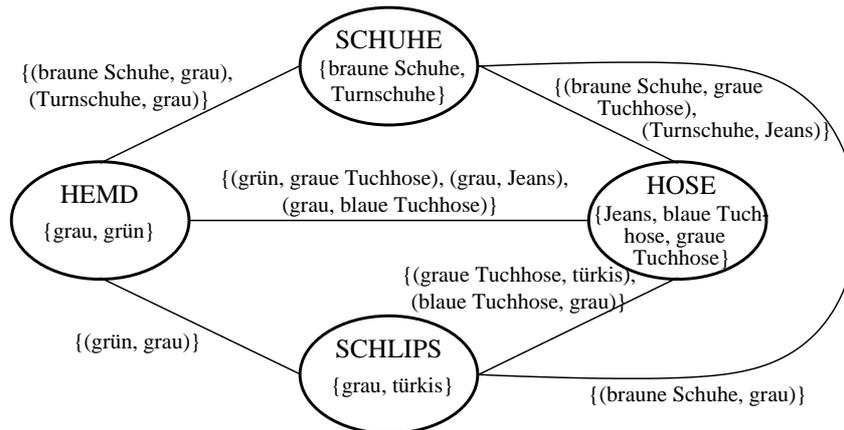
Im folgenden wird gezeigt, daß MaxCSP und Dynamic CSP Spezialfälle der in den Abschnitten 4.3.1 und 4.3.2 vorgestellten Modellierung sind. Um dies zu beweisen, reicht es zu zeigen, daß sowohl MaxCSP als auch Dynamic CSP mit Hilfe des entwickelten Konzepts simuliert werden können.

Zur Realisierung des MaxCSP-Verfahrens reicht es aus, wenn der Wildcard \* aus den Wertebereichen aller Variablen  $v_1, \dots, v_n$  gestrichen wird, und der Wert 0 aus den Wertebereichen aller Statusvariablen  $b_i$  der Constraints  $c'_i$ . Ein Constraint  $c'_i$  modelliert damit nur noch die zwei Fälle, in denen der ursprüngliche Constraint  $c_i$  entweder erfüllt ( $b_i = 1$ ) oder verletzt ist ( $b_i = -1$ ). Die Maximierung der Summe über alle  $b_i$  gemäß Gleichung 4.9 maximiert damit offensichtlich auch die Anzahl erfüllter Constraints.

Dynamic CSP unterscheidet zwischen aktiven und inaktiven Variablen sowie zwischen normalen Constraints und Aktivierungs-Constraints. Bei inaktiven Variablen werden alle Constraints, in denen sie vorkommen, ebenfalls inaktiv. Eine a priori ausgezeichnete Teilmenge der Variablen wird zudem als konstant aktiv angenommen. Bei der Abbildung eines Dynamic CSP auf die in diesem Kapitel vorgestellte Modellierung werden die Variablen und die normalen Constraints unmittelbar übernommen. Da Verletzungen der normalen Constraints zwischen aktiven Variablen nicht erlaubt sind, kann zunächst bereits der Wert  $-1$  aus den Wertebereichen der Statusvariablen  $b_i$  entfernt

werden. Die Deaktivierung einer Variablen läßt sich unmittelbar durch die Zuweisung eines Wildcards erreichen, was nach Inferenzregel 4.5 dazu führt, daß alle zu dieser Variablen inzidenten Constraints durch das Setzen ihrer Statusvariablen auf den Wert 0 dann gemäß Inferenzregel 4.8 ebenfalls relaxiert werden. Bei den konstant aktiven Variablen werden von Anfang an die Wildcards aus den entsprechenden Wertebereichen entfernt. Somit können diese Variablen nicht mehr deaktiviert werden. Aktivierungs-Constraints der Art  $P(v_1, \dots, v_j) \rightarrow active : v_k$  werden transformiert nach  $P'(v_1, \dots, v_j, b) \Leftrightarrow (b = 0) \vee (b = 1 \wedge (P(v_1, \dots, v_j) \rightarrow v_k \neq *))$ . Das bedeutet, daß  $P'$  entweder erfüllt ist, wenn  $b = 0$  gesetzt wird (weil einer der Variablen  $v_1, \dots, v_j$  ein Wildcard zugeordnet wurde bzw. sie deaktiviert wurde) oder wenn  $b = 1$  ist, und die Erfüllung von  $P(v_1, \dots, v_j)$  die Aktivierung der Variablen  $v_k$  (durch Entfernung des Wildcards aus dem Wertebereich von  $v_k$ ) impliziert. Die zweite Art von Aktivierungs-Constraint (siehe Abschnitt 4.2.3) läßt sich auf analoge Weise transformieren.

Der Nutzen der vorgestellten Modellierung für andere Anwendungsgebiete soll nun noch an einem einfachen Konfigurationsproblem illustriert werden, das dem Aufsatz von Freuder und Wallace über MaxCSP ([FW96]) entnommen und leicht erweitert wurde. Dabei geht es darum, eine möglichst zueinander passende Kleiderkombination aus Schuhen, Hemd, Hosen und Schlips aus einem (recht dürftig ausgestatteten) Kleiderschrank auszuwählen. Die Repräsentation des Problems als CSP wird in Abbildung 4.6 dargestellt. Unter den gegebenen Kleidungsstücken und den Beschränkungen, welche davon zueinander passen, besitzt das CSP keine Lösung und ist demnach überbeschränkt.



**Abbildung 4.6:** Beispiel für ein einfaches Konfigurationsproblem, bei dem es darum geht, eine möglichst gut passende Kleiderkombination auszuwählen (nach [FW96]; erweitert um einen Schlips).

Wird nun das MaxCSP-Verfahren zur Lösungsbestimmung verwendet, maximieren die beiden folgenden Lösungen die Anzahl erfüllter Constraints: 1) braune Schuhe, graues Hemd, grauer Schlips und blaue Tuchhosen; 2) braune Schuhe, grünes Hemd, türkisfarbener Schlips und graue Tuchhosen. Beide Lösungen sind im wahrsten Sinne des Wortes *offensichtlich* inakzeptabel. Wird jedoch zusätzlich zur Verletzung einzelner Constraints auch die Eliminierung von Variablen (durch Zuweisung eines Wildcards) zugelassen, wird bei der besten Lösung vorgeschlagen, Turnschuhe, ein graues Hemd und eine Jeans zu tragen, den Schlips aber wegzulassen.

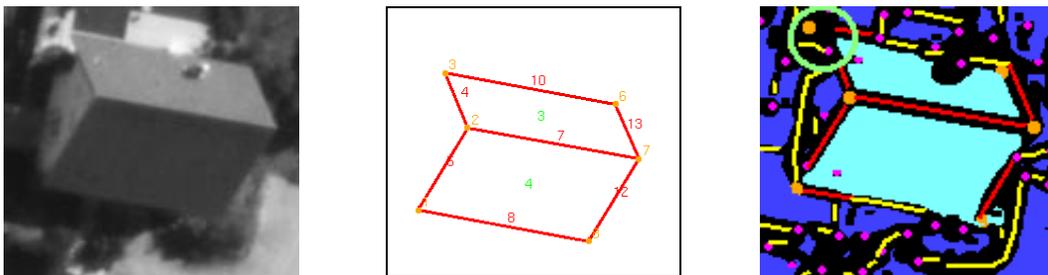


## Kapitel 5

# Bestimmung der wahrscheinlichsten relationalen Zuordnung

Im vorangegangenen Kapitel wurde eine methodische Erweiterung der Constraint-Techniken vorgestellt, die sowohl die Verletzung einzelner Constraints als auch die Unbeobachtbarkeit von Merkmalen explizit handhaben kann. Dazu wurde jeder Constraint durch eine Statusvariable  $b$  mit den drei möglichen Werten 1, 0 und  $-1$  ergänzt, die besagen, ob der ursprüngliche Constraint erfüllt ( $b = 1$ ), aufgrund der Unbeobachtbarkeit eines inzidenten Merkmals nicht vorhanden ( $b = 0$ ) oder verletzt ist ( $b = -1$ ). Zur Bestimmung der besten Lösung wurden diese Werte über alle Constraints aufsummiert, und die Zuordnung bestimmt, die den Wert dieser Bewertungsfunktion maximiert.

Wie im folgenden gezeigt wird, weisen die verschiedenen Constraints unterschiedliche Störanfälligkeiten auf. Beispielsweise werden Merkmalsnachbarschafts-Constraints häufiger verletzt als Linieneinheits-Constraints. Da die Verletzung und Erfüllung für alle Constraints bislang gleich bewertet wird, können dadurch fehlerhafte Zuordnungen bestimmt werden, wie Abbildung 5.1 illustriert.



**Abbildung 5.1:** Beispiel für eine fehlerhafte Zuordnung, die aus der fehlenden Gewichtung der Constraints herrührt. Das mittlere Bild zeigt die generierte Dachhypothese und das rechte Bild die zugeordneten Bildmerkmale der besten Lösung. Die den Modellmerkmalen zugeordneten Flächen sind hellblau, Linien rot und Punkte orange hervorgehoben. Dem linken oberen Eckpunkt des Daches wurde fälschlicherweise der in dem grünen Kreis hervorgehobene Punkt zugeordnet, der jedoch zu weit von der Dachfläche entfernt liegt. Der Grund für diese Zuordnung liegt darin, daß die Kosten für die Verletzung der Merkmalsnachbarschafts-Constraints des Punktes zu den beiden inzidenten Linien und der Dachfläche im Verhältnis zu den anderen Constraints zu niedrig sind.

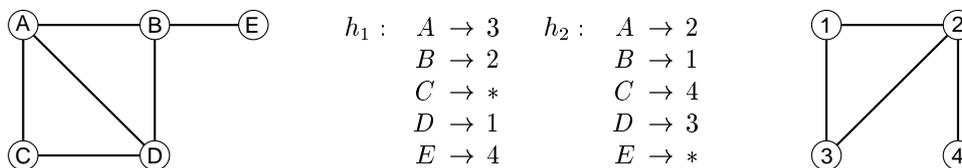
Das Problem besteht nun darin, Gewichte für die Constraints zu bestimmen, die die Störanfälligkeiten angemessen berücksichtigen. Darüberhinaus stellt sich die Frage nach der Semantik der besten Zuordnung, denn es ist bislang nicht klar, ob die Zuordnung mit der höchsten Bewertung tatsächlich auch die im Sinne der Objekterkennung beste Lösung darstellt. Schließlich muß auch ein objektives Kriterium definiert werden können, das entscheidet, wann eine Lösung insgesamt zu schlecht ist, und als unwahrscheinliche Zuordnung zurückgewiesen werden muß. Dies ist beispielsweise für den Fall notwendig, wenn in dem betrachteten Luftbildausschnitt gar kein Gebäude abgebildet ist, das Verfahren aber trotzdem eine Lösung ermittelt hat.

Im Bereich der Objekterkennung wurden bereits verschiedene Bewertungsfunktionen für relationale Zuordnungen entwickelt, die auch Bildstörungen berücksichtigen können. Einen Überblick über die Verfahren gibt Vosselman in [Vos92]. In dieser Arbeit wird ein von Boyer und Kak entwickeltes und von Vosselman erweitertes Bewertungsschema verwendet, welches das Problem der Zuordnung auf ein Kommunikationsproblem abbildet und mit einem informationstheoretischen Maß bewertet. Dieses Bewertungsmaß besitzt eine wahrscheinlichkeitstheoretische Grundlage und ermöglicht die Bestimmung der wahrscheinlichsten Zuordnung von Modell und Bild. Darüberhinaus erlaubt sie die Definition eines Hypothesentests zur Zurückweisung unwahrscheinlicher Zuordnungen.

Nach der Vorstellung des Bewertungsschemas wird gezeigt, wie die Bewertung zur Gewichtung der Constraints verwendet wird. Die Gewichte werden anhand von Wahrscheinlichkeiten berechnet, die sowohl analytisch als auch empirisch anhand von Trainingsdaten ermittelt werden. Zur Kopplung mit den erweiterten Constraints müssen dabei keine Änderungen an dem Constraint-Interpreter oder den in Abschnitt 4.3.3 vorgestellten Inferenzregeln vorgenommen werden. Es werden lediglich die Werte  $-1$  und  $1$  der Statusvariable  $b$  jedes Constraints  $c'(\dots, b)$  durch entsprechende Gewichte ersetzt (vgl. Abschnitt 4.3). Am Ende des Kapitels wird das in Abbildung 5.1 gezeigte Beispiel erneut aufgegriffen, und die Auswirkungen der Gewichtung demonstriert.

## 5.1 Inexakte relationale Zuordnungen

Sowohl die Modelle (Gebäudeansichten) als auch die Daten (extrahierte Bildmerkmale) liegen als relationale Strukturen vor. An den Ergebnissen des vorigen Kapitels wird deutlich, daß bei der Suche von Modellen im Bild im allgemeinen nicht zu erwarten ist, daß die Modelle exakte Subgraphen der Bildgraphen darstellen. Im Gegensatz zu Kapitel 3, bei der diese idealisierende Annahme gemacht wurde, reicht zur Bestimmung der korrekten Zuordnung von Bild- zu Modellmerkmalen demnach die Suche nach exakten Subgraph-Isomorphismen nicht mehr aus. Da Modellmerkmale im Bild unbeobachtbar sein können, und auch Relationen im Bild fehlen können, wird die Bestimmung einer solchen Zuordnung als *inexaktes Relational Matching* bezeichnet (vgl. [SH85, BK88, Vos92]).



**Abbildung 5.2:** Beispiel für zwei verschiedene inexakte relationale Zuordnungen  $h_1$  und  $h_2$ . Beide Zuordnungen bilden die Knoten (Primitive) des linken Graphen auf Knoten (Primitive) des rechten Graphen ab, wobei die zugeordneten Knoten die meisten Relationen erfüllen, die auch zwischen den ursprünglichen Knoten bestehen. In beiden Fällen konnte jeweils für einen Knoten des linken Graphen kein Zuordnungspartner in dem rechten Graphen gefunden werden; diesem Knoten wurde dementsprechend der Fehlwert (Wildcard  $*$ ) zugeordnet.

Es geht im folgenden darum, den Primitiven  $P_1$  einer Relationalstruktur  $D_1 = (P_1, R_1)$ , die z.B. das Modell repräsentiert, entsprechende Primitive  $P_2$  einer zweiten Relationalstruktur  $D_2 = (P_2, R_2)$  (z.B. der Bilddaten) zuzuordnen, so daß eine größtmögliche Ähnlichkeit bzw. Subgraph-Isomorphie zwischen den beiden Strukturen bzgl. der Relationen  $R_1$  und  $R_2$  erreicht wird. Wenn für ein Primitiv kein passender Zuordnungspartner existiert, wird ihm ein Fehlwert, der sogenannte *Wildcard*, zugeordnet (vgl. Abschnitt 4.3.2). Abbildung 5.2 illustriert dieses Problem anhand der inexakten relationalen Zuordnung zweier Graphen.

Das Problem besteht nun darin, die im Sinne der Anwendung beste inexakte Zuordnung zu bestimmen. Dazu wird eine Bewertungsfunktion benötigt, die entweder die Abweichung einer gegebenen Zuordnung von der optimalen, exakten Zuordnung mißt, oder andersherum die Ähnlichkeit der einander zugeordneten Graphstrukturen bewertet.

### 5.1.1 Ein einfaches Distanzmaß für die strukturelle Abweichung

Zur Bestimmung der besten Zuordnung wird von Haralick und Shapiro in [SH85] ein Distanzmaß vorgeschlagen, das die strukturellen Abweichungen zweier Relationalstrukturen  $D_1 = (P_1, R_1)$  und  $D_2 = (P_2, R_2)$  zueinander bewertet. Dazu wird für eine gegebene Zuordnung  $h : P_1 \mapsto P_2$  summiert, wieviele Relationentupel aus  $R_1$  nicht auf  $R_2$  abgebildet werden können und umgekehrt. Die strukturelle Distanz  $\mathcal{E}(h)$  ergibt sich dann zu

$$\mathcal{E}(h) = |R_1 \circ h - R_2| + |R_2 \circ h^{-1} - R_1|. \quad (5.1)$$

Obwohl diese Metrik später noch um die Gewichtung der einzelnen Relationentupel ergänzt wurde (vgl. [HS93a]), eignet sie sich nur bedingt für die Objekterkennung, denn ähnlich wie bei den Verfahren MaxCSP und HCLP zur Lösung überbeschränkter Systeme (siehe Abschnitt 4.2.5), wird auch hier bei der Summierung der Fehler nicht zwischen der Unbeobachtbarkeit ganzer Merkmale und der einfachen Verletzung von Relationen differenziert. Darüberhinaus ist die Gewichtung der Relationentupel problematisch, weil die verschiedenen Relationen Attribute unterschiedlichen Typs (z.B. numerisch oder symbolisch) mit unterschiedlichen Wertebereichen besitzen können (siehe dazu [Vos92]).

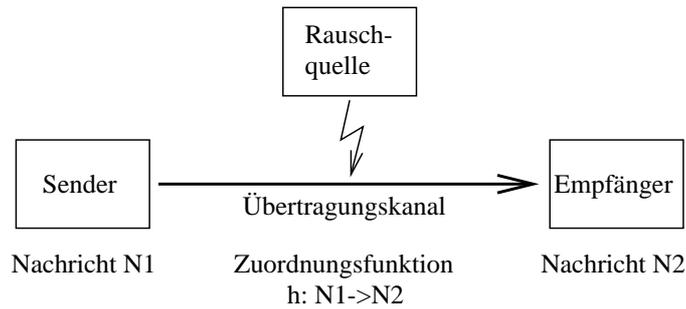
### 5.1.2 Relational Matching als Kommunikationsproblem

Boyer und Kak schlagen in ihrer Arbeit [BK88] vor, das Zuordnungsproblem auf ein Kommunikationsproblem abzubilden. Dabei wird eine relationale Struktur  $D_1 = (P_1, R_1)$  über einen veräuschten Übertragungskanal gesendet, und als verzerrte bzw. veränderte Struktur  $D_2 = (P_2, R_2)$  empfangen. Der Übertragungskanal wird durch die Zuordnungsfunktion  $h : P_1 \mapsto P_2$  repräsentiert, die jedem Primitiv  $p_i \in P_1$  ein Primitiv  $q_j \in P_2$  zuordnet (vgl. Abbildung 5.3). Die Frage nach der Bestimmung der besten Zuordnung stellt sich somit als Suche nach dem besten Kanal dar, der sich durch die kleinste Abweichung von  $D_2$  zu  $D_1$  auszeichnet. Zur Lösung dieses Problems wird im folgenden auf die Informationstheorie zurückgegriffen (siehe [Ing71]).

Shannon stellte 1948 eine umfassende Theorie über die mathematischen Eigenschaften der Kommunikation vor ([Sha48]). Darin wird der Informationsgehalt eines Symbols  $a_i$  aus dem endlichen Alphabet  $A$  definiert durch

$$I(a_i) = \log_2 \frac{1}{P(a_i)} = -\log_2 P(a_i) \quad [\text{Bit}]. \quad (5.2)$$

Die Information ist demnach ein Maß für die Überraschung, daß ein bestimmtes Symbol in einer Nachricht vorkommt, denn bei einer kleinen Auftretenswahrscheinlichkeit ist der Informationswert groß und bei einer hohen Wahrscheinlichkeit gering. Falls es sicher ist, daß eine Nachricht ein bestimmtes Symbol  $a_i$  enthält ( $P(a_i) = 1$ ), ist sein Informationswert 0.



**Abbildung 5.3:** Abbildung des Zuordnungsproblems auf ein Kommunikationsproblem. Die Relationalstruktur  $D_1$  wird als Nachricht N1 interpretiert, über einen verrauschten Kanal  $h$  gesendet und als veränderte Nachricht N2 (mit Relationstruktur  $D_2$  als Inhalt) empfangen.

Wenn für jedes Symbol des Alphabets eines Senders  $A = \{a_1, \dots, a_N\}$  die Wahrscheinlichkeiten bekannt sind, daß das Symbol gesendet wird, beträgt der durchschnittliche Informationsgehalt pro übertragenem Symbol

$$H(A) = \sum_{i=1}^N P(a_i) \cdot I(a_i) = \sum_{i=1}^N -P(a_i) \log_2 P(a_i) \quad [\text{Bit}]. \quad (5.3)$$

$H(A)$  wird auch als *Entropie* des Senders mit Alphabet  $A$  bezeichnet, und ist ein Maß für die Unsicherheit bei der Auswahl der zu übertragenden Symbole<sup>1</sup>.

Die Symbole des Empfängers bilden ebenfalls ein Alphabet  $B = \{b_1, \dots, b_M\}$ , wobei  $P(b_j)$  die Wahrscheinlichkeit angibt, mit der ein Symbol  $b_j$  empfangen wird. Die Übertragungscharakteristik des Kanals kann dann durch die bedingte Wahrscheinlichkeit des gesendeten Symbols  $a_i$ , gegeben einem empfangenen Symbol  $b_j$ , modelliert werden. Bei einem idealen Kanal entspricht jedes gesendete Symbol genau einem bestimmten Symbol des Empfängers. Die bedingte Wahrscheinlichkeit  $P(a_i|b_j)$ , die auch *Übertragungswahrscheinlichkeit* genannt wird, beträgt dabei für ein  $a_i$  bzgl.  $b_j$  genau 1, und für alle anderen  $a_k \neq a_i$  genau 0. Im Fall von Rauschen kann es vorkommen, daß ein Symbol bei der Übertragung verfälscht wird, und ein anderes Symbol als erwartet empfangen wird. Die Überraschung, die entsteht, wenn bekannt wird, daß  $a_i$  gesendet und  $b_j$  empfangen wurde, wird durch die bedingte Information ausgedrückt:

$$I(a_i|b_j) = -\log_2 P(a_i|b_j) \quad [\text{Bit}]. \quad (5.4)$$

Eine wichtige Größe neben der bedingten Information ist die *gemeinsame Information* zwischen einem gesendeten Symbol  $a_i$  und einem empfangenen Symbol  $b_j$ . Sie ist definiert als

$$I(a_i; b_j) = I(a_i) - I(a_i|b_j) = \log_2 \frac{P(a_i|b_j)}{P(a_i)} \quad (5.5)$$

$$= \log_2 \frac{P(a_i, b_j)}{P(a_i)P(b_j)} \quad (5.6)$$

und bezeichnet die Information, die ein Symbol über das andere gibt. Wie an Gleichung 5.6 erkennbar ist, handelt es sich um ein symmetrisches Maß. Die gemeinsame Information spielt eine wichtige Rolle bei der Bewertung von Kommunikationssystemen, da sie den tatsächlich übermittelten Informationsgehalt angibt (siehe [Sha48, Ing71]).

<sup>1</sup>Die Entropie spielt erst im nächsten Kapitel bei der Codierung von relationalen Strukturen eine Rolle (Abschnitt 6.1.2).

### 5.1.2.1 Information einer relationalen Struktur

Nach Boyer und Kak beträgt der Informationsgehalt einer relationalen Struktur  $I(D_1) = I(P_1) + I(R_1)$ , wobei  $I(P_1) = \sum_{p \in P_1} I(p)$  den Informationsgehalt aller Primitive aus  $P_1$  und  $I(R_1) = \sum_{r \in R_1} I(r)$  den Informationsgehalt der Relationen beschreibt. Der Informationsgehalt eines Primitivs  $p_i$  setzt sich wiederum aus den Informationen der einzelnen Attribute  $a_l$  zusammen<sup>2</sup>:  $I(p_i) = \sum_{a_l \in p_i} I(a_l)$ . Die Information eines Primitivattributs hängt dabei von der Wahrscheinlichkeit ab, daß das Attribut einen bestimmten Wert  $wert_l$  annimmt:  $I(a_l = wert_l) = -\log_2 P(a_l = wert_l)$ .

Die Information einer Relation  $r_i$  ist definiert als die Summe der Informationen aller Relationentupel  $I(r_i) = \sum_{t_j \in r_i} I(t_j)$ . Der Informationsgehalt eines Relationentupels setzt sich schließlich aus den Informationen der einzelnen Attribute der Relationentupel zusammen  $I(t_j) = \sum_{a_l \in t_j} I(a_l)$ , wobei  $I(a_l)$  die gleiche Definition besitzt wie oben.

### 5.1.2.2 Bewertung einer relationalen Zuordnung

Die bedingte Information ist ein Maß für die Unsicherheit der Übertragung. Zur Messung der bei der Übertragung einer relationalen Struktur  $D_1 = (P_1, R_1)$  aufgetretenen Störungen, und damit zur Bewertung der Abweichung von  $D_1$  zu der empfangenen Struktur  $D_2 = (P_2, R_2)$ , müssen nur die im vorangegangenen Abschnitt verwendeten Informationswerte  $I(D_1)$  jeweils durch die bedingten Informationen  $I(D_1|D_2)$  ersetzt werden. Die Übertragungsfunktion  $h$  ordnet jedem Primitiv aus  $P_1$  der Relationalstruktur  $D_1$  jeweils ein Primitiv aus  $P_2$  aus der Relationalstruktur  $D_2$  zu. Boyer und Kak haben in [BK88] gezeigt, daß die Zuordnungsfunktion  $\hat{h}$ , die die bedingte Information  $I_h(D_1|D_2)$  minimiert, die wahrscheinlichste Zuordnung im Sinne einer Maximum-Likelihood-Schätzung  $P(h|D_1, D_2)$  darstellt.

Der wesentliche Vorteil dieses Bewertungsschemas bzgl. des in Abschnitt 5.1.1 gezeigten Distanzmaßes liegt darin, daß die Abweichungen zwischen den einzelnen Komponenten zweier Relationalstrukturen probabilistisch modelliert werden, und die beste Zuordnung (mit der kleinsten Abweichung) die wahrscheinlichste ist. Das Problem der Bewertung der Fehlwertzuordnungen besteht jedoch weiterhin, weil es nicht möglich ist, die bedingte Information  $I(p_i|*)$  eines Primitivs  $p_i$  bzgl. eines Fehlwertes zu berechnen.

An dieser Stelle setzt die Arbeit von Vosselman auf, der in [Vos92] vorschlägt, anstelle der bedingten Information  $I_h(D_1|D_2)$  die gemeinsame Information  $I_h(D_1; D_2)$  zu verwenden. Die Bewertungsfunktion nimmt dann folgende Gestalt an:

$$I_h(D_1; D_2) = \underbrace{\sum_{p_i \in P_1, h(p_i) \in P_2} I(p_i; h(p_i))}_{\text{Merkmalskorrespondenz}} + \underbrace{\sum_{r_j \in R_1, h(r_j) \in R_2} I(r_j; h(r_j))}_{\text{relationale Korrespondenz}} \quad (5.7)$$

Diese Bewertungsfunktion mißt nicht mehr die Abweichung der beiden Relationalstrukturen voneinander, sondern ihre Ähnlichkeit. Vosselman hat gezeigt, daß es sich bei der Zuordnungsfunktion  $\hat{h}$ , die den Wert von Gleichung 5.7 maximiert

$$\hat{h} : \arg \max_h I_h(D_1; D_2) \quad (5.8)$$

nach wie vor um die wahrscheinlichste Zuordnung im Sinne der Maximum-Likelihood-Schätzung  $P(h|D_1, D_2)$  handelt.

Die Verwendung der gemeinsamen Information besitzt gegenüber der bedingten Information verschiedene Vorteile (siehe [Vos92]), wobei an dieser Stelle nur auf den für diese Arbeit wesentlichen Aspekt der Fehlwertzuordnungen eingegangen wird. Wird einem Primitiv  $p_i$  ein Fehlwert  $*$  zugeordnet, muß dafür die gemeinsame Information  $I(P_i; *)$  berechnet werden. Da die Eigenschaften

<sup>2</sup>Es wird dabei angenommen, daß die Attribute voneinander unabhängig sind (vgl. [BK88, Vos92]).

des Primitivs statistisch unabhängig von einem allgemeinen Fehlwert sind, ergibt sich aus Gleichung 5.6 die gemeinsame Information zu

$$\begin{aligned} I(p_i; *) &= \log_2 \frac{P(p_i) \cdot P(*)}{P(p_i) \cdot P(*)} \\ &= 0. \end{aligned} \tag{5.9}$$

Das bedeutet, daß eine Fehlwertzuordnung neutral mit 0 bewertet wird. Dies läßt sich auch im informationstheoretischen Sinne begründen: ein Fehlwert sagt nichts über das zugeordnete Primitiv aus, und kann deshalb weder für noch gegen die Zuordnung sprechen. Dies hat auch einen Einfluß auf die gemeinsame Information von Relationen. Da Relationen nur dann bestehen können, wenn alle inzidenten Primitiv vorhanden sind, werden bei einer Fehlwertzuordnung ebenfalls alle zu dem entsprechenden Primitiv inzidenten Relationen mit 0 bewertet.

### 5.1.2.3 Ein Hypothesentest bzgl. der Höhe der gemeinsamen Information

Die Maximum-Likelihood-Schätzung bestimmt die wahrscheinlichste Zuordnung der beiden relationalen Strukturen unabhängig davon, wie ähnlich sie sich tatsächlich sind. Bei der zur Gebäudeerkennung verwendeten Strategie kann es vorkommen, daß keine der generierten Gebäudehypothesen dem im Luftbild abgebildeten Haus entspricht. Andererseits kann es auch passieren, daß der ausgewählte Luftbildausschnitt gar kein Gebäude beinhaltet, und nur zufällig eine 3D-Eckenstruktur rekonstruiert wurde, anhand derer Gebäudehypothesen generiert wurden. Das Zuordnungsverfahren wird in allen Fällen zu jeder Gebäudeansicht eine beste Zuordnung ermitteln. Es muß demnach ein Kriterium definiert werden, das entscheidet, wann eine Zuordnung objektiv als „zu schlecht“ einzustufen ist, und damit die Gebäudeansicht als unzutreffend zurückweist.

Vosselman stellt dazu in [Vos92] ein Konzept vor, das für ein gegebenes Modell die Wahrscheinlichkeitsverteilung der gemeinsamen Information berechnet. Diese Verteilungsfunktion gibt für jeden möglichen gemeinsamen Informationswert  $I$  zu einem Modell die Wahrscheinlichkeit  $P(I_h < I)$  an, mit der die gemeinsame Information  $I_h(D_1; D_2)$  einer Zuordnung  $h$  kleiner ist als  $I$ . Anhand dieser Verteilung kann nun eine untere Schranke für die bei der Zuordnung mindestens zu erreichende gemeinsame Information  $I_h$  in Abhängigkeit einer maximal zu tolerierenden Fehlerwahrscheinlichkeit festgelegt werden. Wenn beispielsweise eine Wahrscheinlichkeit von 5 % für die Zurückweisung einer korrekten Zuordnung in Kauf genommen werden soll, kann in der Verteilung der Informationswert abgelesen werden, an dem dieser den Wert 0,05 überschreitet. Alle Zuordnungen, die unterhalb dieser Schranke liegen, werden als unzutreffend zurückgewiesen.

## 5.2 Berechnung der gemeinsamen Information für die verwendeten Constraints

Die Repräsentation der Gebäudeansichten durch Constraints erfolgt auf der Basis verschiedener topologischer und geometrischer Relationen (vgl. Abschnitt 3.3.1). Die Beziehungen zwischen den Merkmalen einer Gebäudeansicht stellen die Modellrelationen dar und werden künftig mit  $r_m$  bezeichnet. Jedem Modellmerkmal wird jeweils entweder ein extrahiertes Bildmerkmal oder der Fehlwert zugeordnet. Die zwischen den zugeordneten Bildmerkmalen bestehenden Bildrelationen werden mit  $r_i$  bezeichnet ( $i$  steht für *image*). Es gibt nun verschiedene Möglichkeiten, die gemeinsame Information für die verwendeten Constraints bzw. den ihnen zugrundeliegenden Relationen zu bestimmen. Nach Gleichung 5.5 läßt sich die gemeinsame Information durch

$$I(r_i; r_m) = \log_2 \frac{P(r_i|r_m)}{P(r_i)} \quad [\text{Bit}] \tag{5.10}$$

berechnen. Die a priori Wahrscheinlichkeit der Bildrelation  $P(r_i)$  läßt sich durch Anwendung des Gesetzes der totalen Wahrscheinlichkeit (Jeffreys Regel; vgl. [Pea88]) aus der a priori Wahrscheinlichkeit der Modellrelation  $P(r_m)$  und der Übertragungswahrscheinlichkeit  $P(r_i|r_m)$  herleiten, und es ergibt sich

$$I(r_i; r_m) = \log_2 \frac{P(r_i|r_m)}{\sum_{r'_m} P(r_i|r'_m) \cdot P(r'_m)} \quad [\text{Bit}]. \quad (5.11)$$

Zur Berechnung der gemeinsamen Information werden also die a priori Wahrscheinlichkeit  $P(r_m)$  der Modellrelation und die Übertragungswahrscheinlichkeit  $P(r_i|r_m)$  der Bildrelation  $r_i$  bzgl. der korrespondierenden Modellrelation  $r_m$  benötigt.  $P(r_m)$  gibt an, mit welcher Wahrscheinlichkeit die betreffende Relation in einem Modell auftritt bzw. erfüllt ist.  $P(r_i|r_m)$  gibt die Wahrscheinlichkeit an, mit der sich der Zustand einer Modellrelation (erfüllt bzw. nicht erfüllt) auch auf die korrespondierende Bildrelation überträgt.

Die Berechnung soll im folgenden an einem Beispiel verdeutlicht werden. Die a priori Wahrscheinlichkeit der Modellrelation  $P(r_m)$  und die Übertragungswahrscheinlichkeit der Wahrheitswerte der Beispielrelation  $P(r_i|r_m)$  sind dabei durch die in Tabelle 5.1 gezeigten Werte gegeben.

$r_m$	$P(r_m)$	$P(r_i r_m)$	$r_m = \text{wahr}$	$r_m = \text{falsch}$
<i>wahr</i>	0,17	$r_i = \text{wahr}$	0,67	0,01
<i>falsch</i>	0,83	$r_i = \text{falsch}$	0,33	0,99

**Tabelle 5.1:** Beispiel für die a priori Wahrscheinlichkeit einer Modellrelation  $r_m$  (linke Tabelle) und die bedingte Wahrscheinlichkeit der Bildrelation  $r_i$  bzgl. der Modellrelation (rechte Tabelle).

Die sich anhand von Gleichung 5.11 aus Tabelle 5.1 ergebende gemeinsame Information der Modell- und Bildrelation stellt sich dann als  $2 \times 2$ -Tabelle mit den vier möglichen Fällen ( $r_i = \text{wahr}, r_m = \text{wahr}$ ), ( $r_i = \text{falsch}, r_m = \text{wahr}$ ), ( $r_i = \text{wahr}, r_m = \text{falsch}$ ) und ( $r_i = \text{falsch}, r_m = \text{falsch}$ ) dar:

$I(r_i; r_m)$	$r_m = \text{wahr}$	$r_m = \text{falsch}$
$r_i = \text{wahr}$	2,45	-3,61
$r_i = \text{falsch}$	-1,41	0,17

**Tabelle 5.2:** Beispiel für die sich ergebende gemeinsame Information (in Bit), die mittels Gleichung 5.11 aus den in Tabelle 5.1 dargestellten Wahrscheinlichkeiten berechnet wurde.

An diesem Beispiel läßt sich erkennen, daß wenn eine vom Modell erwartete Relation ( $r_m = \text{wahr}$ ) ebenfalls zwischen den zugeordneten Bildmerkmalen erfüllt ist ( $r_i = \text{wahr}$ ), diese die Zuordnung mit 2,45 Bit unterstützt. Wird die Relation von den zugeordneten Bildmerkmalen hingegen nicht erfüllt ( $r_i = \text{falsch}$ ), widerspricht sie der Zuordnung mit -1,41 Bit. Ein anderer Widerspruch zu einer Zuordnung (um -3,61 Bit) entsteht, wenn zwischen zwei Bildmerkmalen eine Relation beobachtet werden kann ( $r_i = \text{wahr}$ ), die aber laut Modell nicht besteht ( $r_m = \text{falsch}$ ). Der vierte Fall bezeichnet die Situation, daß weder zwischen den Modellmerkmalen noch den zugeordneten Bildmerkmalen die entsprechende Relation besteht ( $r_m = \text{falsch}, r_i = \text{falsch}$ ). Er unterstützt die Zuordnung um 0,17 Bit.

In den beiden folgenden Unterabschnitten wird erläutert, wie die benötigten a priori Wahrscheinlichkeiten und bedingten Wahrscheinlichkeiten für die zur Repräsentation der Gebäudeansichten verwendeten Relationen bestimmt werden.

### 5.2.1 Berechnung der a priori Wahrscheinlichkeiten

Die a priori Wahrscheinlichkeiten für die Erfüllung bzw. das Vorhandensein einer Relation hängen vom jeweiligen Modell ab. Sie müssen deshalb für jede Gebäudeansicht neu berechnet werden. Dabei müssen für die Relationen unterschiedlicher Kategorien, wie die Merkmalsnachbarschaft, Linienparallelität, Kollinearität usw., jeweils eigene a priori Wahrscheinlichkeiten bestimmt werden.

Zur Herleitung der Wahrscheinlichkeiten wird ein frequentistischer Ansatz gewählt: Es wird für jede Kategorie gezählt, wieviele entsprechende Relationen in der gegebenen Gebäudeansicht erfüllt sind. Des Weiteren wird die Kardinalität der Relation ermittelt, die die maximale Anzahl möglicher Relationen ausdrückt. Die Wahrscheinlichkeit  $P(r_m = \text{true})$  ergibt sich aus dem Verhältnis der Anzahlen von bestehenden zu möglichen Relationen. Im folgenden bezeichnet  $R_m$  die Menge aller möglichen Relationen einer Kategorie (z.B. Linienparallelität) für ein gegebenes Modell (Gebäudeansicht)  $m$ . Die Kardinalität der Relation ist dann durch  $|R_m|$  gegeben.  $R_m^+ \subseteq R_m$  bezeichnet die Menge der vom Modell  $m$  erfüllten Relationen, und  $R_m^- = R_m \setminus R_m^+$  die dazu komplementäre Menge der nicht erfüllten Relationen. Die a priori Wahrscheinlichkeit ergibt sich dann zu

$$P(r_m = \text{wahr}) = \frac{|R_m^+|}{|R_m|} \quad P(r_m = \text{falsch}) = \frac{|R_m^-|}{|R_m|} = 1 - P(r_m = \text{wahr}). \quad (5.12)$$

Für eine gegebene Menge von Modellmerkmalen  $M$  beträgt die Kardinalität einer beliebigen  $n$ -stelligen Relation  $|R_m| = |M|^n$ , weil jedes Merkmal theoretisch mit allen anderen Merkmalen (inklusive sich selbst) in Relation stehen kann. Weist die Relation hingegen bestimmte Struktureigenschaften auf, wie z.B. Symmetrie oder Reflexivität, so ist die Anzahl möglicher Kombinationen geringer. Beispielsweise sind bei irreflexiven Relationen wie der Merkmalsnachbarschaft keine Relationen zwischen denselben Modellmerkmalen definiert, weil ein Modellmerkmal niemals zu sich selber benachbart sein kann. Bei symmetrischen Relationen wie der Linienparallelität ist es gleichgültig, ob Linie  $L_1$  parallel ist zu  $L_2$  oder  $L_2$  zu  $L_1$ ; es handelt sich in beiden Fällen um dieselbe Parallelitätsrelation. Tabelle 5.3 faßt die Eigenschaften und entsprechenden Kardinalitäten der verschiedenen verwendeten Relationen zusammen.

Relation $R_m$	Stelligkeit	Eigenschaften	Kardinalität $ R_m $
MNG(Linie,Punkt)	binär	irreflexiv, asymmetrisch	$ L  \cdot  P $
MNG(Fläche,Linie)	binär	irreflexiv, asymmetrisch	$ F  \cdot  L $
MNG(Fläche,Punkt)	binär	irreflexiv, asymmetrisch	$ F  \cdot  P $
Linienparallelität	binär	reflexiv, symmetrisch, transitiv	$\frac{ L ( L -1)}{2}$
Kollinearität	binär	irreflexiv, asymmetrisch	$ L  \cdot  P $
Same_Side_Lines	ternär	irreflexiv, 2. und 3. Stelle symm.	$\frac{ L ( L -1)( L -2)}{2}$
$\neq$ (Punkte)	binär	irreflexiv, symmetrisch	$\frac{ P ( P -1)}{2}$
$\neq$ (Linien)	binär	irreflexiv, symmetrisch	$\frac{ L ( L -1)}{2}$
$\neq$ (Flächen)	binär	irreflexiv, symmetrisch	$\frac{ F ( F -1)}{2}$

**Tabelle 5.3:** Eigenschaften und Kardinalitäten der verwendeten Relationen.  $|P|$ ,  $|L|$  und  $|F|$  bezeichnen die Anzahlen der Punkte, Linien und Flächen. MNG steht für die Merkmalsnachbarschafts- und  $\neq$  für die Verschiedenheitsrelation.

Bei der Herleitung der relationalen Repräsentation von Gebäudeansichten werden aufgrund des Bildmodells nur Merkmalsnachbarschaften zwischen Punkten und Linien, Linien und Flächen sowie Flächen und Punkten betrachtet (vgl. Abschnitt 3.3.1.1). Da diese drei Sorten von Nachbarschaften unterschiedliche Wahrscheinlichkeiten sowohl für das Auftreten in einem Modell als auch für die Beobachtbarkeit in den Bilddaten aufweisen, werden sie im folgenden separat betrachtet. Diese

Unterteilung ist auch der Grund für die in Tabelle 5.3 festgestellte Eigenschaft der Asymmetrie der Merkmalsnachbarschaftsrelationen<sup>3</sup>.

### 5.2.1.1 Unterscheidung von Relationen zwischen Dach- und Wandmerkmalen

Die Qualität der extrahierten Bildmerkmale unterscheidet sich für Dächer und Wände mitunter erheblich voneinander. Das liegt im wesentlichen an der Aufnahmeperspektive der Luftbilder, bei der vertikale Strukturen, wie beispielsweise die Wände, sehr viel kleiner abgebildet werden als horizontale Strukturen. Ferner werden Dächer weniger stark durch andere Objekte oder Schatten verdeckt als z.B. Wand- und Bodenteile. Die Unschärfe wirkt sich bei Bildmerkmalen mit kleiner Ausdehnung stärker auf die Genauigkeit bzw. Unsicherheit aus als bei größeren Merkmalen (vgl. [Rad93, Fuc97]). Davon ist auch die Beobachtbarkeit von Relationen zwischen den entsprechenden Merkmalen betroffen. Im folgenden werden deshalb alle Relationen der Modelle (Gebäudeansichten) in die zwei Kategorien Dachrelation und Wandrelation unterschieden. Eine Relation gilt als Dachrelation, wenn es sich bei den inzidenten Merkmalen ausschließlich um Dachmerkmale handelt. Für die Flächenmerkmale wird bei der Generierung der Gebäudeansichten explizit vermerkt, ob sie Dach- oder Wandflächen repräsentieren. Für Linien- und Flächenmerkmale läßt sich dies anhand der Flächeninzidenz bestimmen, wobei es sich dann um Dachmerkmale handelt, wenn diese zu einer Dachfläche inzident sind. Die Menge der Dachmerkmale wird im folgenden mit  $M_{\text{Dach}}$  bezeichnet.

Zur Unterscheidung der a priori Wahrscheinlichkeiten zwischen Dach- und Wandmerkmalen wird die Menge der Relationen  $R_m$  in die zwei Teilmengen der Dach- und Wandrelationen partitioniert, d.h.  $R_m = R_{\text{Dach}(m)} \dot{\cup} R_{\text{Wand}(m)}$ , wobei

$$R_{\text{Dach}(m)} = \{r(v_1, \dots, v_n) \mid r(v_1, \dots, v_n) \in R_m : \bigwedge_{i=1}^n v_i \in M_{\text{Dach}}\} \quad (5.13)$$

$$R_{\text{Wand}(m)} = R_m \setminus R_{\text{Dach}(m)}. \quad (5.14)$$

Analog werden die Mengen der im Modell bestehenden Relationen  $R_m^+$  und deren Komplementärmen- gen  $R_m^-$  in Dach- und Wandrelationen unterschieden:

$$R_{\text{Dach}(m)}^+ = R_m^+ \cap R_{\text{Dach}(m)} \quad R_{\text{Wand}(m)}^+ = R_m^+ \cap R_{\text{Wand}(m)} \quad (5.15)$$

$$R_{\text{Dach}(m)}^- = R_m^- \cap R_{\text{Dach}(m)} \quad R_{\text{Wand}(m)}^- = R_m^- \cap R_{\text{Wand}(m)} \quad (5.16)$$

Damit ergeben sich die a priori Wahrscheinlichkeiten schließlich zu

$$P(r_{\text{Dach}(m)} = \text{wahr}) = \frac{|R_{\text{Dach}(m)}^+|}{|R_{\text{Dach}(m)}|} \quad P(r_{\text{Wand}(m)} = \text{wahr}) = \frac{|R_{\text{Wand}(m)}^+|}{|R_{\text{Wand}(m)}|} \quad (5.17)$$

$$P(r_{\text{Dach}(m)} = \text{falsch}) = \frac{|R_{\text{Dach}(m)}^-|}{|R_{\text{Dach}(m)}|} \quad P(r_{\text{Wand}(m)} = \text{falsch}) = \frac{|R_{\text{Wand}(m)}^-|}{|R_{\text{Wand}(m)}|}. \quad (5.18)$$

## 5.2.2 Training der bedingten Wahrscheinlichkeiten

Die bedingten Wahrscheinlichkeiten  $P(r_i|r_m)$  können nach Vosselman auf drei verschiedene Weisen ermittelt werden: (1) durch analytische Berechnung, (2) durch numerische Simulation aller möglichen Situationen und (3) empirisch anhand von Trainingsdaten (vgl. [Vos92]). Die Wahrscheinlichkeit, daß eine Relation zwischen Modellmerkmalen auch zwischen den ihnen zugeordneten Bildmerkmalen beobachtbar ist, hängt von sehr vielen Einflüssen ab, wie z.B. dem Sensor- und

<sup>3</sup>Im Prinzip handelt es sich bei der Merkmalsnachbarschaft um eine symmetrische Beziehung. Bei der Unterscheidung der Merkmalsnachbarschaften in die drei Arten (Punkt, Linie), (Fläche, Linie) und (Fläche, Punkt) werden aber jeweils zwei Merkmale unterschiedlichen Typs zueinander in Relation gesetzt, wodurch die Relationen asymmetrisch werden.

Bildrauschen, der Perspektive, der Beleuchtung, dem Kontrast, der Bodenauflösung und schließlich dem Merkmalsextraktionsverfahren. Da eine Modellierung dieser Einflüsse zu komplex ist, fallen die beiden ersten Methoden zur Bestimmung der bedingten Wahrscheinlichkeit aus, denn sowohl für eine analytische Herleitung als auch für eine numerische Simulation muß der Zusammenhang der verschiedenen Einflüsse bekannt sein<sup>4</sup>. Aus diesem Grund wird  $P(r_i|r_m)$  für alle Relationen anhand von Trainingsdaten geschätzt.

Für die empirische Evaluation werden dieselben manuellen Zuordnungen von Bild- zu Modellmerkmalen herangezogen, die schon zur Bestimmung der Schwellwerte der geometrischen Constraints verwendet wurden. Dabei wurden für die in mehreren Luftbildausschnitten abgebildeten Gebäude die entsprechenden Gebäudeansichten generiert und manuell in das Bild eingepaßt. Anschließend wurden die den Merkmalen der Gebäudeansichten entsprechenden extrahierten Bildmerkmale einander (ebenfalls manuell) zugeordnet (vgl. Abschnitt 3.3.2). Zur Berechnung der bedingten Wahrscheinlichkeiten wird ausgezählt, wieviele Relationen zwischen zugeordneten extrahierten Bildmerkmalen im Verhältnis zu den jeweils korrespondierenden Modellrelationen beobachtet werden. Tabelle 5.4 zeigt die Ergebnisse für die Auswertung von 28 manuell zugeordneten Gebäudemodellen, wobei jedes der Modelle im Durchschnitt ungefähr aus 25 Modellmerkmalen (Punkte, Linien und Flächen) und 230 Relationen (Constraints) bestand.

Relation	$P(r_i=w r_m=w)$	$P(r_i=f r_m=w)$	$P(r_i=w r_m=f)$	$P(r_i=f r_m=f)$
MNG (Linie,Punkt) <sub>Dach</sub>	0,674	0,326	0,007	0,993
MNG (Linie,Punkt) <sub>Wand</sub>	0,533	0,467	0,014	0,986
MNG (Fläche,Linie) <sub>Dach</sub>	0,866	0,134	0,045	0,955
MNG (Fläche,Linie) <sub>Wand</sub>	0,631	0,369	0,094	0,906
MNG (Fläche,Punkt) <sub>Dach</sub>	0,514	0,486	0,022	0,978
MNG (Fläche,Punkt) <sub>Wand</sub>	0,346	0,654	0,046	0,954
Linienparallelität <sub>Dach</sub>	1,000	0,000	0,032	0,968
Linienparallelität <sub>Wand</sub>	1,000	0,000	0,069	0,931
Kollinearität <sub>Dach</sub>	1,000	0,000	0,063	0,937
Kollinearität <sub>Wand</sub>	1,000	0,000	0,105	0,895
Same_Side_Lines <sub>Dach</sub>	0,967	0,033	0,737	0,263
Same_Side_Lines <sub>Wand</sub>	0,842	0,158	0,756	0,244
$\neq$ (Punkte) <sub>Dach</sub>	0,993	0,007		
$\neq$ (Punkte) <sub>Wand</sub>	0,990	0,010		
$\neq$ (Linien) <sub>Dach</sub>	1,000	0,000		
$\neq$ (Linien) <sub>Wand</sub>	1,000	0,000		
$\neq$ (Flächen) <sub>Dach</sub>	0,980	0,020		
$\neq$ (Flächen) <sub>Wand</sub>	1,000	0,000		

**Tabelle 5.4:** Die bedingten Wahrscheinlichkeiten wurden anhand von 28 manuell zugeordneten Gebäudeansichten geschätzt. Die Bodenauflösung der Bilddaten betrug 15 cm, was die verhältnismäßig schlechten Übertragungswahrscheinlichkeiten für Wandrelationen erklärt. Da die Verschiedenheitsrelation  $\neq$  total ist, d.h. jede Modellvariablen mit jeder anderen in Relation steht, können die Fälle  $r_m = falsch$  nicht auftreten. Dies erklärt die fehlenden Werte in den beiden rechten Spalten.

In der Tabelle fallen zunächst die leeren Einträge der Verschiedenheitsrelationen ( $\neq$ ) auf. Die Wahrscheinlichkeiten für die Fälle  $r_m = falsch$  sind nicht definiert, weil alle Modellmerkmale paarweise voneinander verschieden sein müssen und deshalb alle möglichen Modellrelationen auch tatsächlich vorhanden sind. Es handelt sich hierbei um eine *totale* Relation, bei der für alle möglichen Modelle gilt:  $R_m = R_m^+$  und  $R_m^- = \emptyset$ .

<sup>4</sup>Für das verwendete Verfahren zur Merkmalsextraktion (siehe Abschnitt 2.5.1) liegt bislang nur eine erste Studie vor, die die Qualität von Linienextraktionen bzgl. eines gegebenen Flächenkontrastes schätzt [FLF94].

Eine bedingte Wahrscheinlichkeit von 1, 0 in der ersten Tabellenspalte besagt, daß wenn eine solche Relation im Modell wahr ist, diese immer auch im Bild erfüllt wird, sofern alle inzidenten Bildmerkmale beobachtbar sind. Da die Schwellwerte der geometrischen Constraints anhand derselben Zuordnungen der Trainingsdaten ermittelt worden sind, müssen die bedingten Wahrscheinlichkeiten für die beiden Relationen Linienparallelität und Kollinearität sowohl für die Beziehungen zwischen Dach- als auch zwischen Wandmerkmalen jeweils genau 1, 0 betragen.

Insgesamt läßt sich an den ermittelten Wahrscheinlichkeiten erkennen, daß die topologischen Relationen der Merkmalsnachbarschaft generell störanfälliger sind als die geometrischen Relationen oder die Verschiedenheitsrelationen. Dies hängt im wesentlichen mit der Fragmentierung der Bildmerkmale zusammen, wie bereits in Abschnitt 4.1 erläutert wurde. An den großen Differenzen zwischen den Wahrscheinlichkeiten für Dach- und Wandrelationen läßt sich ferner erkennen, daß sie — wie erwartet — eine unterschiedliche Qualität besitzen, und deshalb die Unterscheidung notwendig ist.

### 5.3 Kopplung mit den erweiterten Constraint-Techniken

In Kapitel 4 wurden die zur Repräsentation und Suche von Gebäudeansichten verwendeten Constraints  $c(v_1, \dots, v_n)$  jeweils um eine Statusvariable  $b \in \{-1, 0, 1\}$  zu Constraints  $c'(v_1, \dots, v_n, b)$  ergänzt, um die Auswirkungen der Unbeobachtbarkeit und Unschärfe explizit zu berücksichtigen. Dabei wird zwischen den drei Zuständen erfüllt ( $b = 1$ ), nicht beobachtbar ( $b = 0$ ) und verletzt ( $b = -1$ ) differenziert. Der erste und der dritte Fall entsprechen genau den Situationen, in denen die Modellrelation erfüllt, und die Bildrelation ebenfalls erfüllt ( $r_i = \text{wahr}, r_m = \text{wahr}$ ) bzw. nicht erfüllt ( $r_i = \text{falsch}, r_m = \text{wahr}$ ) ist. Zur Gewichtung der Constraints werden folglich die Werte  $-1$  und  $1$  durch die gemeinsame Information  $I(r_i = \text{wahr}; r_m = \text{wahr})$  bzw.  $I(r_i = \text{falsch}; r_m = \text{wahr})$  ersetzt (vgl. mit den Werten der linken Spalte von Tabelle 5.2 auf Seite 91).

Bislang wurden Constraints nur für die erfüllten Modellrelationen  $R_m^+$  generiert. Zur Bestimmung der wahrscheinlichsten Zuordnung  $\hat{h}$  müssen nach Gleichung 5.7 in die Bewertung aber auch die nicht erfüllten Modellrelationen  $R_m^-$  eingehen. Dazu wird zu der Menge  $R_m^-$  die entsprechende Menge der komplementären Constraints  $\bar{C}$  bestimmt, die besagen, daß die zwischen Modellmerkmalen nicht erfüllten Relationen zwischen den ihnen zugeordneten Bildmerkmalen ebenfalls nicht erfüllt sein dürfen. Wenn beispielsweise zwei Linien im Modell nicht zueinander parallel sind, sollten es die den Modelllinien zugeordneten extrahierten Linienmerkmale auch nicht sein. Die Werte  $-1$  und  $1$  der Statusvariablen der komplementären Constraints werden durch die gemeinsame Information  $I(r_i = \text{wahr}; r_m = \text{falsch})$  bzw.  $I(r_i = \text{falsch}; r_m = \text{falsch})$  ersetzt (vgl. mit den Werten der rechten Spalte von Tabelle 5.2).

Eine interessante Besonderheit ergibt sich bei der Gewichtung totaler Relationen, d. h. bei Relationen, die für alle möglichen Relationentupel erfüllt sind ( $R_m^+ = R_m$ ). Hier ergibt sich die gemeinsame Information  $I(r_i; r_m = \text{true})$  unabhängig von den bedingten Wahrscheinlichkeiten für sowohl  $r_i = \text{true}$  als auch  $r_i = \text{false}$  zu 0. Der Grund dafür kann an dem Fall ( $r_i = \text{wahr}, r_m = \text{wahr}$ ) verdeutlicht werden. Aus Gleichung 5.11 ergibt sich für diesen Fall

$$I(r_i = w; r_m = w) = \log_2 \frac{P(r_i = w | r_m = w)}{P(r_i = w | r_m = w)P(r_m = w) + P(r_i = w | r_m = f)P(r_m = f)} \quad [\text{Bit}]. \quad (5.19)$$

Da  $P(r_m = \text{wahr}) = 1$  und somit  $P(r_m = \text{falsch}) = 0$ , resultiert daraus

$$I(r_i = w; r_m = w) = \log_2 \frac{P(r_i = w | r_m = w)}{P(r_i = w | r_m = w)} = 0 \quad [\text{Bit}]. \quad (5.20)$$

Die gleiche Situation ergibt sich auch für den Fall ( $r_i = \text{falsch}, r_m = \text{wahr}$ ). Das bedeutet, daß totale Relationen, wie die Verschiedenheitsrelation, niemals eine Zuordnung befürworten noch ihr widersprechen. Sie können aus diesem Grund in der Bewertungsfunktion weggelassen werden.

### 5.3.1 Suchraumeinschränkungen

Obwohl die Verschiedenheitsrelation keine Rolle bei der Bewertung spielt, wird sie dennoch weiterhin zur aktiven Suchraumeinschränkung verwendet. Den bedingten Wahrscheinlichkeiten in Tabelle 5.4 auf Seite 94 läßt sich entnehmen, daß die Verschiedenheitsrelation für Linienmerkmale immer erfüllt ist, d.h. es ist bei den trainierten Zuordnungen keine Situation aufgetreten, bei der zwei verschiedenen Modelllinien dasselbe aus dem Bild extrahierte Linienmerkmal zugeordnet wurde. Die Verletzung der entsprechenden Constraints wird deshalb nicht erlaubt. Dazu wird von vornherein für alle Verschiedenheits-Constraints jeweils der Wert für die Verletzung des Constraints aus dem Wertebereich der Statusvariablen  $b$  entfernt (vgl. Abschnitt 4.3.5). Die Relaxation des Constraints kann damit nur durch den Fall der Unbeobachtbarkeit eines inzidenten Merkmals geschehen. Damit zwischen der Erfüllung des Constraints und der Relaxation aufgrund von Unbeobachtbarkeit unterschieden werden kann, wird die Erfüllung nicht wie in Gleichung 5.20 beschrieben mit 0, sondern mit einer Zahl größer 0 bewertet, die so klein ist, daß die Auswirkungen auf den gemeinsamen Informationswert der Gesamtbewertungsfunktion vernachlässigbar klein sind, und sie praktisch keinen Einfluß auf die Wahl der Zuordnung hat.

Mit den Verschiedenheitsrelationen zwischen Punkten ( $\neq (\text{Punkte})_{\text{Wand}}$  und  $\neq (\text{Punkte})_{\text{Dach}}$ ) sowie zwischen Wandflächen ( $\neq (\text{Flächen})_{\text{Wand}}$ ) wird analog verfahren. Die bedingten Wahrscheinlichkeiten  $P(r_i = \text{wahr} | r_m = \text{wahr})$  betragen für die erstgenannten Relationen zwar nur 0,99; dies kann jedoch schlimmstenfalls dazu führen, daß wenn zwei Modellmerkmalen dasselbe Bildmerkmal entspricht, einem von beiden ein Fehlwert zugeordnet werden muß.

Zusätzlich zu diesen Einschränkungen wird auch weiterhin die Verletzung der geometrischen Kollinearitäts- und Linienparallelitäts-Constraints durch das Entfernen des Wertes  $I(r_i = \text{falsch}; r_m = \text{wahr})$  aus den Wertebereichen der jeweiligen Statusvariablen untersagt (vgl. Abschnitt 4.3.5).

### 5.3.2 Heuristiken

Zur Berechnung der wahrscheinlichsten Zuordnung von Bild- zu Modellmerkmalen muß, wie am Anfang von Abschnitt 5.3 erläutert wurde, auch die Menge der komplementären Constraints bei der Suche berücksichtigt werden. Die Menge der bisher generierten Constraints  $C$  entspricht allerdings nur den erfüllten Modellrelationen  $R_m^+$ ; die nicht erfüllten Modellrelationen  $R_m^-$  sind bislang unberücksichtigt. Würden sie einbezogen, indem für jede Relation  $r_m(v_1, \dots, v_n) \in R_m^-$  ein (negierter) Constraint der Form  $\neg c(v_1, \dots, v_n, b)$  generiert würde, entstünde insgesamt eine Constraint-Menge, deren Größe der Summe der Kardinalitäten aller in Tabelle 5.3 auf Seite 92 aufgeführten Relationen entspräche. Die Anzahl der Constraints würde damit quadratisch in der Anzahl der Modellmerkmale wachsen, und bereits Gebäudeansichten einfacher Haustypen, wie z.B. Satteldachhäuser, würden anstelle von durchschnittlich 230 durch mehrere tausend Constraints beschrieben.

Die Anzahl der Constraints hat einen wesentlichen Einfluß auf die Effizienz der Suche, weil zur Berücksichtigung der Unbeobachtbarkeit und der Unschärfe mit jedem Constraint auch eine dreiwertige Statusvariable  $b$  hinzukommt (vgl. Abschnitt 4.3.5). Ferner stellt jeder Constraint in CLP eine reaktive Datenstruktur dar, mit deren Auswertung ein bestimmtes Maß an Verwaltungsaufwand verbunden ist. Aus diesem Grund werden auch künftig die komplementären Constraints nicht in die Suche nach der besten Zuordnung von Bild- zu Modellmerkmalen einbezogen. Obwohl aufgrund dieser Heuristik nicht mehr garantiert werden kann, daß die beste Zuordnung auch die wahrscheinlichste ist, zeigen die korrekten Zuordnungsergebnisse der Beispiele der vorangegangenen Kapitel die praktische Anwendbarkeit dieses Vorgehens. Um die vollständige gemeinsame Information  $I_h$  zwischen Modell und Bild zu erhalten, wird nach der Bestimmung der besten Zuordnung noch die gemeinsame Information für alle komplementären Constraints berechnet und zu dem Wert der Bewertungsfunktion hinzuaddiert. Auf diese Weise wird der Informationsgehalt der komplementären Constraints zwar nicht bei der Suche berücksichtigt, geht aber dennoch in die abschließende Gesamtbewertung ein, die später in Kapitel 6 noch zur Bestimmung des wahrscheinlichsten Modells benötigt wird.

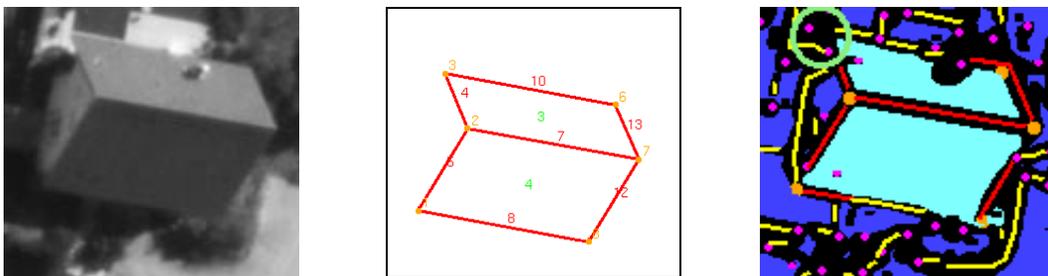
Während der Suche wird für die Auswahl des Modellmerkmals, dem als nächstes ein passendes Bildmerkmal zugeordnet werden soll, bislang die First-Fail-Heuristik verwendet. Diese Heuristik orientiert sich an der Größe der Wertebereiche der Variablen, und nicht an deren Bedeutung oder Wichtigkeit in bezug auf die Bewertungsfunktion. Durch die Gewichtung der Constraints ist es nun möglich, andere Heuristiken zu verwenden, die sich an dem Zuwachs der gemeinsamen Information orientieren. Bei den *Best-First-Heuristiken* (vgl. [Nil82, AHU87]) wird generell versucht, durch die geeignete Auswahl des Modellmerkmals, dem als nächstes ein Bildmerkmal zugeordnet werden soll, den Wert der Bewertungsfunktion größtmöglich zu steigern. Dies hat in Verbindung mit dem Branch & Bound-Ansatz den Vorteil, daß sobald eine vollständige Zuordnung bestimmt wurde, die untere Bewertungsschranke für den nächsten Suchlauf um einen großen Betrag nach oben verschoben wird, und damit insgesamt wahrscheinlich weniger Suchläufe zur Bestimmung der optimalen Lösung benötigt werden (vgl. Abschnitt 4.3.4).

Eine spezielle Variante der Best-First-Suchstrategie ist der sogenannte *A\*-Algorithmus* (vgl. [Nil82, RN95]), der neben dem unmittelbaren Zuwachs der Bewertung durch die Zuordnung eines Bildmerkmals zu der aktuellen Variable, den noch maximal möglichen erreichbaren Zuwachs der noch ausstehenden Zuordnungen abschätzt. Damit wird verhindert, daß am Anfang der Suche zwar Zuordnungen ausgewählt werden, die den Wert der Bewertungsfunktion maximal steigern, aber schließlich nicht zum optimalen Ergebnis führen könnten, weil aufgrund der anfänglichen Entscheidungen später nur noch schlechtbewertete Zuordnungen möglich sein könnten. Die wesentliche Schwierigkeit bei der Verwendung des A\*-Algorithmus besteht in der Definition einer Funktion, die an jeder Stelle der Suche den noch künftig zu erwartenden Zuwachs abschätzt. Vosselman beschreibt in [Vos92] die Verwendung eines (modifizierten) A\*-Algorithmus für die Bestimmung der besten relationalen Zuordnung, und zeigt, wie der jeweils noch zu erwartende Informationszuwachs der künftigen Zuordnungen abgeschätzt werden kann.

Eine umfassende Untersuchung über die Verwendbarkeit dieser Heuristiken und deren Auswirkungen auf die jeweils benötigte Rechenzeit liegt nicht im Rahmen dieser Arbeit, denn durch die Verwendung der Constraint-Techniken können bereits mit der einfachen First-Fail-Heuristik die meisten Zuordnungen in Sekunden oder wenigen Minuten bestimmt werden. Es steht jedoch zu erwarten, daß die zusätzliche Einbeziehung der beschriebenen Heuristiken den Suchaufwand nochmals deutlich reduzieren würde.

### 5.3.3 Praktische Umsetzung

Der verwendete Constraint-Interpreter von CLP(FD) arbeitet auf ganzzahligen Wertebereichen (vgl. Abschnitt 3.2.3). Aus diesem Grund werden die Werte der gemeinsamen Information jeweils durch die maximal gewünschte Auflösung  $\varepsilon$  dividiert und dann gerundet.  $\varepsilon$  hat für alle folgenden Beispiele dieses und des nächsten Kapitels den Wert 0,001, d.h. die Genauigkeit der Bewertungsfunktion ist für die Beispiele auf  $\frac{1}{1000}$  Bit beschränkt.



**Abbildung 5.4:** Ergebnis der Zuordnung von Satteldachhypothese und Bildmerkmalen, wobei die Constraints mit der gemeinsamen Information gewichtet wurden.

Abschließend wird noch einmal das Beispiel vom Kapitelanfang aufgegriffen (Abbildung 5.1 auf Seite 85), bei dem die Ansicht einer Satteldachhypothese in einem Luftbildausschnitt mit ei-

nem Satteldachhaus zu identifizieren war. Aufgrund der ungewichteten Constraints wurde ein Dacheckpunkt, der im Bild zu weit von der Dachfläche entfernt lag, dennoch zugeordnet, weil er die Kollinearitäts-Constraints erfüllen konnte.

Abbildung 5.4 zeigt, daß unter Berücksichtigung der Gewichtung korrekt erkannt wurde, daß der linke obere Dacheckpunkt nicht extrahiert wurde und unbeobachtbar ist. Die gesamte gemeinsame Information  $I_h$  der Zuordnung beträgt 78,4 Bit, wobei 42,99 Bit von den für die Suche verwendeten Constraints und 35,41 Bit von den danach ausgewerteten komplementären Constraints stammen (vgl. Abschnitt 5.3.2).

## Kapitel 6

# Bestimmung des wahrscheinlichsten Modells

In den drei vorangegangenen Kapiteln wurde bislang gezeigt, wie man für eine einzelne Gebäudeansicht mittels erweiterter Constraint-Techniken auf effiziente Weise die beste Zuordnung von Bild- zu Modellmerkmalen bestimmen kann. Durch die Anwendung eines informationstheoretisch begründeten Ähnlichkeitsmaßes wird sichergestellt, daß es sich bei der besten Zuordnung auch um die wahrscheinlichste Zuordnung im Sinne einer Maximum-Likelihood-Schätzung handelt.

Nun kommen bei der Erkennung eines Gebäudes oftmals mehrere verschiedene Gebäudehypothesen in Frage. Diese Hypothesen können darüberhinaus jeweils in verschiedenen möglichen Ansichten im Bild erscheinen (vgl. Abschnitt 2.5.6). Zur Klassifikation des abgebildeten Gebäudes muß somit entschieden werden, um welche Gebäudeansicht es sich am wahrscheinlichsten bzgl. der gegebenen Bilddaten handelt. Wenn die verschiedenen Modelle durch  $M_i$  und die Bilddaten durch  $D$  repräsentiert werden, läßt sich diese Aufgabe formal beschreiben durch

$$\hat{M} = \arg \max_{M_i} P(M_i|D). \quad (6.1)$$

Dieses Vorgehen wird als *Maximum A Posteriori-Schätzung* (MAP) bezeichnet (vgl. [Koc97]), weil sich  $P(M_i|D)$  nach dem Bayesschen Theorem als a posteriori Wahrscheinlichkeit aus den a priori Wahrscheinlichkeiten der Daten  $P(D)$  und des Modells  $P(M_i)$  sowie der bedingten Wahrscheinlichkeit  $P(D|M_i)$  ergibt (siehe [Spi93, Pea88]):

$$P(M_i|D) = \frac{P(D|M_i)P(M_i)}{P(D)} \quad (6.2)$$

Die a priori Wahrscheinlichkeit der Daten  $P(D)$  hängt nur von den Bilddaten ab und ist für alle Modelle, die in demselben Bild gesucht werden, konstant. Sie kann deshalb bei der Maximierung vernachlässigt werden, und es ergibt sich ausgehend von Gleichung 6.1 und 6.2:

$$\hat{M} = \arg \max_{M_i} P(D|M_i)P(M_i). \quad (6.3)$$

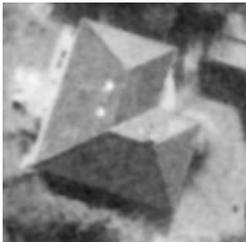
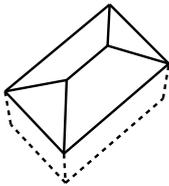
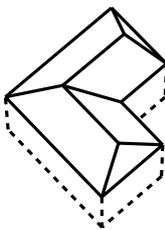
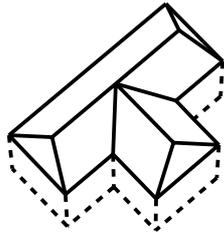
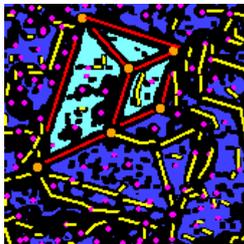
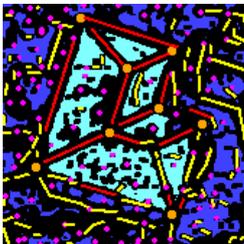
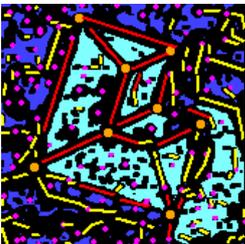
Diese Formel bildet den Ausgangspunkt für den weiteren Verlauf des Kapitels, das sich im wesentlichen in drei Teile gliedert: Zuerst wird gezeigt, wie sich die im vorigen Kapitel vorgestellte Bewertungsfunktion unter Verwendung des *Prinzips der minimalen Beschreibungslänge* so erweitern läßt, daß sie die MAP-Strategie realisiert und damit auch zwischen verschiedenen Modellen unterscheiden kann. Dabei wird es hauptsächlich darum gehen, wie die Komplexität der verschiedenen Modelle adäquat einzubeziehen ist.

Die Messung der Modellkomplexität wird über die Beschreibungslänge der Modelle erfolgen, die dazu optimal codiert sein müssen. Im zweiten Teil dieses Kapitels geht es deshalb um die Entwicklung eines Codierschemas für die Gebäudeansichten, das eine sehr kompakte Darstellung der Modelle ermöglicht.

Im dritten Teil wird die Qualität der vorgeschlagenen Codierung anhand von praktischen Tests untersucht. Dazu werden Klassifikationen unterschiedlicher Gebäudetypen und -ansichten in drei Szenarien mit synthetischen Bildern durchgeführt und diskutiert. Am Ende des Kapitels, das auch gleichzeitig den Abschluß dieser Arbeit darstellt, bilden zwei an realen Luftbildausschnitten durchgeführte Klassifikationen, welche die Anwendbarkeit des entwickelten Konzepts auf reale Daten beispielhaft demonstrieren sollen.

## 6.1 MAP-Klassifikation mit Hilfe des Prinzips der minimalen Beschreibungslänge

Die in den beiden vorangegangenen Kapiteln vorgestellte Methode für die Bestimmung inexakter Matchings findet zu jeder Gebäudehypothese immer eine beste Zuordnung<sup>1</sup>. Während sich die Zurückweisung völlig unpassender Zuordnungen, wie in Abschnitt 5.3.3 erläutert, durch einen Hypothesentest realisieren läßt, stellt sich nunmehr die Frage, wie aus einer Menge von Modellen, deren Zuordnungen zu den Bilddaten den Hypothesentest erfolgreich passiert haben, das wahrscheinlichste ausgewählt wird.

Bild	Modelle		
			
zugeordnete Bildmerkmale des jeweils besten Matchings $\hat{h}$			
$I_{\hat{h}}(\text{Modell}; \text{Bild})$ [Bit]	85,2	224,24	237,56

**Abbildung 6.1:** Ergebnisse der Zuordnungen von drei verschiedenen Ansichten unterschiedlicher Gebäudehypothese zu demselben Bildausschnitt. Es läßt sich erkennen, daß die gemeinsame Information der besten Zuordnung mit der Modellkomplexität wächst. Sie kann deshalb nicht unmittelbar zur Auswahl des wahrscheinlichsten Modells verwendet werden (das mittlere Modell wäre korrekt, aber das rechte weist den größten Informationswert auf).

Die Bewertungsfunktion mißt den Grad der Gemeinsamkeit durch Aufsummierung aller Einzelgemeinsamkeiten (vgl. Formel 5.7 auf Seite 89). Wenn ein Modell, das komplexer ist als das im Bild abgebildete Haus, den Bilddaten zugeordnet wird, besteht eine hohe Wahrscheinlichkeit, daß „überschüssige“ Ansichtsmerkmale des Modells Bildmerkmalen zugeordnet werden, die zwar nicht zu dem abgebildeten Haus gehören, jedoch zufällig ein paar der geforderten Relationen erfüllen. Das bedeutet wiederum, daß der zu erwartende Wert für die gemeinsame Information des besten Matchings für ein komplexes Modell generell höher sein wird als der eines einfacheren Modells.

<sup>1</sup>Eine Zuordnung, die immer möglich ist, besteht darin, daß jedem Modellmerkmal ein Wildcard zugeordnet wurde. Die gemeinsame Information ist in diesem Fall gleich 0.

Abbildung 6.1 demonstriert diese Situation anhand dreier Zuordnungen verschiedener Gebäudeansichten zu demselben Bildausschnitt.

Das bedeutet, daß man nicht einfach das beste Modell durch Vergleich der gemeinsamen Information der einzelnen besten Zuordnungen ermitteln kann. Offensichtlich muß die Bewertung entsprechend dem jeweiligen Modell geeignet normalisiert werden, um die Vergleichbarkeit des besten Matchings verschiedener und unterschiedlich komplexer Modelle zu erreichen. Ein Ansatz, der genau dies realisiert, wird im folgenden vorgestellt.

### 6.1.1 Das Prinzip der minimalen Beschreibungslänge

Das *Prinzip der minimalen Beschreibungslänge* (englisch: *minimum description length principle*, künftig abgekürzt als MDL-Prinzip) ist ein von Rissanen entwickeltes, allgemeines statistisches Verfahren zur Klassifikation (siehe [Ris83, Ris87]). Man kann es als mathematische Umsetzung des von dem Philosophen Ockham im 14. Jahrhundert formulierten und nach ihm benannten Gesetz *Ockham's Razor* betrachten: „Ohne Zwang sollte man keine Vielheiten annehmen.“ (vgl. [Mac98]). Dieses Prinzip besagt, daß eine Theorie (in unserem Fall ein Modell) nicht aufwendiger sein sollte, als es zur Erklärung der Beobachtungen notwendig ist.

Rissanen hat dieses Prinzip auf die Statistik übertragen, in dem er zeigt, daß jene Klassifikation die beste ist, deren Summe aus der Beschreibungslänge des Modells und der Beschreibungslänge notwendiger Erklärungen einzelner Abweichungen (z.B. Sonderfälle) von diesem Modell minimal ist. Zur Formalisierung der Beschreibungslänge wird auf die Codierungstheorie, ein Teilgebiet der Informationstheorie, zurückgegriffen, und es ergibt sich die folgende formale Darstellung des MDL-Prinzips:

$$\hat{M} = \arg \min_{M_i} L(D|M_i) + L(M_i) \quad (6.4)$$

Dabei bezeichnet

- $L(D|M_i)$  die Beschreibungslänge der Abweichungen der Daten  $D$  vom Modell  $M_i$ , und
- $L(M_i)$  die Beschreibungslänge des Modells  $M_i$ .

Aus der Codierungstheorie ist bekannt, daß der Informationsgehalt eines Symbols genau seiner Codierungslänge entspricht, falls alle Symbole optimal codiert werden (siehe [Ing71] und folgenden Abschnitt 6.1.2). Unter der Annahme einer optimalen Codierung kann Gleichung 6.4 somit auch formuliert werden als

$$\hat{M} = \arg \min_{M_i} I(D|M_i) + I(M_i) \quad (6.5)$$

Laut Definition der Information ist  $I(M_i) = -\log_2 P(M_i)$  und  $I(D|M_i) = -\log_2 P(D|M_i)$ . Damit entspricht die Minimierung von Gleichung 6.5 der Maximierung von

$$\hat{M} = \arg \max_{M_i} P(D|M_i)P(M_i). \quad (6.6)$$

Diese Gleichung ist identisch mit Gleichung 6.3 zur Bestimmung des wahrscheinlichsten Modells mit der größten a posteriori Wahrscheinlichkeit. Das Prinzip der minimalen Beschreibungslänge realisiert also einen MAP-Schätzer, unter der Annahme, daß die Modelle optimal codiert sind.

Bevor nun näher auf die optimale Codierung eingegangen wird, soll an dieser Stelle noch auf zwei Eigenschaften hingewiesen werden, die die Universalität des MDL-Prinzips verdeutlichen:

1. Bei konstanter Beschreibungslänge des Modells entspricht das MDL-Prinzip der Maximum-Likelihood-Schätzung (vgl. [DH73, Ris87]).
2. Es wurde von Feder gezeigt ([Fed86]), daß das Maximum-Entropy-Verfahren, ein anderer häufig verwendeter statistischer Ansatz zur Klassifikation, einen Spezialfall des MDL-Prinzips darstellt.

### 6.1.2 Optimale Codes

Sei  $A = \{a_1, \dots, a_N\}$  eine Menge von Symbolen. Ein Code  $C$  ordnet nun den Symbolen aus  $A$  Codewörter, die aus Bitvektoren variabler Länge bestehen, zu. Ein Code ist *eindeutig dekodierbar*, wenn jedem Symbol genau ein Codewort  $c(a_i) \in \{0, 1\}^+$  zugeordnet wird, das von allen anderen verschieden ist.  $L(a_i)$  bezeichnet dabei die Länge der jeweiligen Codewörter. Ein Code erfüllt darüberhinaus die sogenannte *Präfixeigenschaft*, wenn kein Codewort Präfix eines anderen ist. Präfixcodes sind unmittelbar dekodierbar, d.h. zur Dekodierung eines Symbols muß immer nur das aktuelle Codewort betrachtet und keine Vorausschau über anschließende Codewörter erfolgen. Präfixcodes erlauben also die Codierung ganzer Symbolfolgen durch Aneinanderfügung der einzelnen Codeworte, ohne daß zur Dekodierung zusätzliche Trennzeichen eingefügt werden müßten.

**Definition 6.1 (Optimaler Code [Ing71])**

Ein Code ist optimal gdw.

1. er unmittelbar dekodierbar ist, d.h. die Präfixeigenschaft erfüllt, und
2. seine mittlere Länge  $\bar{L} = \sum_{a_i \in A} P(a_i)L(a_i)$  minimal ist, wobei  $P(a_i)$  die Auftretenswahrscheinlichkeit des Symbols  $a_i$  angibt. □

Der Nachweis der Präfixeigenschaft eines Codes kann anhand der *Kraftschen Ungleichung*

$$\sum_{a_i \in A} 2^{-L(a_i)} \leq 1 \quad (6.7)$$

erfolgen, die eine notwendige und hinreichende Bedingung für die eindeutige Dekodierbarkeit eines Codes ist. Bleibt noch die Frage zu klären, wie die zweite Bedingung zu erfüllen ist. Shannon hat bereits 1948 in seinem Aufsatz [Sha48] gezeigt, daß die Entropie  $H(A)$  die untere Schranke für die minimale mittlere Codierungslänge  $\bar{L}$  bildet (siehe auch [Ing71]):

$$\underbrace{- \sum_{a_i \in A} P(a_i) \log_2 2^{-L(a_i)}}_{\bar{L}} \geq \underbrace{- \sum_{a_i \in A} P(a_i) \log_2 P(a_i)}_{H(A)} \quad (6.8)$$

Die linke Seite der Ungleichung kann höchstens so klein werden, wie die rechte Seite, und zwar ist das genau dann der Fall, wenn alle Symbole  $a_i$  durch Codewörter der Länge  $-\log_2 P(a_i)$  codiert werden, was der Definition des Informationsgehaltes der Symbole entspricht. Das bedeutet, daß die Information eines Symbols nicht nur ein Maß für die Überraschung ist; sie entspricht auch ihrer Codierungslänge, wenn sie optimal codiert wird.

Auf dieser Erkenntnis basieren im übrigen auch die zur Kompression von Daten oftmals verwendeten Huffman-Codes. Sie sortieren die vorkommenden Symbole nach ihrer Auftretenshäufigkeit und verwenden für die Symbole mit absteigender Häufigkeit längere Codes.

### 6.1.3 Anwendungen des MDL-Prinzips zur Klassifikation und Objekterkennung

Das MDL-Prinzip wurde im Bereich der Photogrammetrie und der Objekterkennung seit dem Ende der 80er Jahre in einer Reihe von Anwendungen erfolgreich eingesetzt, von denen einige hier kurz vorgestellt werden sollen. Auf zwei Anwendungen im Bereich des maschinellen Lernens wird später noch eingegangen.

In ihrem Artikel [GW84] entscheiden Georgeff und Wallace anhand der Beschreibungslänge über das optimale Modell zur Interpretation von Punktmengen. Dabei wird versucht, eine Menge einzelner Punkte durch eine Menge verschiedener Hypothesen wie z.B. „die Punkte liegen auf einer, zwei

oder mehr Linien“, oder „sie sind zufällig verteilt“ zu erklären. Die Erklärung mit der kürzesten Beschreibungslänge wird als die beste angenommen.

Förstner greift in [För89] diese Idee auf und stellt eine kompakte Codierung sowohl für die Hypothese einer Linie mit Ausreißerpunkten als auch die Hypothese einer zufällig verteilten Punktmenge vor. Mit dieser Codierung können dann Punktmengen unter Verwendung des MDL-Prinzips robust klassifiziert werden. Während die von Förstner vorgeschlagene Codierung von der Auflösung der Punktkoordinaten abhing, zeigte Vosselman später in [Vos92], daß die Auflösung bei Verwendung eines von Rissanen vorgeschlagenen allgemeinen Codierschemas für Modellparameter (vgl. [Ris83]) keine Rolle bei der Klassifikation spielt.

Fua und Hanson beschreiben in [FH89a, FH89b] ein Verfahren zur Extraktion von Gebäuden, wobei das MDL-Kriterium zur Auswahl aus verschiedenen alternativen Hypothesen verwendet wird. Hausdächer werden generisch als Flächen geschlossener rechtwinkliger Polygone modelliert, was zur Beschreibung von Gebäuden mit flachen Dächern wie z.B. Hochhäuser und Verwaltungsgebäude ausreicht. Zur Bestimmung der Hypothesen wird zunächst eine Kantenextraktion durchgeführt. Anschließend wird versucht aus kollinearen und zueinander orthogonalen Kanten größere Polygone zu konstruieren, die dann Hypothesen für Hausdächer darstellen. Die Bewertungsfunktion summiert nun auf der einen Seite für jeden von einem Polygon eingeschlossenen Bildpunkt ein auf der gemeinsamen Information basierendes Flächenhogenitätsmaß auf. Auf der anderen Seite wird davon wiederum die Beschreibungslänge des Modells subtrahiert. Für jeden Bildbereich, zu dem mehrere sich überlappende und ausschließende Hypothesen bestimmt wurden, wird schließlich das Polygon mit der größten Gesamtbewertung als wahrscheinlichste Erklärung der beobachteten Bilddaten angesehen und als Ergebnis der Gebäudeextraktion geliefert.

In [BWF95, WF95, Wei95] beschreiben Weidner, Brunn und Förstner einen Ansatz zur MDL-basierten Formrekonstruktion, der später von Weidner zur Rekonstruktion von Gebäuden aus Digitalen Oberflächenmodellen verwendet wurde (siehe [Wei97b] und Abschnitt 2.3.1). Dabei geht es um die modellbasierte Bestimmung von Flächenumrandungen, die auch hier als geschlossene Polygone, die jedoch nicht notwendigerweise rechteckig sein müssen, modelliert sind. Nach der Extraktion der Kantenpixel einer Fläche in einem Binärbild, wird versucht, diese Kantenpixel durch so wenige gerade Linien wie möglich zu repräsentieren. Zur Vereinfachung der Form werden Annahmen über die Parallelität, Kollinearität und Orthogonalität von Polygonlinien gemacht. Das MDL-Kriterium balanciert nun zwischen der Einfachheit der rekonstruierten Form auf der einen Seite und der Abweichung dieser Form von den Daten auf der anderen Seite.

Wie an den vorgestellten Arbeiten deutlich wird, liegt einer der Hauptvorteile des MDL-Prinzips darin, daß über die Beschreibungslänge die verschiedenartigen Erklärungen von Beobachtungen durch unterschiedliche Modelle unmittelbar miteinander verglichen werden können. Darüberhinaus ist es immer möglich, ein Codiervorgehen für ein Modell anzugeben, dessen Optimalität zwar nicht unbedingt bewiesen werden kann, welches aber als Näherung für die wahre Codierungslänge verwendet werden kann. Dagegen kann es schwierig oder gar unmöglich sein, direkt die a priori Wahrscheinlichkeit eines Modells zu schätzen.

Keines der vorgestellten Konzepte kann direkt auf die Klassifikation der Gebäudeansichten übertragen werden, weil deren relationale Struktur deutlich komplexer ist, als die oben beschriebenen Modelle. Allerdings bietet das Konzept von Fua und Hanson einen interessanten Ansatzpunkt, da sie ebenfalls auf der einen Seite die Ähnlichkeit zwischen Modell und Bild über die gemeinsame Information messen und auf der anderen Seite die Modellkomplexität über seine Beschreibungslänge berücksichtigen. Im folgenden Abschnitt wird nun gezeigt, wie analog zu diesem Ansatz die MAP-Strategie für die Bestimmung der wahrscheinlichsten Gebäudeansicht realisiert werden kann.

### 6.1.3.1 Erweiterung der Matching-Bewertungsfunktion

In dem Artikel, in dem Fua und Hanson die der oben vorgestellten Anwendung zugrundeliegende Theorie erläutern ([FH89b]), zeigen sie, wie die wahrscheinlichste Modellhypothese mittels des

Bayesschen Theorems formuliert werden kann. Dabei sei  $M_i$  ein Modell, bestehend aus Modellteilen  $M_i = \{m_1, \dots, m_n\}$ , und  $E = \{e_1, \dots, e_n\}$  die den Modellteilen zugeordneten Beobachtungen. Die wahrscheinlichste Zuordnung von Modell- zu Bildmerkmalen wird durch die Maximierung der a posteriori Wahrscheinlichkeit  $P(m_0, m_1, \dots, m_n | e_1, \dots, e_n)$  erreicht, die sich wie folgt darstellt:

$$P = P(m_0, m_1, \dots, m_n | e_1, \dots, e_n) = P(m_0, m_1, \dots, m_n) \prod_{i=1}^n \frac{P(e_i | m_i)}{P(e_i)} \quad (6.9)$$

Dabei ist  $m_0$  ein Modell für den Bildhintergrund, d.h. eine Erklärung für alle Teile, die nicht zu dem eigentlich gesuchten Objekt gehören. Die a priori Wahrscheinlichkeit  $P(m_0, m_1, \dots, m_n)$  wird weiter aufgespalten in

$$P(m_0, m_1, \dots, m_n) = P(m_0 | m_1, \dots, m_n) P(m_1, \dots, m_n) \quad (6.10)$$

$$= P_0 P(m_1, \dots, m_n) \quad (6.11)$$

wobei  $P(m_1, \dots, m_n)$  die a priori Wahrscheinlichkeit für das Auftreten der  $n$  Modellteile bezeichnet, und  $P_0$  die davon unabhängige Wahrscheinlichkeit für den Hintergrund. Nun kann die Gleichung 6.9 auch als Information ausgedrückt werden, und es ergibt sich die Bewertungsfunktion  $S$  zu

$$S = \log_2 \frac{P}{P_0} = F - G \quad (6.12)$$

Da der Bildhintergrund nicht explizit modelliert wird, wird  $P_0$  als konstant angenommen und kann bei der Maximierung von  $S$  vernachlässigt werden. Gemäß Gleichung 6.9 sind  $F$  und  $G$  folgendermaßen definiert:

$$F = \sum_{i=1}^n F_i = \sum_{i=1}^n [-\log_2 P(e_i) + \log_2 P(e_i | m_i)] \quad (6.13)$$

$$G = -\log_2 P(m_1, \dots, m_n) \quad (6.14)$$

Eine genauere Betrachtung der Summanden  $F_i$  zeigt, daß diese der Definition der gemeinsamen Information  $I(e_i; m_i)$  zwischen Modellteil und zugeordnetem Bildmerkmal entspricht (vgl. Gleichung 5.5 auf Seite 88). Unter der Annahme einer optimalen Codierung entspricht  $G$  außerdem der Codierungslänge des Modells  $M_i$ . Zur Bestimmung des wahrscheinlichsten Modells reicht es gemäß Gleichung 6.12 also aus, für jedes Modell die beste Zuordnung zu bestimmen und von der gemeinsamen Information dieser Zuordnung die Codierungslänge des Modells zu subtrahieren. Die resultierenden Ergebnisse sind dann miteinander vergleichbar, und das Modell mit der höchsten Gesamtbewertung  $S$  stellt das wahrscheinlichste unter den gegebenen Daten dar.

Die von Fua und Hanson vorgestellte probabilistische Herleitung dieses MAP-Schätzers läßt sich aber auch einfacher über die Informationstheorie erreichen. Laut Definition der gemeinsamen Information ist

$$I(D; M) = I(D) - I(D|M). \quad (6.15)$$

Wenn von der gemeinsamen Information zwischen Modell und Bild (Gleichung 6.15) der Informationsgehalt der Daten  $I(D)$  sowie der Informationsgehalt des Modells  $I(M)$  subtrahiert wird, ergibt sich

$$I(D; M) - I(D) - I(M) = -I(D|M) - I(M). \quad (6.16)$$

Die Maximierung von (6.16) entspricht der Minimierung von  $I(D|M) + I(M)$ , was wiederum genau Gleichung 6.5 entspricht. Unter der Annahme einer optimalen Codierung bestimmt die Maximierung von Gleichung (6.16) über alle möglichen Modelle  $M_i$  also das wahrscheinlichste Modell. Es bleibt noch die Frage nach  $I(D)$  und  $I(M_i)$  zu klären, die dazu von der gemeinsamen Information subtrahiert werden müssen.

Der Informationsgehalt der Daten  $I(D)$  hängt ausschließlich vom Bild ab. Er ist deshalb bzgl. der verschiedenen Modelle konstant und kann bei der Maximierung vernachlässigt werden. Die Definition von  $I(M_i)$  entspricht der Definition von  $G$  in Gleichung 6.14 und entspricht bei optimaler Codierung der Codierungslänge des Modells. Somit läßt sich die Bestimmung des wahrscheinlichsten Modells  $\hat{M}$  aus einer Menge von Modellen  $\{M_1, \dots, M_k\}$  mit den Codierungslängen  $L(M_i)$  und den jeweils dazu bestimmten wahrscheinlichsten Zuordnungen  $\{\hat{h}_1, \dots, \hat{h}_k\}$  mit gemeinsamer Information  $I_{\hat{h}_i}(D; M_i)$  mittels des folgenden MAP-Schätzers realisieren:

$$\hat{M} = \arg \max_{M_i} (I_{\hat{h}_i}(D; M_i) - L(M_i)) \quad (6.17)$$

Da die Berechnung der gemeinsamen Information der besten Zuordnung eines Modells zu den Bild-daten auf Basis der relationalen Repräsentation der Gebäudeansichten erfolgt, muß sich auch die Codierung der einzelnen Modelle  $M_i$  auf diese Repräsentation beziehen. Es geht im folgenden deshalb darum, ein Verfahren zur Bestimmung der Codierungslänge relationaler Gebäudestrukturen zu entwickeln.

## 6.2 Codierung relationaler Gebäudestrukturen

Für eine optimale Codierung der Modelle müßten die a priori Wahrscheinlichkeiten für das Auftreten jedes einzelnen Modells bekannt sein (s.o., Abschnitt 6.1.2). Da diese Wahrscheinlichkeiten in den meisten Fällen gar nicht oder zumindest nicht vollständig bekannt sind, gibt es zwei Möglichkeiten Näherungen für die Codierungslängen zu erhalten:

1. Die a priori Wahrscheinlichkeiten werden direkt empirisch geschätzt. Unter der Annahme, daß die geschätzten Werte den wahren Wahrscheinlichkeiten entsprechen, kann man die Codierungslänge des optimalen Codes unmittelbar aus diesen Wahrscheinlichkeiten berechnen.
2. Es wird ein Codiervorgang angegeben, das für alle auftretenden Modelle eine möglichst kurze Beschreibungslänge generiert, die nahe an der wahren Beschreibungslänge und damit auch der wahren a priori Wahrscheinlichkeit liegt.

Beim ersten Ansatz wird also die Codierungslänge geschätzt, ohne daß dazu überhaupt ein Codiervorgang angegeben werden muß. Jedoch muß eine gute Schätzung der a priori Wahrscheinlichkeiten vorliegen, was eher nur für einfachere Modelle möglich ist.

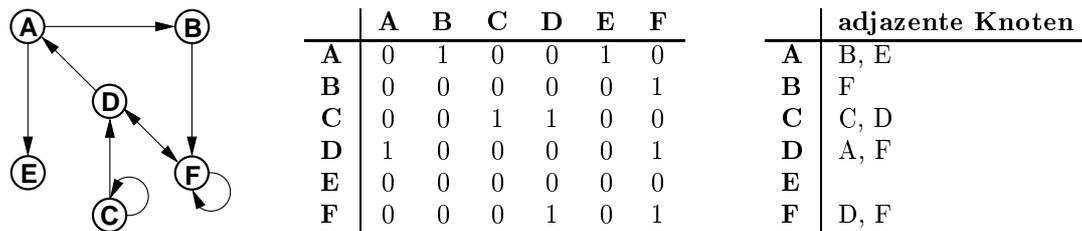
Das Hauptproblem bei dem zweiten Ansatz besteht darin, daß es in der Regel schwierig ist, ein Codiervorgang anzugeben, das auf der einen Seite möglichst optimal und auf der anderen Seite allgemeingültig bzgl. aller möglichen Modelle ist. Eine nicht-optimale Codierung führt aber schlimmstenfalls dazu, daß einfachere Modelle bevorzugt werden, weil ihre Längen überschätzt werden. Die deutlich problematischere Situation des Over-Fittings (vgl. [DH73]), bei dem beliebig komplexe Modelle generiert werden, wird somit generell vermieden.

Aufgrund der Komplexität der relationalen Repräsentation von Gebäudeansichten, kommt der erste Ansatz zur Bestimmung der Codierungslänge nicht in Frage. Die a priori Wahrscheinlichkeiten können nicht für jede mögliche relationale Gebäudestruktur empirisch geschätzt werden, weil die Gebäudeansichten gleichzeitig durch verschiedenartige Relationen wie Linienparallelität, Merkmalsnachbarschaft, Kollinearität etc. repräsentiert werden, deren Auftretenswahrscheinlichkeiten voneinander abhängen. Aus diesem Grund muß ein allgemeines, für alle möglichen Gebäudeansichten gültiges Codierschema konstruiert werden, das die relationalen Strukturen so kompakt wie möglich repräsentiert, und jeweils möglichst nahe an die wahre minimale Codierungslänge und somit auch nahe an die a priori Wahrscheinlichkeit des Modells kommt.

Im nächsten Abschnitt werden dazu zunächst drei allgemeine Schemata für die Repräsentation von Graphen vorgestellt und die von ihnen benötigten Codierungslängen diskutiert. Anschließend wird gezeigt, wie sich Gebäudeansichten unter Nutzung dieser Codierschemata und insbesondere unter expliziter Berücksichtigung des Wissens über das verwendete Gebäudemodell sehr kompakt codieren lassen.

### 6.2.1 Allgemeine Codierung von Graphen

Zur kompakten Darstellung relationaler Strukturen wird auf das mathematische Konzept des Graphen zurückgegriffen (s. [CH94a, BSMM97]). Ein Graph  $G = (V, E)$  besteht aus einer Menge von Knoten  $V$  und dazwischen bestehenden Kanten  $E = \{(v_i, v_j) | v_i, v_j \in V\}$ . Für die Repräsentation von Graphen werden in der Informatik hauptsächlich zwei Standarddatenstrukturen eingesetzt; dies ist zum einen die Adjazenzmatrix und zum anderen die Adjazenzliste (s. [AHU87]). Beide Datenstrukturen codieren Graphen und damit die zugrundeliegenden Relationen auf unterschiedliche Weise. Abbildung 6.2 zeigt links einen Beispielgraphen und rechts davon seine Repräsentationen als Adjazenzmatrix bzw. Adjazenzliste.



**Abbildung 6.2:** Beispielgraph mit  $|V| = 6$  Knoten und  $|E| = 9$  Kanten (links); Repräsentation des Graphen durch eine Adjazenzmatrix (Mitte) und eine Adjazenzliste (rechts).

Die Adjazenzmatrix besitzt für jeden Knoten eine Zeile und Spalte. Die Matrixeinträge können jeweils nur die Werte 0 oder 1 annehmen<sup>2</sup>. Das Vorhandensein einer Graphkante bzw. einer Relation von Knoten A zu Knoten B wird dann durch das Setzen der Matrixzelle an der Position (Zeile A, Spalte B) auf den Wert 1 ausgedrückt. Die Adjazenzliste hingegen notiert für jeden Knoten die Menge der Knoten, zu denen eine Graphkante bzw. Relation besteht.

#### 6.2.1.1 Codierungslängen der Adjazenzmatrix und Adjazenzliste

Die Adjazenzmatrix besitzt für jede mögliche Graphkante ein Bit, das besagt, ob die entsprechende Kante besteht. Das bedeutet, daß zur Codierung eines Graphen mit  $|V|$  Knoten  $|V|^2$  Bits benötigt werden. Da der Code die Präfixeigenschaft erfüllen, d.h. unmittelbar dekodierbar sein muß, wird vorher noch die Anzahl der Knoten codiert. Damit Sender und Empfänger wissen, wieviele Bits zur Codierung der Anzahl verwendet werden, wird vorher eine maximale Knotenanzahl  $|V|_{\max}$  vereinbart. Die Gesamtcodierungslänge  $L(\text{Adj. Matrix})$  beträgt somit

$$L(\text{Adj. Matrix}) = \log_2(|V|_{\max} + 1) + |V|^2 \quad [\text{Bit}]. \quad (6.18)$$

Bei der Adjazenzliste müssen für jeden Knoten zunächst die Anzahl seiner adjazenten Knoten und dann die Knoten selber aufgelistet werden. Wenn man die Knoten bei 0 beginnend durchnumeriert, werden zur Codierung der Anzahlen adjazenter Knoten insgesamt  $|V| \log_2(|V| + 1)$  und für die einzelnen Kanten insgesamt  $|E| \log_2 |V|$  Bits benötigt. Zusammen mit der vorherigen Codierung der Knotenanzahl ergibt sich die Codierungslänge  $L(\text{Adj. Liste})$  zu

$$L(\text{Adj. Liste}) = \log_2(|V|_{\max} + 1) + |V| \log_2(|V| + 1) + |E| \log_2 |V| \quad [\text{Bit}]. \quad (6.19)$$

Unter der Annahme von  $|V|_{\max} = 7$  ergibt sich für den in Abbildung 6.2 gezeigten Graphen für die Adjazenzliste eine Codierungslänge von 39 Bit und für die Adjazenzmatrix eine Länge von 43,11 Bit.

<sup>2</sup>Dert Wert 1 steht für *wahr* und 0 für *falsch*.

### 6.2.1.2 Graphcodierung mittels Permutationsreihe

Ein anderer Ansatz zur kompakten Codierung von Graphen wurde von Cook und Holder im Rahmen einer Forschungsarbeit im Bereich des maschinellen Lernens vorgestellt (siehe [CH94b]). Dabei wurde ein Verfahren zur Erkennung wiederkehrender Substrukturen, d.h. Subgraphen innerhalb großer Graphen entwickelt. Zur Erkennung, ob bestimmte Teilgraphen signifikante Substrukturen innerhalb des großen Graphen bilden, werden die in Frage kommenden Substrukturen kompakt codiert und alle Vorkommnisse dieser Strukturen in dem großen Graph durch die kompaktifizierte Beschreibung ersetzt. Wird dabei die Codierungslänge der Gesamtstruktur kleiner, so kann für die betrachteten Teilgraphen gemäß des Prinzips der minimalen Beschreibungslänge angenommen werden, daß es sich dabei um signifikante Substrukturen handelt.

Das in dem Ansatz verwendete Codierschema für Graphen baut auf einem von Quinlan und Rivest vorgeschlagenem Codierschema für Strings auf (vgl. [QR89]). Diese verwendeten das MDL-Prinzip zur Herleitung von Entscheidungsbäumen, und mußten dazu Strings möglichst optimal codieren. Für die Codierung der Strings wird eine sogenannte *Permutationsreihe* verwendet, die einen Bitstring als ein durch eine Permutation einer festen Anzahl von Einsen und Nullen entstandenes Tupel betrachtet. Da es nur  $\binom{n}{k}$  Möglichkeiten gibt,  $k$  Einsen auf einen  $n$  Bit langen String zu verteilen, und diese Permutationen in bestimmter Reihenfolge algorithmisch aufzählbar sind, reicht es zur Codierung des Strings, die Nummer der Permutation anzugeben. Vor der Codierung der Permutationsnummer müssen allerdings noch die Länge des Bitstrings und die Anzahl der Einsen codiert werden. Die Gesamtlänge beträgt demnach

$$L(\text{Bitstring}) = \log_2 n_{\max} + \log_2(n+1) + \log_2 \binom{n}{k} \quad [\text{Bit}]. \quad (6.20)$$

**Beispiel:** Der String (0,0,1,0,0,0,0,1,1,0,0,0,0,0,1,0,0,0) besitzt die Länge  $n = 18$  und beinhaltet  $k = 4$  Einsen. Im Gegensatz zur Repräsentation als Bitvektor, die  $\log_2 n_{\max} + 18$  Bits lang ist, werden für die Codierung mittels Permutationsreihe nur  $\log_2 n_{\max} + 15,83$  Bits benötigt.

Cook und Holder nutzen dieses Stringcodierschema nun zur Codierung der Adjazenzmatrizen von Graphen. Dabei wird jede Zeile eines Graphen als Bitstring aufgefaßt, und es werden für jede Zeile die Anzahl der Einsen sowie die Nummer der Permutation codiert. Die maximale Anzahl von Einsen pro Zeile ist durch den maximalen Knotengrad des Graphen beschränkt. Deshalb wird nach der Codierung der Knotenanzahl des Graphen zunächst der maximale Knotengrad *MaxKnotengrad* codiert. Die Angabe der Anzahl der Einsen benötigt dann pro Zeile nur noch  $\log_2(\text{MaxKnotengrad} + 1)$  Bits. Die Gesamtlänge eines mittels Permutationsreihe codierten Graphen beträgt somit

$$L(\text{Perm.Reihe}) = \underbrace{\log_2(|V|_{\max} + 1)}_{\text{Anzahl Graphknoten}} + \underbrace{\log_2(|V| + 1)}_{\text{max. Knotengrad}} + \sum_{i=1}^{|V|} \left( \underbrace{\log_2(\text{MaxKnotengrad} + 1)}_{\text{Grad des Knotens } v_i} + \underbrace{\log_2 \binom{|V|}{\text{degree}(v_i)}}_{\text{Nr. der Permutation}} \right) \quad (6.21)$$

Dies können wir aber noch verbessern: Die komplizierte Codierung kann vereinfacht werden, indem man alle Zeilen der Adjazenzmatrix hintereinanderhängt und den Graph dann als Bitstring der Länge  $|V|^2$  interpretiert. Der Vorteil gegenüber der von Cook und Holder vorgeschlagenen Codierung besteht zum einen darin, daß man die Bits zur Codierung der Anzahl der Einsen pro Zeile einsparen kann. Zum anderen hängt die Codierungslänge dann nur noch von der Knoten- und Kantenanzahl ab, und nicht mehr zusätzlich von der Verteilung der Kanten im Graph. Die

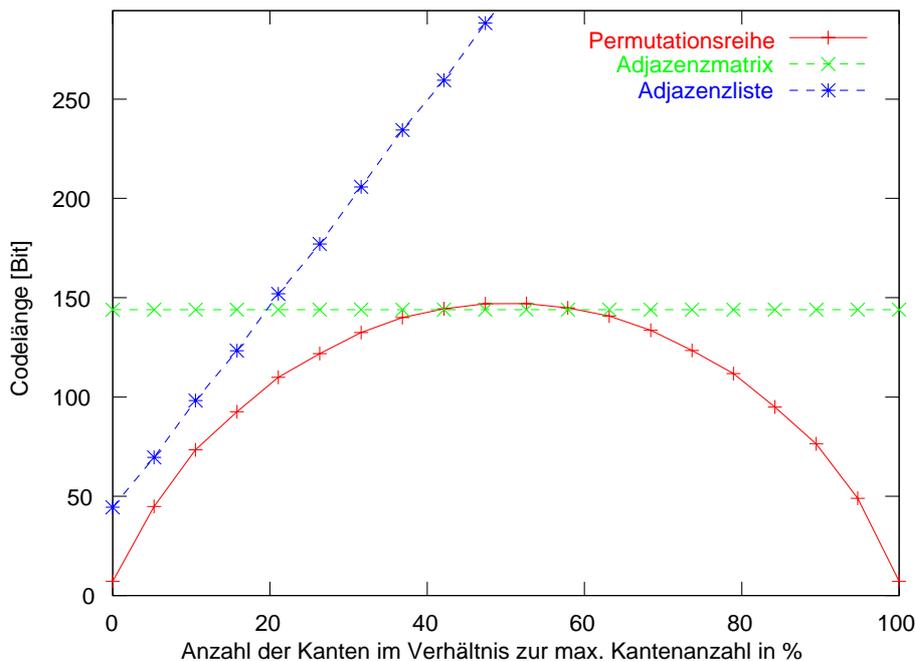
Codierungslänge  $L(\text{Neue Perm.Reihe})$  eines Graphen beträgt dann

$$L(\text{Neue Perm.Reihe}) = \log_2(|V|_{\max} + 1) + \log_2(|V| + 1) + \log_2\left(\frac{|V|^2}{|E|}\right) \quad [\text{Bit}]. \quad (6.22)$$

Unter der Annahme von  $|V|_{\max} = 7$  ergibt sich für den in Abbildung 6.2 gezeigten Graphen für die von Cook und Holder vorgeschlagene Codierung eine Länge von 33,53 Bit und für die neue Permutationsreihencodierung eine Länge von 32,3 Bit.

### 6.2.1.3 Diskussion der verschiedenen Graphrepräsentationen

Die Codierungslängen der drei vorgestellten Repräsentationsformen hängen zwar von der Knotenanzahl und zum Teil auch von der Kantenanzahl ab, jedoch sind sie unabhängig von der Verteilung der Kanten im Graph, d.h. es spielt keine Rolle, zwischen welchen Knoten die Kanten bestehen. Zum Vergleich der drei Repräsentationen wird deshalb untersucht, wie sich die Codierungslängen bzgl. eines Graphen mit fester Knoten- und variierender Kantenanzahl verhalten. Abbildung 6.3 zeigt die Entwicklung der Codierungslängen, angefangen bei einem Graphen mit 0 Kanten bis hin zu einem Graphen mit der maximalen Anzahl von  $|V|^2$  Kanten.



**Abbildung 6.3:** Codierungslänge eines Graphen mit 12 Knoten in bezug auf die Anzahl seiner Kanten, aufgetragen für die drei Codierschemata Adjazenzmatrix, Adjazenzliste und Permutationsreihe. Es ist erkennbar, daß die Codierungslänge der Permutationsreihe bzgl. der Adjazenzliste immer, und bzgl. der Adjazenzmatrix fast immer kleiner ist. Die Codierung als Adjazenzmatrix ist nur kürzer, wenn die Anzahl der Graphkanten zwischen ca. 43 % und ca. 57 % der maximalen Kantenanzahl (in diesem Beispiel ist  $|E|_{\max} = |V|^2 = 144$ ) liegt.

Die Codierungslänge der Adjazenzmatrix entspricht der Anzahl der möglichen Graphkanten und ist somit über den gesamten Bereich konstant. Die durch die tatsächlich vorhandenen Graphkanten ausgedrückte relationale Komplexität bleibt bei der Codierung unberücksichtigt. Adjazenzmatrizen sind deswegen für die Bestimmung des wahrscheinlichsten Modells ungeeignet, weil sie verschiedene Graphen, die dieselbe Knotenanzahl besitzen, nicht unterscheiden können.

Die Adjazenzliste berücksichtigt zwar die Anzahl der Kanten, jedoch steigt die Codierungslänge mit wachsender Kantenanzahl stark an. Bereits ab ca. 18 % der möglichen Kantenanzahl werden mehr Bits als für die Adjazenzmatrix benötigt.

Wie man Abbildung 6.3 entnehmen kann, ist die Permutationsreihencodierung immer kürzer als die beiden anderen, bis auf einen Bereich, in dem der Graph ungefähr die Hälfte aller möglichen Kanten besitzt. In diesem Fall ist die Adjazenzmatrixcodierung kürzer. Es läßt sich jedoch folgendes berechnen: Je größer die Anzahl der Knoten wird, desto kleiner wird das Intervall, in dem die Codierung der Permutationsreihe länger ist. Interessanterweise nimmt die Codierungslänge ab der Hälfte der möglichen Kanten wieder ab, was sich durch die Anzahl der Permutationen erklären läßt: es gibt genausoviele Möglichkeiten beispielsweise vier Einsen auf einen Bitstring zu verteilen wie vier Nullen. Der Grund dafür, daß die Adjazenzliste bereits von 0 Kanten an mehr Bits als die Permutationsreihe benötigt, liegt darin, daß sie für jeden Knoten die Anzahl seiner Kanten codieren muß (mittlerer Summand in Gleichung 6.19).

#### 6.2.1.4 Einfluß von Relationeneigenschaften auf die Codierungslänge

Wenn ein Graph ausschließlich durch Relationen einer Art gebildet wird, wie z.B. die Linienparallelität, und die Definition dieser Relation bestimmte mathematische Eigenschaften aufweist, wie z.B. Reflexivität, Symmetrie oder Transitivität, so spiegelt sich dies auch in der jeweiligen Ausprägung des Graphen in der Datenstruktur wider. Symmetrische Relationen führen beispielsweise dazu, daß die Einträge der Adjazenzmatrix an der Hauptdiagonalen gespiegelt sind. Reflexivität bedeutet, daß alle Einträge der Hauptdiagonalen 1 sind (bei Irreflexivität sind sie 0). Wenn eine Relation universal ist, d.h. jeder Knoten mit jedem anderen in Relation steht, besteht die Adjazenzmatrix nur aus Einsen, und bei der Adjazenzliste steht zu jedem Knoten die Liste aller Knoten aufgezählt. Transitive Relationen bilden zusammenhängende Komponenten innerhalb des Graphen; transitive und symmetrische Relationen bilden Cliques, innerhalb derer jeder Knoten mit jedem anderen verknüpft ist.

Diese Eigenschaften können nun für eine kürzere Codierung ausgenutzt werden. Wenn beispielsweise sowohl der Kodierer (Sender) als auch der Dekodierer (Empfänger) eines codierten Graphen wissen, daß dieser auf eine symmetrische Relation zurückgeht, reicht es, neben der Anzahl der Knoten nur die Matrixeinträge oberhalb und einschließlich der Hauptdiagonalen zu codieren. Die untere Hälfte kann der Empfänger durch Spiegelung an der Hauptdiagonalen wieder rekonstruieren. Für universale Relationen reicht es gar, nur die Anzahl der Knoten zu codieren, weil der Empfänger weiß, daß zur Rekonstruktion des Graphen von jedem Knoten zu jedem anderen eine Kante besteht.

Wie oben bereits angesprochen bilden alle Objekte, die durch eine symmetrische und transitive Relation zueinander in Beziehung stehen, Cliques im Graph. Jede Clique mit  $n$  Knoten hat dabei  $n(n - 1)/2$  Kanten. Diese Cliques können vom Sender jeweils zu einem minimalen spannenden Baum ausgedünnt werden, der nur noch  $n - 1$  Kanten besitzt (vgl. [AHU87, CH94a]). Da der Empfänger weiß, daß es sich um eine transitive Relation handelt, kann er die fehlenden Kanten in den empfangenen Graph durch Berechnung der transitiven Hülle wieder einfügen. Durch diese Maßnahme reduziert sich die Anzahl der Kanten des zu codierenden Graphen je nach Größe der einzelnen Cliques mitunter sehr beträchtlich.

## 6.2.2 Codierung der Relationen des Gebäudemodells

Die relationale Struktur von Gebäudeansichten setzt sich aus mehreren unterschiedlichen Relationen wie Linienparallelität, Merkmalsnachbarschaft usw. zusammen. Bei der vollständigen Codierung einer Gebäudeansicht zu einer Nachricht müssen deshalb alle verwendeten Relationen berücksichtigt werden. Dazu reicht es aber nicht aus, einfach die Graphen für jede Relation getrennt zu codieren und hintereinander in die Nachricht zu schreiben. Die verwendeten Relationen weisen nämlich z.T. größere funktionale Abhängigkeiten voneinander auf, d.h. man kann aus

dem Vorhandensein bestimmter Relationen des Modells u.U. auf das Vorhandensein bzw. Nicht-Vorhandensein von Relationen eines anderen Typs schließen.

Diese Redundanzen müssen bei der Konstruktion einer möglichst kurzen Codierung berücksichtigt werden, denn optimal codierte Nachrichten beinhalten keine Redundanz. Der Informationsgehalt redundanter Informationen in einer Nachricht ist gleich 0, weil daraus keine neuen Erkenntnisse gewonnen werden können, die sich nicht aus den anderen Teilen der Nachricht herleiten ließen. Da im Fall der optimalen Codierung der Informationsgehalt einer Nachricht seiner Codierungslänge entspricht, folgt daraus, daß optimal codierte Nachrichten keine Redundanzen aufweisen.

Es geht also zunächst darum, die verwendeten Relationen auf Abhängigkeiten zueinander zu untersuchen. Anschließend gilt es ein Gesamtcodierschema für Gebäudemodelle zu spezifizieren, das die identifizierten Abhängigkeiten gezielt zur Verkürzung der Codierung verwendet. An dieser Stelle spielt die Semantik des Gebäudemodells eine entscheidende Rolle, weil diese beschreibt, wie Gebäude aufgebaut sind, und damit auch implizit definiert, wann welche Relationen auftreten und welche Abhängigkeiten zwischen ihnen existieren.

In den folgenden fünf Unterabschnitten werden die Eigenschaften, Abhängigkeiten und Codierungen der Merkmalsnachbarschafts-, Linienparallelitäts-, Kollinearitäts-, Same\_Side\_Lines- und Verschiedenheitsrelationen erläutert. Dabei wird zusätzlich das Wissen um die Struktureigenschaften Symmetrie, Reflexivität und Transitivität der einzelnen Relationen, wie im vorangegangenen Abschnitt 6.2.1 erläutert, genutzt, um kompaktere Repräsentationen zu erhalten. Die Eigenschaften der jeweiligen Relationen waren in Tabelle 5.3 auf Seite 92 zusammengefaßt.

Bei der im folgenden verwendeten Notation bezeichnet  $P$  die Menge der Punkte,  $L$  die Menge der Linien und  $F$  die Menge der Flächen einer Gebäudeansicht. Dementsprechend bezeichnen  $|P|$ ,  $|L|$  und  $|F|$  die jeweiligen Anzahlen. Für die verschiedenen Codierungslängen werden  $\Omega$  und  $\Psi$  mit entsprechenden Indizes verwendet.

### 6.2.2.1 Codierung der Merkmalsnachbarschaft

Das verwendete Bildmodell berücksichtigt nur Nachbarschaften zwischen Merkmalen jeweils verschiedener Klassen. Das sind die Nachbarschaften  $MNG(L,P)$  zwischen Linien und Punkten,  $MNG(F,L)$  zwischen Flächen und Linien und  $MNG(F,P)$  zwischen Flächen und Punkten. Die

	Punkte	Linien	Flächen
	$p_1, \dots, p_{ P }$	$l_1, \dots, l_{ L }$	$f_1, \dots, f_{ F }$
Punkte	$p_1$ $\vdots$ $p_{ P }$	0	$(L, P)$
Linien	$l_1$ $\vdots$ $l_{ L }$	$(L, P)$	0
Flächen	$f_1$ $\vdots$ $f_{ F }$	$(F, P)$	$(F, L)$

**Abbildung 6.4:** Partitionierung der Adjazenzmatrix des Merkmalsnachbarschaftsgraphen (MNG) in die drei Teilgraphen (Linie, Punkt), (Fläche, Linie) und (Fläche, Punkt).

Adjazenzmatrix des Merkmalsnachbarschaftsgraphen (MNG) besitzt aus diesem Grund die in Abbildung 6.4 gezeigte Form. Wegen der Symmetrie der Nachbarschaftsrelation sind die Teilgraphen oberhalb der Diagonalen (bis auf eine Spiegelung an den jeweiligen Hauptdiagonalen der Teilgraphen) identisch. Das bedeutet, daß im folgenden nur die drei Teilgraphen  $MNG(L,P)$ ,  $MNG(F,L)$

und  $\text{MNG}(\mathbf{F},\mathbf{P})$  codiert werden müssen, um den gesamten Merkmalsnachbarschaftsgraph zu beschreiben.

### Merkmalsnachbarschaft zwischen Linie und Punkt $\text{MNG}(\mathbf{L},\mathbf{P})$ :

Das Bildmodell besagt, daß jede Linie von zwei Endpunkten begrenzt wird. Damit ist jede Linie zu genau zwei Punkten benachbart. Zur Codierung bietet sich in diesem Fall die Adjazenzliste an, weil die Anzahl adjazenter Knoten zu jedem Knoten feststeht und sie deshalb nicht für jeden Knoten einzeln angegeben werden muß. Für die Codierung werden einfach nacheinander für jede Linie die beiden Endpunkte benannt. Die Codierungslänge  $\Omega_{\text{MNG}(\mathbf{L},\mathbf{P})}$  des  $\text{MNG}(\mathbf{L},\mathbf{P})$ -Teilgraphen beträgt somit

$$\Omega_{\text{MNG}(\mathbf{L},\mathbf{P})} = 2 \cdot |L| \cdot \log_2 |P| \quad [\text{Bit}]. \quad (6.23)$$

### Merkmalsnachbarschaft zwischen Fläche und Linie $\text{MNG}(\mathbf{F},\mathbf{L})$ :

Jede Linie ist normalerweise zu genau zwei Flächen benachbart, es sei denn, es handelt sich um eine Randlinie. Zur Codierung dieser Relation wird deshalb zunächst ein Bitstring der Länge  $|L|$  angegeben, der für jede Linie besagt, ob sie eine Randlinie ist (und damit nur eine Nachbarfläche besitzt) oder nicht. Anschließend werden nacheinander für jede Linie gemäß des vorher codierten Bitstrings entweder eine oder zwei Nachbarflächen angegeben.

Sei  $L_R \subseteq L$  die Menge der Randlinien und  $L_I = L \setminus L_R$  die Menge der innenliegenden Linien der Gebäudeansicht. Dann beträgt die Codierungslänge  $\Omega_{\text{MNG}(\mathbf{F},\mathbf{L})}$  des  $\text{MNG}(\mathbf{F},\mathbf{L})$ -Teilgraphen

$$\Omega_{\text{MNG}(\mathbf{F},\mathbf{L})} = |L| + (|L_R| + 2|L_I|) \log_2 |F| \quad [\text{Bit}]. \quad (6.24)$$

Die  $(\mathbf{F},\mathbf{L})$ -Relation könnte auch mit Hilfe einer Permutationsreihe codiert werden, jedoch haben praktische Tests gezeigt, daß diese in allen Fällen um wenige Bits länger ist. Der Grund dafür liegt darin, daß sich die oben vorgestellte Codierung das Wissen um die Beschränkung der Anzahl der Nachbarschaften auf entweder eine oder zwei Flächen pro Linie explizit zunutze macht.

### Merkmalsnachbarschaft zwischen Fläche und Punkt $\text{MNG}(\mathbf{F},\mathbf{P})$ :

Da die  $\text{MNG}(\mathbf{F},\mathbf{L})$ -Relationen beschreiben, von welchen Linien eine Fläche jeweils umschlossen wird, und die  $\text{MNG}(\mathbf{L},\mathbf{P})$ -Relationen die Endpunkte dieser Linien benennen, lassen sich die Nachbarschaftsrelationen zwischen Flächen und Punkten wie folgt herleiten:

$$\begin{aligned} \text{MNG}(\mathbf{F},\mathbf{P}) = \{ (f,p) \mid \exists f \exists p \exists l : f \in F \wedge p \in P \wedge l \in L \\ \wedge (f,l) \in \text{MNG}(\mathbf{F},\mathbf{L}) \wedge (l,p) \in \text{MNG}(\mathbf{L},\mathbf{P}) \} \end{aligned} \quad (6.25)$$

Der  $(\mathbf{F},\mathbf{P})$ -Teilgraph muß also nicht gesondert codiert werden; die Codierungslänge  $\Omega_{\text{MNG}(\mathbf{F},\mathbf{P})}$  beträgt demzufolge

$$\Omega_{\text{MNG}(\mathbf{F},\mathbf{P})} = 0 \quad [\text{Bit}]. \quad (6.26)$$

### Gesamtcodierungslänge des $\text{MNG}$ :

Die Gesamtcodierungslänge des Merkmalsnachbarschaftsgraphen ergibt sich gemäß der Gleichungen 6.23, 6.24 und 6.26 zu

$$\Omega_{\text{MNG}} = \Omega_{\text{MNG}(\mathbf{L},\mathbf{P})} + \Omega_{\text{MNG}(\mathbf{F},\mathbf{L})} \quad [\text{Bit}]. \quad (6.27)$$

### 6.2.2.2 Codierung der Linienparallelität

Wie in Abschnitt 3.3.1.2 erläutert, handelt es sich bei der Linienparallelitätsrelation des Gebäude-modells um eine Äquivalenzrelation. Aufgrund der Symmetrie und Reflexivität ist die Kardinalität der Relation auf  $n = |L|(|L| - 1)/2$  beschränkt. Wegen der Transitivität zerfällt der Graph in die Cliques  $C_1, \dots, C_c$ , wobei jede Clique mit  $|C_i|$  Knoten durch einen minimalen Spannbaum mit  $|C_i| - 1$  Kanten ersetzt wird (vgl. Abschnitt 6.2.1). Sei  $k = \sum_{i=1}^c (|C_i| - 1)$  die Anzahl der verbleibenden Kanten des Graphen. Dann beträgt die Codierungslänge der Linienparallelität  $\Omega_{\text{parallel}}$  bei Verwendung der Permutationsreihencodierung

$$\Omega_{\text{parallel}} = \log_2(n + 1) + \log_2 \binom{n}{k} \quad [\text{Bit}]. \quad (6.28)$$

### 6.2.2.3 Codierung der Kollinearität zwischen Punkt und Linie

Die Kollinearitätsrelation wird genau wie die Merkmalsnachbarschaft zwischen Punkt und Linie aus der Inzidenzbeziehung der Linien zu den Punkten der Gebäudeansichten hergeleitet. Beide Graphen sind also zueinander isomorph, womit der eine Graph aus dem anderen konstruiert werden kann. Das bedeutet, daß die Kollinearitätsrelation nicht extra codiert werden muß. Die Codierungslänge  $\Omega_{\text{collin}}$  beträgt demzufolge

$$\Omega_{\text{collin}} = 0 \quad [\text{Bit}]. \quad (6.29)$$

### 6.2.2.4 Codierung der Same\_Side\_Lines-Relation

Jede Same\_Side\_Lines-Relation (kurz SSL-Relation) wird aus drei zusammenhängenden, hintereinanderfolgenden Linien des Umringspolygons einer Fläche gebildet. Die mittlere Linie wird dabei jeweils als  $L_{Ref}$  bezeichnet. Die SSL-Relation beschreibt nun für jede Linie des Polygons, ob die an den beiden Enden von  $L_{Ref}$  angrenzenden Linien in derselben Halbebene liegen, die von der durch  $L_{Ref}$  verlaufenden Geraden aufgespannt wird.

Nun können auf den beiden Seiten der Linie  $L_{Ref}$  genau eine oder zwei Flächen liegen, wobei der erste Fall dann eintritt, wenn es sich um eine Randlinie, d.h. eine Außenkante der Gebäudeansicht handelt. Diese Information muß nicht extra codiert werden, sondern läßt sich für jede Linie aus dem Merkmalsnachbarschaftsgraphen  $MNG(F,L)$  ablesen. Ferner wird auch die Adjazenz der drei Linien aus dem MNG hergeleitet.

Jede Linie kann höchstens zweimal als  $L_{Ref}$  auftreten, einmal für die Fläche auf der einen und evtl. noch einmal für die Fläche auf der anderen Seite. Da aus dem MNG bestimmt werden kann, zu wievielen Flächen die Linie inzident ist, muß dies nicht extra codiert werden und es reichen für jede Linie ein bzw. zwei Bits zur Codierung, ob die SSL-Relation für die jeweilige Seite wahr ist. Bei zwei Flächen steht dabei das erste Bit für die Fläche mit der kleineren Nummer. Da eine Fläche von mindestens drei Linien eingeschlossen wird, muß auch jede Gebäudeansicht mindestens drei Außenkanten besitzen. Die Länge  $n$  des Bitvektors liegt demnach im Intervall  $|L| \leq n \leq 2|L| - 3$ .

Es läßt sich beobachten, daß für einfache Gebäude alle und für komplexere Gebäude zumindest der Großteil der Hausflächen konvex sind. Aus diesem Grund sind die SSL-Relationen zu den meisten Linien der Umringspolygone wahr. Das bedeutet wiederum, daß der codierte Bitvektor zum größten Teil aus Einsen besteht. Zur Codierung bietet sich deshalb die Codierung in Form einer Permutationsreihe an, weil diese ja im Fall von verhältnismäßig wenigen sowie von vielen Einsen effizienter ist, als die Codierung als Adjazenzliste oder Adjazenzmatrix (vgl. Abbildung 6.3 auf Seite 108). Wenn  $k$  die Anzahl der Einsen ist, dann beträgt die Codierungslänge  $\Omega_{\text{SSL}}$  der Same\_Side\_Line-Relation

$$\Omega_{\text{SSL}} = \log_2(n + 1) + \log_2 \binom{n}{k} \quad [\text{Bit}]. \quad (6.30)$$

An dieser Stelle wird der Einfluß der Semantik des Gebäudemodells auf die Codierung sehr deutlich: Die Ergänzung des MNG um nur wenige Bits reicht aus, um diese geometrische Relation qualitativ zu beschreiben. Hier wird also Modellwissen dazu genutzt, um die Codierung besonders effizient zu machen und damit auch näher an die wahre Beschreibungslänge zu kommen.

### 6.2.2.5 Codierung der Verschiedenheitsrelation

Da jede Linie von jeder anderen verschieden sein muß, besteht für jedes Linienmerkmal eine Verschiedenheitsrelation zu allen anderen Linien. Das gleiche gilt für die Punkte und die Flächen. Die Verschiedenheitsrelation muß also nicht codiert werden, und die Codierungslängen ergeben sich somit zu

$$\Omega_{\neq(\text{Punkte})} = 0 \quad [\text{Bit}], \quad \Omega_{\neq(\text{Linien})} = 0 \quad [\text{Bit}], \quad \Omega_{\neq(\text{Flächen})} = 0 \quad [\text{Bit}]. \quad (6.31)$$

## 6.2.3 Codierung der gesamten Gebäudeansicht

In den vorangegangenen Unterabschnitten wurde die Codierung der verwendeten topologischen und geometrischen Relationen vorgestellt. Zur vollständigen Codierung einer Gebäudeansicht müssen darüberhinaus auch die Anzahl der Punkte, Linien und Flächen der Gebäudeansicht sowie die Wertetabellen der gemeinsamen Informationen zu den jeweiligen Relationen codiert werden. Ferner muß für die Merkmale codiert werden, ob es sich um Dach- oder Wandmerkmale handelt.

Da einige Anteile der Codierung wie z.B. die Anzahl der Punkte, Linien und Flächen für die verschiedenen Gebäudeansichten immer die gleiche Länge besitzen, andere Teile jedoch, wie die oben gezeigten Relationen, eine variable Länge aufweisen, wird die Gesamtcodierungslänge als Summe der Anteile konstanter Länge  $\Psi$  und der Anteile variabler Länge  $\Omega$  ausgedrückt:

$$\text{Codierungslänge} = \Psi + \Omega \quad (6.32)$$

In den beiden folgenden Unterabschnitten wird zunächst die Zusammensetzung des konstanten Anteils  $\Psi$  und anschließend die des variablen Anteils  $\Omega$  erläutert.

### 6.2.3.1 Konstante Anteile

Zur Erfüllung der Präfixeigenschaft muß der gesamte Code der Gebäudeansicht unmittelbar dekodierbar sein. Da die Codelängen der verschiedenen Relationen neben der jeweiligen Anzahl von Relationentupeln unmittelbar von der Anzahl der Punkte, Linien und Flächen der Gebäudeansicht abhängen, müssen diese zuerst codiert werden. Damit Sender und Empfänger wissen, wieviele Bits für die Codierung dieser Anzahlen verwendet werden, einigt man sich vorher auf die über alle möglichen Modelle maximalen Anzahlen von Punkten  $|P|_{\max}$ , Linien  $|L|_{\max}$  und Flächen  $|F|_{\max}$ . Die notwendigen Codierungslängen ergeben sich dann zu

$$\Psi_{|P|} = \log_2 |P|_{\max} \quad [\text{Bit}] \quad \Psi_{|L|} = \log_2 |L|_{\max} \quad [\text{Bit}] \quad \Psi_{|F|} = \log_2 |F|_{\max} \quad [\text{Bit}] \quad (6.33)$$

Neben den strukturellen Ausprägungen der Relationen müssen auch ihre jeweiligen Gewichte codiert werden. Wie in Abschnitt 5.2 beschrieben, wird bei den Gewichten der einzelnen Graphkanten einer Relation zwischen Beziehungen zwischen Dachmerkmalen und Beziehungen zwischen Wandmerkmalen unterschieden. Zu beiden Beziehungsarten liegen die Gewichte in Form einer  $2 \times 2$ -Tabelle der gemeinsamen Information zwischen Bild- und Modellrelation vor (vgl. Tabelle 5.2 auf Seite 91). Zur Verwendung dieser Werte mit dem CLP(FD)-Constraint-Solver wurden diese, wie in Abschnitt 5.3 erläutert, durch die Division mit der Auflösung  $\varepsilon$  und anschließender Rundung zu ganzen Zahlen diskretisiert. Wenn nun  $R_{I_{\max}}$  den Wertebereich der gemeinsamen Information bezeichnet, der sich aus der Differenz des maximalen und minimalen Wertes über alle Tabellen

ergibt, ergeben sich die Codierungslängen für die Gewichte der Merkmalsnachbarschaft zwischen Linien und Punkten zu

$$\Psi_{\text{MNG}(\text{L,P}),\text{Dach}} = 4 \log_2 \frac{R_{I_{\max}}}{\varepsilon} \quad [\text{Bit}] \quad \Psi_{\text{MNG}(\text{L,P}),\text{Wand}} = 4 \log_2 \frac{R_{I_{\max}}}{\varepsilon} \quad [\text{Bit}] \quad (6.34)$$

Die Gewichte der Merkmalsnachbarschaften zwischen Fläche und Linie sowie Fläche und Punkt werden auf die gleiche Weise berechnet. Damit beträgt die Gesamtcodierungslänge der Gewichte des Merkmalsnachbarschaftsgraphen

$$\begin{aligned} \Psi_{\text{MNG}} &= \Psi_{\text{MNG}(\text{L,P}),\text{Dach}} + \Psi_{\text{MNG}(\text{L,P}),\text{Wand}} \\ &\quad + \Psi_{\text{MNG}(\text{F,L}),\text{Dach}} + \Psi_{\text{MNG}(\text{F,L}),\text{Wand}} \\ &\quad + \Psi_{\text{MNG}(\text{F,P}),\text{Dach}} + \Psi_{\text{MNG}(\text{F,P}),\text{Wand}} \quad [\text{Bit}]. \end{aligned} \quad (6.35)$$

Die Codierung der Gewichte der Linienparallelitäts-, Kollinearitäts- und Same\_Side\_Line-Relationen erfolgt analog. Wie in Abschnitt 5.2 gezeigt wurde, sind die Gewichte der drei Verschiedenheitsrelationen Null, so daß sich die Codierungslänge  $\Psi$  des konstanten Anteils der Gesamtcodierungslänge wie folgt zusammensetzt:

$$\Psi = \Psi_{|P|} + \Psi_{|L|} + \Psi_{|F|} + \Psi_{\text{MNG}} + \Psi_{\text{parallel}} + \Psi_{\text{collin}} + \Psi_{\text{SSL}} \quad [\text{Bit}]. \quad (6.36)$$

### 6.2.3.2 Variable Anteile

Neben der Codierung der Relationen und ihrer Gewichte muß noch codiert werden, für welche Relation welches Gewicht zu verwenden ist. Dabei geht es für die Relationen eines Typs (z.B. der Linienparallelität) um die Unterscheidung, ob es sich jeweils um eine Beziehung zwischen Wandmerkmalen oder Dachmerkmalen handelt. Diese Unterscheidung muß nicht extra codiert werden, wenn bekannt ist, ob es sich bei den jeweils beteiligten Merkmalen um Dach- oder Wandmerkmale handelt. Wie bereits in Abschnitt 5.2 auf Seite 90 erläutert wurde, läßt sich die Typunterscheidung, ob es sich bei einem Punkt- bzw. Linienmerkmal um ein Dach- oder ein Wandmerkmal handelt, allein aus dem Typ der angrenzenden Flächenmerkmale ableiten. Es muß also lediglich für die Flächen codiert werden, ob es sich um Wand- oder Dachflächen handelt.

Zur Codierung werden die Flächen nun so numeriert, daß alle Dachflächen eine kleinere Nummer erhalten als die Wandflächen, d.h. die ersten  $x$  Flächen sind Dachflächen. Um zu beschreiben, welche Flächen Dachflächen sind, reicht es somit aus, nur die Anzahl der Dachflächen anzugeben. Da die Gebäudeansicht aus  $|F|$  Flächen besteht, können auch nur maximal soviele Flächen Dachflächen sein, und die Codierungslänge  $\Omega_{|\text{Dachflächen}|}$  ergibt sich zu

$$\Omega_{|\text{Dachflächen}|} = \log_2 |F| \quad [\text{Bit}]. \quad (6.37)$$

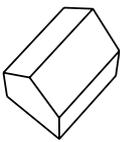
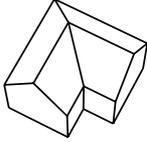
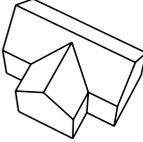
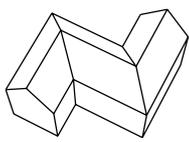
Die Codierungslänge  $\Omega$  des variablen Anteils der Gesamtcodierungslänge ergibt sich somit zu

$$\Omega = \Omega_{|\text{Dachflächen}|} + \Omega_{\text{MNG}} + \Omega_{\text{parallel}} + \Omega_{\text{SSL}} \quad [\text{Bit}]. \quad (6.38)$$

### 6.2.3.3 Beispiele für die Codierungslängen ausgewählter Gebäudeansichten

In diesem Abschnitt sollen die Auswirkungen verschiedener Codierschemata auf die Codierungslänge anhand einiger Beispiele illustriert werden. Ferner werden die Codierungslängen der Modelle gezeigt, die im darauffolgenden Abschnitt 6.3 für die Klassifikation von Gebäudeansichten in synthetischen und realen Bildern verwendet werden.

Der durch  $\Psi$  ausgedrückte Anteil der Codierungslänge besitzt — unabhängig von den konkreten Ausprägungen der einzelnen Gebäudeansichten — für alle Modelle dieselbe Länge. Er spielt beim Vergleich verschiedener Gebäudemodelle und somit bei der Bestimmung des wahrscheinlichsten Modells keine Rolle, und wird deshalb im folgenden nicht weiter betrachtet.

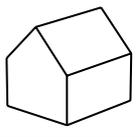
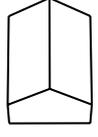
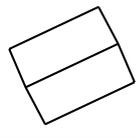
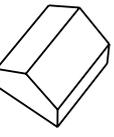
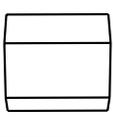
												
Punktzahl $ P $	9	14	19	17								
Linienanzahl $ L $	12	21	27	26								
Flächenanzahl $ F $	4	8	9	10								
Anzahl d. Relationen												
$ MNG(\text{Fläche}, \text{Linie}) $	17	34	42	41								
$ MNG(\text{Linie}, \text{Punkt}) $	24	42	54	52								
$ MNG(\text{Fläche}, \text{Punkt}) $	17	34	42	41								
$ \text{Linienparallelität} $	11	30	53	52								
$ \text{Kollinearität} $	24	42	54	52								
$ \text{Same\_Side\_Line} $	17	34	38	41								
Codierschema	AdjMat.	AdjLst.	Perm.	AdjMat.	AdjLst.	Perm.	AdjMat.	AdjLst.	Perm.	AdjMat.	AdjLst.	Perm.
$\Omega_{MNG(\text{Fläche}, \text{Linie})}$	48,00	<b>46,00</b>	47,53	168,00	<b>123,00</b>	125,76	243,00	<b>160,14</b>	165,43	260,00	<b>162,20</b>	167,62
$\Omega_{MNG(\text{Linie}, \text{Punkt})}$	108,00	<b>76,08</b>	85,85	294,00	<b>159,91</b>	178,24	513,00	<b>229,39</b>	253,92	442,00	<b>212,55</b>	235,67
$\Omega_{\text{parallel}}$	66,00	53,93	<b>35,58</b>	210,00	118,18	<b>70,99</b>	351,00	178,72	<b>107,50</b>	325,00	177,69	<b>109,36</b>
$\Omega_{SSL}$	660,00	17,00	<b>4,09</b>	3990,00	34,00	<b>5,09</b>	8775,00	42,00	<b>22,16</b>	7800,00	41,00	<b>5,36</b>
$\Omega_{ \text{Dachflächen} }$	<b>2,00</b>			<b>3,00</b>			<b>3,17</b>			<b>3,32</b>		
Code-Gesamtlänge $\Omega$	<b>163,75</b>			<b>361,98</b>			<b>522,36</b>			<b>492,78</b>		

**Tabelle 6.1:** Codierungslängen der Gebäudeansichten von vier unterschiedlich komplexen Haustypen in derselben Perspektive (von links nach rechts: einfaches, L-, T- und S-förmiges Satteldachhaus). Alle Codierungslängen sind in Bit angegeben.

**Tabelle 6.1** zeigt die Codierungslängen  $\Omega$  für vier Gebäudeansichten in ähnlicher Perspektive von vier unterschiedlich komplexen Gebäudehypothesen. Für jede Gebäudeansicht wird die Anzahl der Punkte, Linien und Flächen sowie die Anzahl der Relationen der unterschiedlichen Relationentypen aufgelistet. Darunter werden die Codierungslängen der einzelnen Relationentypen aufgeführt. Diese sind zum Vergleich für die drei Codierungsformen Adjazenzmatrix, Adjazenzliste und Permutationsreihe jeweils nebeneinander angegeben. Die letztlich zur Berechnung der Gesamtcodierungslänge  $\Omega$  verwendeten Werte sind fett gedruckt. Die Kardinalität der einzelnen Relationentypen, d.h. die maximal mögliche Anzahl von Relationen eines Typs für eine Gebäudeansicht, entspricht der Codierungslänge der jeweiligen Adjazenzmatrix.

An den jeweiligen Gesamtcodierungslängen  $\Omega$  läßt sich zunächst erwartungsgemäß erkennen, daß die Codierungslänge mit wachsender Komplexität des Modells größer wird. Es läßt sich auch erkennen, daß die Anzahl der Relationen eines Typs im Verhältnis zu der Kardinalität klein ist. Dies fällt insbesondere bei der Same\_Side\_Lines-Relation auf, wo beim S-förmigen Gebäudemodell nur 41 der theoretisch 7800 möglichen Relationen gegeben sind. Hier wird die Wichtigkeit der Berücksichtigung von Abhängigkeiten an der gewaltigen Einsparung in der Codierungslänge deutlich; es werden lediglich 5,36 Bit bei der Codierung als Permutationsreihe benötigt statt 7800 Bit als Adjazenzmatrix. Ferner kann man anhand des Verhältnisses der Codierungslänge der Permutationsreihe zur Länge der Adjazenzliste der SSL-Relationen erkennen, daß die meisten Flächen konvex sind. Beim T-förmigen Haus, bei dem es eine konkave Dachfläche gibt, fällt der Unterschied zwischen den beiden Codierungsformen entsprechend kleiner aus.

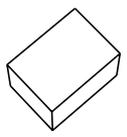
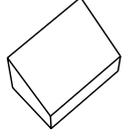
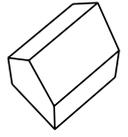
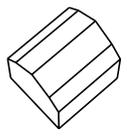
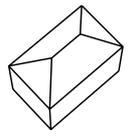
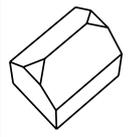
Bei den Codierungslängen der Linienparallelitätsrelation zeigt sich an der jeweils großen Differenz zwischen Permutationsreihe und Adjazenzliste, daß die Ersetzung der Cliques paralleler Linien durch minimale Spannbäume die Graphen stark ausdünn und damit die Codierungslänge substantiell verkürzt wird.

					
Punktanzahl $ P $	8	8	6	9	8
Linienanzahl $ L $	10	10	7	12	10
Flächenanzahl $ F $	3	3	2	4	3
$\Omega_{\text{MNG}}(\text{Fläche,Linie})$	30,60	30,60	15,00	46,00	29,02
$\Omega_{\text{MNG}}(\text{Linie,Punkt})$	60,00	60,00	36,19	76,08	60,00
$\Omega_{\text{parallel}}$	25,71	28,45	18,70	35,58	28,45
$\Omega_{\text{SSL}}$	3,70	3,70	3,00	4,09	3,58
$\Omega_{ \text{Dachflächen} }$	1,58	1,58	1,00	2,00	1,58
Gesamtcodierungslänge $\Omega$	121,59	124,33	73,89	163,75	122,63

**Tabelle 6.2:** Codierungslängen der fünf verschiedenen in Luftbildern beobachtbaren Ansichten eines Satteldachhauses. Alle Codierungslängen sind in Bit angegeben.

**Tabelle 6.2** zeigt die Codierungslängen der fünf verschiedenen in Luftbildern beobachtbaren Ansichten eines Satteldachhauses. Im Gegensatz zu Tabelle 6.1 werden nur noch die Punkt-, Linien- und Flächenanzahlen sowie die gemäß der Unterabschnitte 6.2.2.1 bis 6.2.2.4 berechneten Codierungslängen der einzelnen Relationen angegeben. Die ersten beiden Ansichten und die letzte Ansicht bestehen aus gleichvielen Punkten, Linien und Flächen. Die dritte Ansicht besitzt weniger und die vierte Ansicht mehr Merkmale, was sich auch entsprechend auf die Gesamtcodierungslängen  $\Omega$  auswirkt.

Die ersten beiden Ansichten unterscheiden sich nur in der Anzahl der Linienparallelitäten, alle anderen Relationen kommen in beiden Ansichten gleichhäufig vor. Bei der zweiten und der letzten Ansicht sind gleichviele Linien zueinander parallel, jedoch unterscheiden sich hier die Codierungslängen der SSL-Relation und der Merkmalsnachbarschaftsrelation zwischen Flächen und Linien, weil die letzte Ansicht nur zwei innenliegende Linien besitzt, im Gegensatz zur zweiten und ersten Ansicht, die drei Innenlinien besitzen.

						
Punktanzahl $ P $	7	7	9	13	9	13
Linienanzahl $ L $	9	9	12	18	14	18
Flächenanzahl $ F $	3	3	4	6	6	6
$\Omega_{\text{MNG}}(\text{Fläche,Linie})$	28,02	28,02	46,00	87,79	70,87	90,38
$\Omega_{\text{MNG}}(\text{Linie,Punkt})$	50,53	50,53	76,08	133,22	88,76	133,22
$\Omega_{\text{parallel}}$	26,06	23,69	35,58	61,31	39,42	53,76
$\Omega_{\text{SSL}}$	3,58	3,58	4,09	4,75	4,46	4,81
$\Omega_{ \text{Dachflächen} }$	1,58	1,58	2,00	2,58	2,58	2,58
Gesamtcodierungslänge $\Omega$	109,77	107,40	163,75	289,65	206,09	284,75

**Tabelle 6.3:** Codierungslängen der Gebäudeansichten von sechs verschiedenen einfachen Haustypen in derselben Perspektive (von links nach rechts: Flachdach-, Pultdach-, Satteldach-, Mansarddach-, Walmdach- und Krüppelwalmdachhaus). Alle Codierungslängen sind in Bit angegeben.

In **Tabelle 6.3** sind die Codierungslängen der Gebäudeansichten von sechs verschiedenen einfachen Haustypen in derselben Perspektive angegeben. Die Ansichten des Mansarddach- und Krüppelwalmdachhauses bzw. die des Flachdach- und Pultdachhauses bestehen jeweils aus gleich-

vielen Punkten, Linien und Flächen. Die ähnliche Größe der Codierungslängen  $\Omega$  besagt, daß die Gebäudeansichten für beide Paare jeweils ungefähr dieselbe strukturelle Komplexität aufweisen. Während sich das Pultdach- und Flachdachhaus ausschließlich in der Codierungslänge der Linienparallelitätsrelation unterscheiden, weichen das Mansarddach- und Krüppelwalmdachhaus darüberhinaus noch (minimal) bei der Anzahl der Merkmalsnachbarschaften voneinander ab.

Obwohl sich ein Walmdach- und Krüppelwalmdachhaus für den menschlichen Betrachter als sehr ähnlich darstellt, weichen die Codierungslängen deutlich voneinander ab. Der Grund besteht in den erheblichen strukturellen Unterschieden zwischen den beiden Haustypen, z.B. besitzt die Ansicht des Krüppelwalmdachhauses mehr Punkte und Linien. Dies führt zu einer signifikant größeren Codierungslänge der Merkmalsnachbarschaft zwischen Linien und Punkten. Die Codierung der Linienparallelitätsrelation ist nicht nur wegen der größeren Linienanzahl länger, sondern auch deshalb, weil beim Krüppelwalmdachhaus die Dachkanten an beiden Stirnseiten paarweise zueinander parallel sind.

## 6.3 Praktische Evaluation

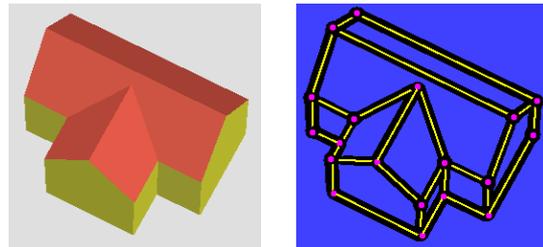
Da, wie in Abschnitt 6.2 erläutert wurde, die Optimalität der Codierung nicht analytisch bewiesen werden kann, soll nun die Tauglichkeit des entwickelten Konzepts anhand praktischer Tests überprüft werden. Es geht dabei um die Klärung der Frage, ob die vorgeschlagene Codierung nahe genug an der wahren Codierungslänge und damit der wahren a priori Wahrscheinlichkeit der verschiedenen Modelle liegt, um korrekte Klassifikationen zu ermöglichen. Bei der praktische Evaluation entscheidet sich jedoch nicht nur, ob die Codierung gut genug ist, sondern darüberhinaus, ob die Auswahl der topologischen und geometrischen Constraints überhaupt aussagekräftig bzw. spezifisch genug sind, auch sehr ähnliche Gebäude voneinander zu unterscheiden. Dazu werden im folgenden Unterabschnitt zunächst beispielhaft Tests mit synthetischen Bildern, die Projektionen typischer Gebäude zeigen, durchgeführt. Abschließend wird noch exemplarisch gezeigt, daß das Verfahren auch auf realen Daten eine korrekte Klassifikation ermöglicht.

Der Grund dafür, daß zunächst synthetische Daten genommen werden, besteht darin, daß diese Daten störungsfrei sind und deshalb eine perfekte Segmentierung und Merkmalsextraktion zu erwarten ist. Wenn sich zum einen herausstellen würde, daß die Bewertung bereits für optimales Bildmaterial nicht korrekt arbeitete, würde es gar keinen Sinn machen, diese auf reale Daten anzuwenden. Zum anderen ist bei Verwendung unverrauschter Daten die größte Differenz in der Bewertung der optimalen Zuordnung eines Bildes zu den einzelnen Modellen zu erwarten. Man erhält damit also auch einen Anhaltspunkt für die Trennschärfe der verwendeten Constraints und des entwickelten Codiervorgangs.

### 6.3.1 Gebäudeerkennung in synthetischen Bildern

Es wurden für drei Meßreihen synthetische Bilder von Häusern generiert. Auf diese Bilder wurde das in Kapitel 2.5.1 beschriebene Verfahren zur Merkmalsextraktion angewendet.

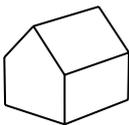
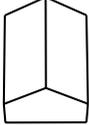
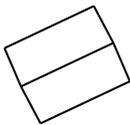
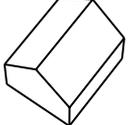
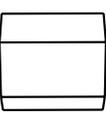
Abbildung 6.5 zeigt links das generierte Bild eines T-förmigen Satteldachhauses und rechts daneben die bei der Segmentierung extrahierten Bildmerkmale. Insgesamt entspricht das Ergebnis dem idealen Bildmodell des verwendeten Segmentierungsverfahrens, weil 1) alle Merkmale des Modells extrahiert werden konnten, 2) keine Fremd- oder Störmerkmale auftraten, 3) alle gemäß des Modells erwarteten Relationen beobachtet wurden und 4) keine Relationen gegeben sind, die nicht vom Modell erwartet wurden wie z.B. eine direkte Nachbarschaft zweier Liniensegmente.



**Abbildung 6.5:** Synthetisches Bild eines T-förmigen Satteldachhauses (links) und Ergebnis der Merkmalsextraktion (rechts).

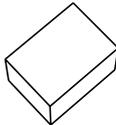
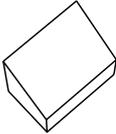
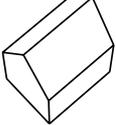
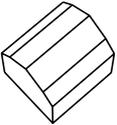
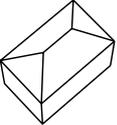
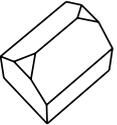
Für die bedingten Wahrscheinlichkeiten, die zur Berechnung der gemeinsamen Information benötigt werden, wurden für alle Meßreihen die in Tabelle 5.4 auf Seite 94 gezeigten, an realen Daten trainierten Wahrscheinlichkeiten verwendet. Auch die Constraint-Schwellwerte wurden aus diesen Trainingsdaten bestimmt. Das bedeutet, daß bei diesen Tests allein die Bilder künstlich erzeugt sind, und sich alle anderen Parameter an realen Daten orientieren.

Anhand der **ersten Meßreihe** sollte nun überprüft werden, ob die entwickelte Bewertungsfunktion in der Lage ist, korrekt zu entscheiden, welche Ansicht einer Gebäudehypothese im Bild zu sehen ist. Es wurden fünf Bilder generiert, welche die fünf möglichen verschiedenen Ansichten zeigen, in denen ein Satteldachhaus im Luftbild erscheinen kann. Für jedes Bild wurde dann eine Merkmalsextraktion durchgeführt. Anschließend wurde für jedes Bild die beste Zuordnung für alle fünf Modelle bestimmt und gemäß Gleichung 6.17 von der gemeinsamen Information dieser Zuordnungen jeweils die Codierungslängen der entsprechenden Modelle subtrahiert. Die Ergebnisse sind in Tabelle 6.4 in gegenüberstellender Form aufgeführt. Aus Gründen der Übersichtlichkeit wird nur die Gesamtbewertung angegeben; die gemeinsame Information der besten Zuordnung kann rekonstruiert werden, indem die Codierungslänge des entsprechenden Modells wieder zum jeweiligen Tabelleneintrag hinzuaddiert wird. Die Codierungslängen für diese Meßreihe können Tabelle 6.2 entnommen werden.

Bild	Modell (alle möglichen Luftbildansichten)				
					
	<b>26, 59</b>	-40, 01	-12, 59	-16, 91	-29, 58
	-27, 96	<b>40, 9</b>	-6, 13	-18, 64	-53, 46
	-49, 64	-50, 02	<b>21, 93</b>	-86, 45	-37, 24
	24, 56	-1, 1	-10, 73	<b>49, 04</b>	-38, 21
	-44, 64	-38, 28	19, 13	-46, 13	<b>32, 44</b>

**Tabelle 6.4:** Satteldachhaus in fünf verschiedenen Ansichten. Das wahrscheinlichste Modell mit der höchsten Bewertung wurde für jede Zeile fett markiert. Der Anordnung der besten Zuordnungen auf der Diagonalen ist entnehmbar, daß für die fünf Bilder jeweils die richtige Ansicht erkannt wurde.

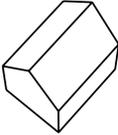
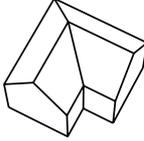
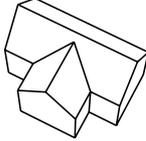
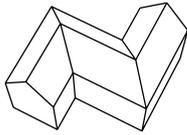
In jeder Zeile bezeichnet der Eintrag mit der höchsten Bewertung das wahrscheinlichste Modell für das entsprechende Bild. An der Dominanz der Hauptdiagonalen der Tabelle kann man erkennen, daß alle Bilder richtig klassifiziert wurden. Dabei ist der Abstand des höchstbewerteten Modells zu den anderen Modellen in jeder Zeile verhältnismäßig groß. Beim letzten Bild in der untersten Zeile ist der Unterschied mit 13,31 Bit zwar am geringsten, dieser ist jedoch im Verhältnis zu den absoluten Werten immer noch als eindeutig zu bezeichnen. Der Grund, warum ein Großteil der Tabelleneinträge negativ ist, liegt darin, daß in diesen Fällen jeweils die Codierungslänge des Modells größer war als die gemeinsame Information der besten Zuordnung des Modells zu den Bilddaten.

Bild	Modell (jeweils nur eine Ansicht)					
						
	<b>17,63</b>	12,5	-45,8	-175,35	-79,37	-170,33
	1,98	<b>13,61</b>	-55,31	-159,29	-98,9	-173,59
	-24,39	3,68	<b>43,12</b>	-86,81	-15,82	-103,8
	-13,81	6,56	15,0	<b>70,08</b>	-60,72	-73,35
	20,84	15,72	5,4	-162,04	<b>41,57</b>	-113,89
	-17,37	-0,12	23,63	-73,53	-53,66	<b>115,02</b>

**Tabelle 6.5:** Sechs verschiedene einfache Haustypen in derselben Perspektive werden einander zugeordnet (von links nach rechts bzw. von oben nach unten: Flachdach-, Pulldach-, Satteldach-, Mansarddach-, Walmdach- und Krüppelwalmdachhaus). Das wahrscheinlichste Modell mit der höchsten Bewertung wurde für jede Zeile fett markiert. Der Anordnung der besten Zuordnungen auf der Diagonalen ist entnehmbar, daß für die sechs Bilder jeweils die richtige Ansicht und damit auch der richtige Haustyp erkannt wurde.

Die **zweite Meßreihe** sollte feststellen, wie gut die Trennschärfe des Verfahrens bei ähnlichen Haustypen ist. Dazu wurde von den sechs einfachen Haustypen Flach-, Pult-, Sattel-, Mansard-, Walm- und Krüppelwalmdachhaus jeweils ein Bild aus derselben Perspektive generiert und eine Merkmalsextraktion durchgeführt. Die Ergebnisse der anschließenden Erkennung und Klassifikation sind in Tabelle 6.5 aufgelistet.

Analog zur ersten Meßreihe bezeichnet auch hier jeweils der Eintrag mit der höchsten Bewertung einer Zeile das wahrscheinlichste Modell, und auch hier läßt sich wiederum der Dominanz der Hauptdiagonalen entnehmen, daß alle abgebildeten Häuser richtig klassifiziert wurden. Im Unterschied zur ersten Meßreihe wurde dabei nicht zwischen den verschiedenen Ansichten einer Gebäudehypothese sondern aus den gleichperspektivischen Ansichten verschiedener Gebäudehypothesen gewählt. Diese Situation ist typisch für den Fall, daß ursprünglich nur wenige 3D-Gebäudeecken rekonstruiert wurden, und deshalb verschiedene passende Gebäudehypothesen zu diesen Eckenbeobachtungen generiert werden (vgl. Abbildung 2.12 auf Seite 29). Für diese Hypothesen ist oftmals die Perspektive aufgrund der zugeordneten 3D-Ecken bereits eindeutig bestimmt (siehe Abschnitt 2.5.5).

Bild	Modell (jeweils nur eine Ansicht)			
				
	<b>50,43</b>	-151,37	-328,68	-339,83
	7,9	<b>102,6</b>	-137,99	-85,44
	36,18	70,08	<b>163,8</b>	-100,05
	-8,59	61,8	-141,37	<b>164,57</b>

**Tabelle 6.6:** Klassifikationsergebnisse für vier Ansichten von Gebäuden unterschiedlicher Komplexität. Das wahrscheinlichste Modell mit der höchsten Bewertung ist für jede Zeile fett gedruckt. Der Anordnung der besten Zuordnungen auf der Diagonalen ist entnehmbar, daß für die vier Bilder jeweils die richtige Gebäudehypothese erkannt wurde.

Der Abstand der Bewertung zwischen Pult- und Flachdachhaus ist erwartungsgemäß relativ klein. Dies läßt sich dadurch erklären, daß der einzige Unterschied zwischen den relationalen Strukturen der beiden gezeigten Ansichten in den Linienparallelitäten zwischen Dach- und Bodenkanten an den Gebäudestirnseiten besteht. Der Abstand der Bewertung zwischen Walmdach- und Krüppelwalmdachhaus ist hingegen in jeder Zeile sehr groß. Dies mag auf den ersten Blick überraschend

erscheinen, da der menschliche Betrachter die beiden Haustypen als relativ ähnlich wahrnimmt. Der Grund für die große Differenz liegt darin begründet, daß der strukturelle Unterschied zwischen diesen beiden Haustypen in bezug auf die verwendeten Relationen jedoch sehr groß ist. Dies läßt sich an den in Tabelle 6.3 auf Seite 116 gezeigten Codierungslängen der sechs Gebäudeansichten nachvollziehen.

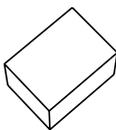
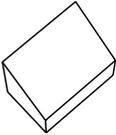
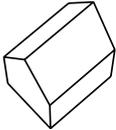
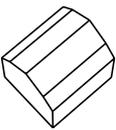
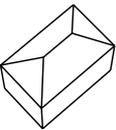
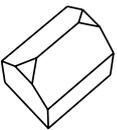
Nachdem an den ersten beiden Meßreihen erkennbar ist, daß die Bewertungsfunktion sowohl die richtige Ansicht eines abgebildeten einfachen Haustyps zu identifizieren vermag, als auch gleiche Ansichten verschiedener, jedoch ähnlich komplexer Häuser richtig klassifizieren kann, sollte mit der **dritten Meßreihe** nachgeprüft werden, ob auch verschieden komplexe Häuser korrekt klassifiziert werden können. Dazu wurde jeweils ein Bild eines einfachen, L-, T- und S-förmigen Satteldachhauses in derselben Perspektive generiert. Auch für diese Bilder wurde eine Merkmalsextraktion und anschließende Klassifikation durchgeführt. Die Ergebnisse sind in Tabelle 6.6 angegeben.

Wie bei den beiden Meßreihen zuvor, wurden auch hier alle Gebäudeansichten korrekt klassifiziert, wobei die Differenz zwischen der jeweils bestbewertetsten in bezug zu den anderen Ansichten aufgrund der großen strukturellen Unterschiede sehr groß sind.

Insgesamt läßt sich feststellen, daß in allen drei Meßreihen die Gebäudeansichten und in den Meßreihen zwei und drei damit implizit auch der Gebäudetyp in allen Fällen korrekt klassifiziert wurde. Der Abstand der bestbewertetsten Hypothesen zu den jeweils anderen ist selbst bei der Unterscheidung zwischen den ähnlichen Gebäudetypen Pult- und Flachdachhaus in der zweiten Meßreihe verhältnismäßig groß. Dies ist auch wichtig, um einen genügend großen Spielraum für die aufgrund von Rauschen und anderen Störungen zu erwartende Reduktion der gemeinsamen Information der einzelnen Zuordnungen zu besitzen, wenn das Verfahren auf reale Luftbilddaten angewendet wird.

### 6.3.1.1 Empfindlichkeit der Klassifikation gegenüber der Codiermethode

Die Wichtigkeit der sorgfältigen Wahl der Codierverfahren für die einzelnen Relationen soll nun noch an einem Beispiel verdeutlicht werden. Tabelle 6.7 zeigt die Auswirkungen der Verwendung schlechterer Codierverfahren auf die Bewertungen bei der Klassifikation eines Walmdachhauses.

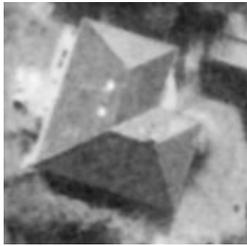
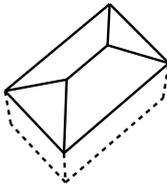
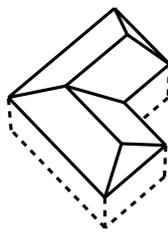
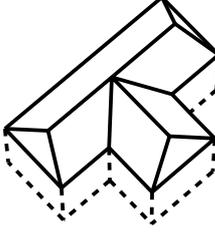
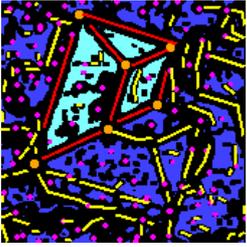
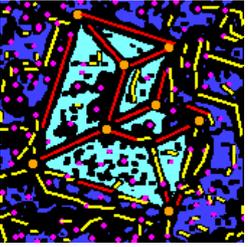
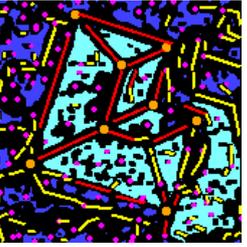
Bild		Modelle					
							
<b>A</b>	$\Delta\Omega_{\text{parallel}}$	11,43	10,63	18,35	37,07	23,57	36,28
	$I_{\tilde{h}} - \Omega - \Delta\Omega$	9,41	5,09	-12,95	-199,11	<b>18,00</b>	-150,17
<b>B</b>	$\Delta\Omega_{\text{SSL}}$	8,42	8,42	12,91	22,25	17,54	23,19
	$I_{\tilde{h}} - \Omega - \Delta\Omega$	12,42	7,3	-7,51	-184,29	<b>24,03</b>	-137,08
<b>C</b>	$I_{\tilde{h}} - \Omega$ $-\Delta\Omega_A - \Delta\Omega_B$	<b>0,99</b>	-3,33	-25,86	-221,36	0,46	-173,36
<b>D</b>	$\Delta\Omega_{\text{MNG(L,P)}}$	12,47	12,47	31,92	100,78	37,24	100,78
	$I_{\tilde{h}} - \Omega - \Delta\Omega$	<b>8,37</b>	3,25	-26,52	-262,82	4,33	-214,67

**Tabelle 6.7:** Auswirkungen längerer Codierungen auf das Klassifikationsergebnis. Zeile A zeigt oben die Vergrößerung der Codierungslänge  $\Delta\Omega_{\text{parallel}}$ , wenn die Linienparallelitätsrelation als Adjazenzliste statt Permutationsreihe codiert wird. Darunter steht das jeweils damit erreichte Gesamtergebnis. Zeile B zeigt den Fall, daß die Same\_Side\_Lines-Relation als Adjazenzliste statt Permutationsreihe codiert wird, und Zeile C die Kombination der Zeilen A und B. Die Codierung der Merkmalsnachbarschaft zwischen Punkt und Linie als Adjazenzmatrix statt Adjazenzliste wird in Zeile D gezeigt. In den Fällen C und D kommt es durch die längere Codierung zu Fehlklassifikationen.

Es werden vier Fälle überprüft, deren Auswirkungen jeweils in den Zeilen A bis D aufgetragen sind. Im ersten Fall (A) wird die Linienparallelitätsrelation nicht mittels einer Permutationsreihe sondern durch eine Adjazenzliste codiert.  $\Delta\Omega_{\text{parallel}}$  gibt dabei für jede Gebäudehypothese an, um wieviele Bits sich die Codierungslänge vergrößert. Darunter befinden sich die entsprechend angepaßten Gesamtbewertungen (vgl. mit der vorletzten Zeile von Tabelle 6.5). Im zweiten Fall (B) wird angenommen, daß die Same\_Side\_Lines-Relation durch eine Adjazenzliste statt durch eine Permutationsreihe codiert wird, wobei  $\Delta\Omega_{\text{SSL}}$  die entsprechende Codeverlängerung mißt. In beiden Fällen wird nachwievordas korrekte Modell identifiziert. Werden die Codierungen A und B jedoch gleichzeitig verwendet (Fall C), kommt es zu einer Fehlklassifikation. Der letzte Fall (D) zeigt schließlich, daß bereits die Codierung der Merkmalsnachbarschaft zwischen Punkt und Linie als Adjazenzmatrix statt Adjazenzliste zu einer Fehlklassifikation führt. Aufgrund der starken Überschätzung der Codierungslänge wird bei beiden Fehlklassifikation C und D übrigens das einfachste Modell gewählt.

### 6.3.2 Gebäudeerkennung in realen Bildern

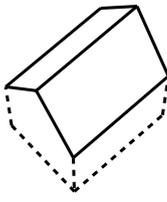
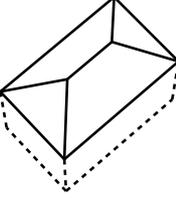
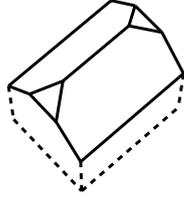
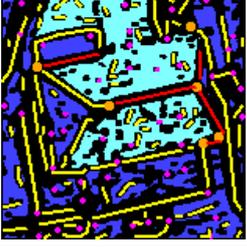
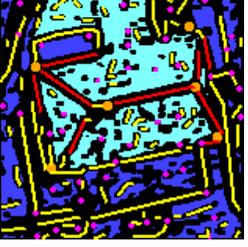
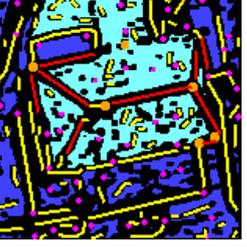
Im vorangegangenen Abschnitt wurde festgestellt, daß sich das entwickelte Codierschema für die Erkennung und Klassifikation von Gebäuden in synthetischen Bildern eignet. Obwohl eine empirische Untersuchung über die erreichbare Qualität und Klassifikationsgenauigkeit für größere Luftbildszenen nicht im Rahmen dieser Arbeit liegt, soll an dieser Stelle abschließend die Übertragbarkeit auf reale Daten beispielhaft an zwei Luftbildausschnitten demonstriert werden.

Bild	Modelle		
			
beste Zuordnung $\hat{h}$			
$I_{\hat{h}}(\text{Modell}; \text{Bild})$	85,2	224,24	237,56
$\Omega$	107,31	206,09	325,19
$I_{\hat{h}}(\text{Modell}; \text{Bild}) - \Omega$	-22,11	<b>18,15</b>	-87,63

**Abbildung 6.6:** Identifikation des wahrscheinlichsten Modells für das am Anfang des Kapitels in Abb. 6.1 auf Seite 100 gezeigte Beispiel. In jeder Spalte werden das Modell, das im Bild gesucht wurde, und die Bildmerkmale der diesbzgl. besten Zuordnung  $\hat{h}$  gezeigt. Darunter wird die gemeinsame Information  $I_{\hat{h}}$  der besten Zuordnung zwischen Bild und Modell, sowie die Codierungslänge  $\Omega$  des Modells gezeigt. Der untersten Zeile läßt sich entnehmen, daß das abgebildete Haus korrekt als L-förmiges Gebäude klassifiziert wird.

Dazu wird als erstes das am Anfang dieses Kapitels gezeigte Beispiel wieder aufgegriffen, in dem ein L-förmiges Walmdachhaus abgebildet ist, und für welches aus den drei generierten Gebäudehypothesen Walmdachhaus, L-förmiges Walmdachhaus und T-förmiges Walmdachhaus die richtige Ansicht identifiziert werden sollte (siehe Abbildung 6.1 auf Seite 100). Der untersten Zeile von Abbildung 6.6 ist entnehmbar, daß nach der Subtraktion der Codierungslängen der Modelle von den jeweiligen Zuordnungsergebnissen, das L-förmige Haus die höchste Gesamtbewertung erhält und somit als wahrscheinlichste Interpretation der Szene eingeschätzt wird.

Bei dem zweiten Beispiel geht es — analog zur zweiten Meßreihe mit den synthetischen Daten — darum, aus einer Menge ähnlich komplexer Gebäudehypothesen die dem abgebildeten Haus entsprechende Ansicht zu identifizieren. Abbildung 6.7 zeigt ein Walmdachhaus, für das die drei Gebäudehypothesen Satteldach-, Walmdach- und Krüppelwalmdachhaus aufgestellt wurden. Wie beim vorigen Beispiel erhält auch hier die dem abgebildeten Haus entsprechende Ansicht nach der Subtraktion der Codierungslängen von den jeweiligen Zuordnungsergebnissen die höchste Gesamtbewertung.

Bild	Modelle		
			
beste Zuordnung $\hat{h}$			
$I_{\hat{h}}(\text{Modell}; \text{Bild})$	63,27	115,64	98,82
$\Omega$	72,13	107,31	172,16
$I_{\hat{h}}(\text{Modell}; \text{Bild}) - \Omega$	-8,86	<b>8,33</b>	-73,34

**Abbildung 6.7:** Identifikation der wahrscheinlichsten Ansicht (und damit implizit der entsprechenden Gebäudehypothesen) für reale Bilddaten. In jeder Spalte werden das Modell, das im Bild gesucht wurde, und die Bildmerkmale der diesbzgl. besten Zuordnung  $\hat{h}$  gezeigt. Darunter wird die gemeinsame Information  $I_{\hat{h}}$  der besten Zuordnung zwischen Bild und Modell, sowie die Codierungslänge  $\Omega$  des Modells gezeigt. Der untersten Zeile läßt sich entnehmen, daß das abgebildete Gebäude korrekt als Walmdachhaus klassifiziert wird.

Aufgrund der geringen Bodenauflösung der Luftbildausschnitte wurden in beiden Beispielen nur die Dächer der Modelle gesucht und bewertet. Konsequenterweise wurden bei der Berechnung der Codierungslängen gleichfalls nur Dachmerkmale und Relationen zwischen Dachmerkmalen berücksichtigt. Die Ergebnisse zeigen dabei, daß das Verfahren auch für die Klassifikation von Teilstrukturen verwendet werden kann.



# Kapitel 7

## Zusammenfassung

In dieser Arbeit wurde ein Verfahren zur automatischen Identifikation und Rekonstruktion von Gebäuden in Luftbildern vorgestellt. Dieses Konzept basiert auf der Verwendung eines komponentenbasierten, generischen Gebäudemodells, das die Repräsentation und Unterscheidung einer Vielzahl verschiedener Gebäudetypen erlaubt. Die Erkennungsstrategie basiert auf der Generierung und Verifikation von Gebäudehypothesen, wobei der Fokus dieser Arbeit auf der Verifikation liegt. Dazu mußten insbesondere die Probleme der Suche nach Gebäuden im Bild und der Bewertung der Zuordnungen von Gebäudehypothesen zu den Bilddaten gelöst werden.

Es wurde gezeigt, daß Constraint-Techniken eine effiziente Suche der Gebäude im Bild ermöglichen. Zu diesem Zweck wurden verschiedene geometrische und topologische Constraints mit ihren Eigenschaften vorgestellt, die einerseits zur relationalen Repräsentation von Gebäudehypothesen dienen und andererseits im Kontext des Constraint Logic Programming gleichzeitig auch ein Programm zur Suche dieser Hypothesen im Bild definieren. Um die in realen Bilddaten häufig auftretenden Störungen handhaben zu können, wurde eine Erweiterung der Constraint-Modellierung vorgenommen, die die Effekte der Unbeobachtbarkeit und Unschärfe bzw. Unsicherheit explizit berücksichtigt.

Durch die Verwendung einer informationstheoretisch begründeten Bewertungsfunktion wurden Constraints so gewichtet, daß die beste Lösung der Maximum-Likelihood-Schätzung zwischen Modell und Bild entspricht. Die Auswahl der wahrscheinlichsten Gebäudehypothese aus einer Reihe möglicher Gebäudemodelle erfolgt mittels des Prinzips der minimalen Beschreibungslänge. Dazu wurde in dieser Arbeit ein Codierverfahren für relationale Gebäudemodelle entwickelt, das in Verbindung mit der informationstheoretischen Bewertung relationaler Zuordnungen die Bestimmung der wahrscheinlichsten Gebäudehypothese für eine gegebene Szene ermöglicht. Anhand einer praktischen Evaluation an synthetischen und realen Bilddaten wurde beispielhaft gezeigt, daß die entstandene Bewertungsfunktion in der Lage ist, sowohl ähnlich komplexe als auch völlig verschieden komplexe Gebäude miteinander zu vergleichen und die zutreffende Hypothese auszuwählen.

Die Anwendbarkeit des vorgestellten Gesamtkonzepts zur automatischen Extraktion von Gebäuden aus Luftbildern, insbesondere der Modellierung sowie der Strategie — und damit auch der einzelnen Teilschritte — wurde erfolgreich an dem Avenches-Luftbilddatensatz<sup>1</sup> getestet, der aus einer vierfach überdeckten Szene mit zwölf freistehenden Gebäuden unterschiedlicher Art besteht. Für die Merkmalsextraktion wurde eine Bodenpixelauflösung von 15 cm gewählt. Im Durchschnitt konnten nur knapp die Hälfte der Gebäudeecken durch die 3D-Eckenrekonstruktion bestimmt werden. Anschließend wurden für jedes Gebäude mehrere verschiedene Gebäudehypothesen mit ihren

---

<sup>1</sup>Der Avenches-Datensatz wurde im Vorfeld des Workshops „Automatic Extraction of Man-Made Objects from Aerial and Space Images“ in Ascona 1995 (siehe [GKA95]) von der ETH Zürich bereitgestellt. Er ist über das Internet erhältlich und wird in [MBS94] beschrieben.

entsprechenden Ansichten generiert, die durchweg noch freie Parameter aufwiesen. Für alle abgebildeten Gebäude konnte dann mittels der in dieser Arbeit entwickelten Zuordnungs- und Klassifikationsverfahren die richtige Gebäudehypothese erkannt werden. Insgesamt wurden alle zwölf Gebäude korrekt identifiziert und vollständig rekonstruiert, wobei die Unsicherheit der Parameter im Bereich weniger Zentimeter liegt. Die Ergebnisse sind im einzelnen in [FKL<sup>+</sup>98] aufgeführt.

## 7.1 Rückblick und Bewertung

In diesem Abschnitt wird untersucht, inwieweit die am Anfang der Arbeit formulierten Ziele erreicht werden konnten. Die drei wesentlichen Ziele waren dabei: (1) Die Identifikation geeigneter Relationen zu Gebäudeerkennung, (2) die Verwendung von Constraint-Techniken zur Suche nach Gebäuden und (3) die Entwicklung einer Bewertungsfunktion, die zudem mit den Constraint-Techniken gekoppelt werden sollte. Eine generelle Vorgabe war, daß die zu entwickelnden Konzepte nicht nur unter idealisierenden Annahmen funktionieren, sondern sich insbesondere für die Anwendung auf reale Luftbilddaten eignen sollen. Im folgenden werden, nach den drei Hauptzielen gegliedert, die entwickelten Konzepte im einzelnen diskutiert, wobei auf die besonderen Schwierigkeiten jeweils näher eingegangen wird.

### Relationale Repräsentation von Gebäudeansichten

Zum Zeitpunkt der Suche nach den Gebäudehypothesen im Bild weisen diese in der Regel noch verschiedene Freiheitsgrade in bezug auf die Ausdehnung, Orientierung und Lage im Bild auf. Darüberhinaus können durch Verdeckungen oder Bildstörungen Teile der Gebäudeansichten unbeobachtbar sein. Aus diesem Grund wurden die Gebäudeansichten in ihre Bestandteile, den Punkt-, Linien- und Flächenmerkmalen sowie den dazwischen bestehenden Relationen zerlegt. Zur Modellierung der Beziehungen zwischen den Merkmalen wurden insgesamt sechs verschiedene geometrische und topologische Relationen identifiziert, die invariant gegenüber der konkreten geometrischen Ausprägung der Gebäudeansichten und gleichzeitig relativ robust gegenüber Störungen sind. Es wurde beschrieben, wie aus der Randflächendarstellung der Gebäudeansichten die relationale Repräsentation für die verschiedenen Relationen hergeleitet werden kann.

Alle Relationen sind so definiert, daß sie kategorisch zwischen „erfüllt“ und „verletzt“ unterscheiden. Bei den geometrischen Relationen werden dazu Schwellwerte verwendet, die anhand von Trainingsdaten bestimmt werden. Die Schwellwerte relativieren die verschiedenen Effekte der Unschärfe und Unsicherheit sowie geringe Abweichungen abgebildeter Häuser von der Gebäudemodellierung (z.B. kleinere Dachüberstände). Darüberhinaus egalisiert der Schwellwert der Linienparallelitätsrelation auch die perspektivische Verzerrung der Zentralprojektion. Aufgrund der nicht weiter differenzierten Bewertung der einzelnen Relationen definieren diese ein verhältnismäßig grobes Raster. Um eine für die Gebäudeerkennung ausreichende Spezifität zu erreichen, müssen deshalb die zu erkennenden relationalen Modellstrukturen eine ausreichende Größe besitzen. Aus diesem Grund wird nicht nach einzelnen Gebäudeteilen, sondern nach vollständigen Gebäudeansichten in den Bildern gesucht.

Es wurde an praktischen Beispielen mit realen Luftbilddaten gezeigt, daß die Repräsentation von Gebäudeansichten durch Relationen der sechs verschiedenen Arten eine ausreichende Spezifität besitzt, um die Bildmerkmale, die zu den Merkmalen der Gebäudeansichten korrespondieren, vom Bildhintergrund diskriminieren zu können. Ferner wurde bei der Bestimmung des wahrscheinlichsten Modells demonstriert, daß die ausgewählten Relationen eine ausreichende Trennschärfe besitzen, um ähnliche Gebäudemodelle allein auf Basis ihrer relationalen Repräsentation voneinander unterscheiden zu können.

### Verwendung von Constraint-Techniken und Constraint Logic Programming

Zur Suche nach den Gebäudeansichten werden die Luftbilder segmentiert, wobei Punkt-, Linien und Flächenmerkmale extrahiert werden. Die Identifikation einer Gebäudeansicht im Bild erfolgt durch die Zuordnung von extrahierten Bildmerkmalen zu den Modellmerkmalen, so daß die Bildmerkmale die vom Modell erwarteten Relationen erfüllen (Relational Matching). Die kategorische

Unterscheidung der Relationen in „erfüllt“ und „nicht erfüllt“ erlaubt die Darstellung des Zuordnungsproblems als Constraint Satisfaction Problem, wobei die Modellmerkmale als Variablen und die Modellrelationen als Constraints auf diesen interpretiert werden. Die Aufgabe besteht darin, jeder Modellvariablen ein passendes Bildmerkmal zuzuweisen, so daß alle Constraints erfüllt sind. Der Vorteil der Constraint-Techniken im Gegensatz zu anderen Suchverfahren besteht darin, daß Constraints die netzwerkartige Struktur der Constraint-Probleme zur aktiven Propagierung von Einschränkung ausnutzen, und damit den Suchraum deutlich beschränken können.

Zwar ist die Verwendung von Constraint-Techniken für die Bilderkennung nicht neu, die Effekte der Unbeobachtbarkeit und Unschärfe wurden aber bislang nicht methodisch berücksichtigt. Meistens wurde implizit die idealisierende Annahme gemacht, daß entweder keine störenden Bildfehler auftreten oder sich diese nicht auf die Lösbarkeit des gestellten Constraint Satisfaction Problems auswirken (vgl. [Wal75, Bro83, Sch96, DH97]). Während in Kapitel 3 zur Einführung der grundlegenden Ideen der Constraint-Techniken und der Entwicklung einer relationalen Gebäudemodellierung ebenfalls zunächst eine solche Annahme gemacht wurde, wird in Kapitel 4 dieser Arbeit hingegen gezeigt, daß reale Luftbilddaten viele verschiedene Formen von Bildstörungen aufweisen können, die eine Unlösbarkeit des Constraint-Problems zur Folge haben, und somit die entsprechenden Gebäudeansichten nicht gefunden werden. Die verschiedenen Bildstörungen wurden systematisch daraufhin untersucht, welche Auswirkungen sie auf die verschiedenen Constraints haben können. Es wurde festgestellt, daß zur Berücksichtigung der Störungen letztlich explizit zwischen der Unbeobachtbarkeit von Merkmalen und der einfachen Verletzung von Relationen zwischen beobachteten Merkmalen unterschieden werden muß.

In den letzten Jahren wurden im Bereich des Constraint Satisfaction verschiedene Ansätze zur Lösung überbeschränkter Constraint-Probleme entwickelt. In dieser Arbeit wurde gezeigt, daß keines der bestehenden Konzepte in der Lage ist, die Effekte der Unbeobachtbarkeit und Unschärfe bzw. Unsicherheit adäquat zu berücksichtigen. Es wurde deshalb eine Erweiterung der Constraint-Modellierung entwickelt, die beide Phänomene explizit repräsentiert und integriert. Da mit dieser Erweiterung eine Vergrößerung des Suchraumes einhergeht, spielt die Einbeziehung von Vorwissen eine wichtige Rolle. Das verwendete Vorwissen stammt hauptsächlich aus der 3D-Eckenrekonstruktion, bei der bereits Bildmerkmale den Gebäudeecken zugeordnet wurden. Wie die praktischen Beispiele gezeigt haben, wirkt sich die Anzahl a priori bekannter Zuweisungen deutlich auf die Laufzeit zur Bestimmung der besten Zuordnung aus.

Die Realisierung der Constraint-Verfahren erfolgte auf Basis des Constraint Logic Programming (CLP). CLP stellt eine Erweiterung der logischen Programmierung um Strategien zur effizienten Ausnutzung und Verarbeitung von Constraints auf speziellen Wertebereichen dar. Die Repräsentation von Gebäudeansichten durch die konjunktive Verknüpfung der Constraints ist in CLP gleichzeitig als Wissensrepräsentation und Programm zu sehen. In CLP stellt jeder Constraint eine reaktive Datenstruktur dar, die eine eigene Auswertestrategie besitzt und auf Veränderungen der Variablen unterschiedlich reagieren kann. Die Auswertestrategien werden mittels Inferenzregeln beschrieben, die definieren, wann Suchraumeinschränkungen aktiv propagiert werden können. Es wurde gezeigt, daß die differenzierte Verwendung der Forward-Checking- und Look-Ahead-Inferenzregeln für die einzelnen Constraints einen erheblichen Performanzgewinn bei der Suche erbringt. Darüberhinaus wurde gezeigt, wie unter Nutzung der CLP-Techniken das Problem der Fragmentierung durch die Einbeziehung einer modellbasierten Gruppierung in den Suchprozeß gelöst werden kann, ohne die Komplexität des Verfahrens wesentlich zu erhöhen.

### **Bewertungsfunktion**

Aufgrund von Störungen in den Bildern sind oftmals nicht alle Merkmale und Relationen der Gebäudeansichten beobachtbar. Deshalb wird die bestmögliche Zuordnung von Bild- zu Modellmerkmalen gesucht. Dazu wird eine Bewertungsfunktion benötigt, die für eine gegebene Zuordnung die Ähnlichkeit der Modell- und zugeordneten Bildstrukturen mißt. Darüberhinaus muß eine vergleichende Bewertung verschiedener Gebäudeansichten möglich sein, da für eine Luftbildszene unterschiedliche Gebäudehypothesen in Frage kommen können, und die zutreffende ausgewählt werden muß.

In dieser Arbeit wird — angefangen bei der Suche nach den einzelnen Gebäudeansichten bis zur Bestimmung des wahrscheinlichsten Modells — durchgehend ein auf der Informationstheorie basierendes Bewertungsschema verwendet. Die Bewertung bezieht sich bildseitig immer auf die einmalig extrahierten Bildmerkmale und modellseitig ausschließlich auf die relationale Repräsentation der Gebäudeansichten. Zur Messung der Ähnlichkeit von Modell- und zugeordneten Bildmerkmalen wird eine von Boyer und Kak vorgestellte sowie von Vosselman weiterentwickelte Bewertungsfunktion verwendet. Diese mißt die Ähnlichkeit zweier relationaler Strukturen durch Summation der Ähnlichkeiten der einzelnen Merkmale sowie deren Relationen, wobei eine explizite Berücksichtigung der Unbeobachtbarkeit einzelner Relationen und ganzer Merkmale erfolgt. Mit dieser Bewertungsfunktion wird eine Maximum-Likelihood-Schätzung zwischen Modell und Bild erreicht.

Die Bewertungsfunktion wurde in dieser Arbeit so erweitert, daß sie zur Bestimmung des wahrscheinlichsten Modells (Maximum-A-Posteriori-Schätzung) verwendet werden kann. Dazu wurde das Prinzip der minimalen Beschreibungslänge verwendet, welches es ermöglicht, auch unterschiedlich komplexe Modelle miteinander vergleichen zu können. Zu diesem Zweck wurde ein Verfahren zur kompakten Codierung der relationalen Gebäuderepräsentationen entwickelt. Der Vorteil dieses Vorgehens besteht darin, daß die Ergebnisse der Bestimmung der besten Zuordnungen von Gebäudeansichten zu den Daten nicht nur qualitativ sondern auch quantitativ — und insbesondere auf wahrscheinlichkeitstheoretisch konsistente Weise — in die anschließende Klassifikation eingehen. Die praktischen Tests haben gezeigt, daß zur Verifikation der Hypothesen die polymorphen extrahierten Bildmerkmale ausreichen und nicht, wie z.B. in den Ansätzen von Schutte et. al. [Sch96, SSH97b] oder Fua und Hanson [FH87, FH89a], erneut auf die Radiometrie der Bilddaten zurückgegriffen werden muß. Letzteres könnte man auch als Nachteil auslegen, mit der Begründung, daß die extrahierten Merkmale u.U. nicht reichhaltig genug attribuiert oder zu fehlerbehaftet sein könnten, und deshalb nochmal die Bilddaten angeschaut werden sollten. Die praktischen Tests haben jedoch gezeigt, daß die verwendeten Merkmalseigenschaften und Relationen eine ausreichende Spezifität sowohl für die Diskriminierung von Gebäuden bzgl. des Bildhintergrunds als auch deren Klassifikation in verschiedene Haustypen bereitstellen.

Der wesentliche Vorteil des vorgestellten Verfahrens besteht darin, daß es gänzlich ohne frei zu wählende und abzustimmende Parameter auskommt. Ferner werden im Gegensatz zu anderen Arbeiten für die Bewertung keine semantisch fragwürdigen Normalisierungen benötigt. Alle für das vorgestellte Konzept benötigten Schwellwerte und Wahrscheinlichkeiten werden entweder analytisch hergeleitet oder aus Trainingsdaten empirisch geschätzt. Obwohl eine direkte Kopplung der Bewertungsfunktion mit den erweiterten Constraint-Techniken realisiert wurde, bleibt das Bewertungsschema unabhängig vom verwendeten Suchverfahren. Es könnte auch ein anderes Verfahren wie z.B. Simulated Annealing oder der A\*-Algorithmus verwendet werden.

## 7.2 Beiträge der Dissertation

Nach dem Rückblick und der Bewertung, bei der die entwickelten Konzepte im Hinblick auf die formulierten Ziele dieser Arbeit betrachtet wurden, sollen an dieser Stelle die Beiträge der Dissertation zur wissenschaftlichen Forschung herausgearbeitet werden. Dabei geht es hauptsächlich um die konzeptionellen sowie methodischen Entwicklungen, und ihren Nutzen für bzw. ihre Übertragbarkeit auf andere Anwendungs- und Forschungsgebiete. Insgesamt lassen sich die Beiträge im wesentlichen den drei Bereichen (1) automatische Gebäudeerkennung, (2) Constraint Satisfaction und überbeschränkte Constraint-Systeme und (3) Relational Matching zuordnen. Sie werden im folgenden nach diesen Bereichen gegliedert aufgezählt.

### Beiträge für den Bereich Gebäudeerkennung

Zur Suche der Gebäude im Bild werden für die dreidimensionalen Gebäudehypothesen Ansichtshierarchien generiert, die eine besondere Form der Aspekthierarchien darstellen. Da die Gebäude-

hypothesen freie Form- und Lageparameter besitzen können, weisen die generierten Gebäudeansichten ebenfalls verschiedene Freiheitsgrade bzgl. ihrer geometrischen Ausprägung in bezug auf die Ausdehnung, Position und Orientierung im Bild auf. In dieser Arbeit wurde eine geeignete relationale Repräsentation für Gebäudeansichten vorgestellt, die invariant gegenüber dieser Freiheitsgrade ist. Zu diesem Zweck wurden verschiedene geometrische und topologische Relationen bzw. Constraints definiert, wobei für jeden Constraint im einzelnen die Invarianz gegenüber der variablen geometrischen Ausprägung einer Ansicht gezeigt wurde. Diese Invarianz ermöglicht die Anwendung relationaler Zuordnungsverfahren zur Suche der entsprechenden Gebäudeansichten im Bild. Es wurde anhand von praktischen Beispielen mit realen Luftbilddaten demonstriert, daß die ausgewählten Constraints eine ausreichende Trennschärfe besitzen, um relational repräsentierte Gebäudeansichten im Bild identifizieren, d.h. vom Bildhintergrund diskriminieren zu können.

Die Verwendung von Standard-Constraint-Techniken zur Objekterkennung ist bislang an den (unvermeidbaren) Störungen in den Bilddaten gescheitert. Constraints können durch Rauschen und andere Störungen verletzt werden, was zu einem Fehlschlagen der Suche führt. Diese Arbeit stellt ein neues Konzept vor, bei dem die Effekte der Unbeobachtbarkeit und der Unschärfe explizit berücksichtigt und differenziert bewertet werden. Die Erweiterung ermöglicht die Anwendung klassischer Constraint-Verfahren zur effizienten Suche von Gebäudeansichten in Luftbildern. Darüberhinaus erlaubt sie die Kopplung mit einer informationstheoretisch basierten Bewertungsfunktion, die die Bestimmung der wahrscheinlichsten Zuordnung gewährleistet.

Die Auswahl der wahrscheinlichsten Gebäudehypothese erfolgt mittels des Prinzips der minimalen Beschreibungslänge (MDL-Prinzip). Die Anwendung dieses Verfahrens zur Klassifikation relationaler Gebäuderepräsentationen ist neu und wird in dieser Arbeit erstmalig vorgestellt. Das MDL-Prinzip erfordert die optimale Codierung der Modelle. In dieser Arbeit wurde ein Schema entwickelt, das eine kompakte Codierung der relationalen Repräsentation von Gebäudeansichten realisiert, welche die verschiedenen Abhängigkeiten der einzelnen geometrischen und topologischen Relationen voneinander berücksichtigt. Die durchgeführten praktischen Tests an synthetischen und realen Daten haben ergeben, daß die entwickelte Codierung hinreichend kompakt ist, um selbst sehr ähnliche Gebäude wie z.B. ein Pultdach- und ein Flachdachhaus differenzieren zu können. Durch die Entwicklung des Codierschemas ist es nun möglich, auch unterschiedlich komplexe Gebäudeansichten ausschließlich auf der Basis ihrer relationalen Repräsentationen miteinander zu vergleichen, und probabilistisch fundiert zu klassifizieren.

Die Fragmentierung von Bildmerkmalen stellt ein schwerwiegendes Problem bei der Verwendung von relationalen Zuordnungsverfahren zur Objekterkennung dar. Es wurde eine Lösung vorgestellt, die eine modellbasierte Gruppierung fragmentierter Linienmerkmale während der Suche durchführt, wobei die Komplexität des Suchverfahrens nur unwesentlich erhöht wird. Durch die Gruppierung wird neben der Erhöhung der Genauigkeit bei der 3D-Rekonstruktion insbesondere erreicht, daß nur tatsächlich verschiedene Gebäude in den Bildern als einzelne Ergebnisse ausgegeben werden.

### **Beiträge für den Bereich Constraint Satisfaction und Over-Constrained Systems**

Die Verwendung von Techniken zur Lösung überbeschränkter Constraint-Probleme (Over-Constrained Systems, OCS) für die Objekterkennung ist neu. Bestehende OCS-Verfahren wurden im Hinblick auf die spezifische — mit bisherigen Ansätzen nicht repräsentierbare — Integration der Phänomene der Unbeobachtbarkeit und Unschärfe bzw. Unsicherheit konzeptionell erweitert. Es wurden Inferenzregeln für die Umsetzung des Konzepts angegeben, die auf Standard-Konsistenztechniken des Constraint Logic Programming aufsetzen.

Ferner wurde gezeigt, daß dieses neue Konzept eine natürliche Verbindung zwischen der Relaxation von Constraints, wie sie z.B. bei MaxCSP und HCLP vorgeschlagen wird, und der Eliminierung von Variablen, wie es Dynamic CSP vorsieht, darstellt. Die Modellierung läßt sich, wie in Abschnitt 4.3.7 demonstriert wurde, auch für andere Anwendungen wie z.B. Konfigurationsprobleme

einsetzen, bei denen abgewägt werden muß, ob es besser ist, einzelne Constraints zu verletzen oder bestimmte Variablen (die Objekte repräsentieren) wegzulassen.

Die Auswertung der geometrischen Constraints erfordert z.T. verhältnismäßig aufwendige Berechnungen. Um wiederholte Berechnungen bei Konsistenzüberprüfungen zu vermeiden, wurde die Implementierung des Standard-Constraint-Interpreters um einen Cache-Mechanismus erweitert, der vollständige Wertebereichsreduktionen zwischenspeichert, die im späteren Verlauf der Suche wiederverwendet werden können. Der Cache-Mechanismus sorgt insbesondere in Verbindung mit der erweiterten Constraint-Modellierung für eine deutliche Performanzsteigerung bei der Suche.

### Beiträge für den Bereich Relational Matching

In dieser Arbeit wird gezeigt, wie sich die Probleme der exakten und insbesondere inexakten relationalen Zuordnung (Relational Matching) mittels Constraint Logic Programming auf effiziente Weise lösen lassen. Sofern „gutartige“ Daten vorliegen, die die Erfüllung aller Constraints erlauben, ist die Objekterkennung trotz großer Suchräume bereits mit Standard-CSP-Heuristiken sehr effizient durchführbar. Es wurde gezeigt, wie eine informationstheoretische Bewertungsfunktion für relationale Zuordnungen zur Berechnung von Constraint-Gewichten eingesetzt werden kann, so daß die beste Lösung des überbeschränkten Constraint-Problems die wahrscheinlichste darstellt.

Durch die in dieser Arbeit entwickelte kompakte Codierung relationaler Gebäudestrukturen wurde die Verwendung des MDL-Prinzips zur Klassifikation unterschiedlich komplexer relationaler Gebäudestrukturen ermöglicht. Es wurde gezeigt, daß die Einbeziehung der Codierungslänge verschiedener Modelle in die informationstheoretische Bewertungsfunktion eine Maximum-A-Posteriori-Schätzung realisiert, die das wahrscheinlichste Modell bestimmt.

Diese Vorgehensweise ist übertragbar auf andere Problembereiche der Objekterkennung mit komplexen Modellen, wie z.B. in der Robotik. Ferner sind Anwendungen in anderen Bereichen denkbar, die relationale Modelle zuordnen und anschließend klassifizieren müssen, wie z.B. in der Chemie oder der Molekularbiologie. Dabei muß jedoch beachtet werden, daß das Codierschema von der Semantik der verwendeten Relationen abhängt, und diese wiederum von der konkreten Modellierung der zu erkennenden Objekte. Dadurch ist die Bestimmung eines (sub-)optimalen Codierschemas jeweils stark von der Anwendungsmodellierung abhängig. Das bedeutet, daß für andere Anwendungsgebiete u.U. ein völlig neues Codierschema angegeben werden muß. Hierbei können aber die in Abschnitt 6.2 vorgestellten Regeln bzgl. der Berücksichtigung von Relationeneigenschaften wie Symmetrie, Reflexivität und Transitivität sowie die Diskussion der verschiedenen Graphrepräsentationen wie Adjazenzmatrix, Adjazenzliste und Permutationsreihe hilfreich sein.

## 7.3 Ausblick

Bislang erfolgt die Identifikation von Gebäudehypothesen in einem Schritt, d.h. es werden vollständige Gebäudeansichten mit dem Bild verglichen. Dies führt bei komplexen Gebäuden zu sehr großen Suchräumen. Die Komplexität ließe sich reduzieren, indem zunächst nach einzelnen Gebäudeteilen gesucht wird, und diese dann zu einem Gebäude aggregiert werden. Die vorgestellte Bewertungsfunktion würde sich nach wie vor für die abschließende Gesamtbewertung aller Hypothesen und der Identifikation der wahrscheinlichsten Hypothese eignen. Bei einer komponentenweisen Erkennung wäre aber zu beachten, daß die Gebäudeteile nicht zu klein werden dürfen, da ansonsten die daraus hergeleiteten Constraint-Repräsentationen u.U. zu unspezifisch sind, und somit viele Fehlzusammenhänge auftreten können.

Zur Erhöhung der Spezifität kann die Modellierung um weitere Constraints, wie z.B. Flächensymmetrien, ergänzt werden. Ferner wurden bislang keine Flächenattribute verwendet. Sobald ein Beleuchtungsmodell integriert würde, könnte das Verhältnis der durchschnittlichen Grauwerte benachbarter Flächen als neuer Constraint verwendet werden. Damit Constraints auf den Form- und

Flächeneigenschaften verwendet werden können, müßte zunächst die Segmentierung soweit verbessert werden, daß die Flächen nicht mehr so stark fragmentieren oder „auslaufen“. Dies könnte z.B. durch Berücksichtigung von Farbinformationen erfolgen.

Eine andere Möglichkeit, der Komplexität zu begegnen, besteht in der Verwendung stärkerer Heuristiken. Die bislang verwendete First-Fail-Heuristik berücksichtigt keine Information über die Constraint-Gewichte. Hier wäre zu untersuchen, wie Best-First-Heuristiken, wie z.B. der A\*-Algorithmus, mit der dem CLP-Konzept zugrundeliegenden Tiefensuche integriert werden können.

Ein Problem, das in Verbindung mit der Zerlegung von Gebäudeansichten in einzelne Merkmale auftritt, ist, daß die Randflächendarstellung der Gebäudeansichten Merkmale beinhalten kann, die im Bild so klein sind, daß sie mit Sicherheit nicht beobachtet werden können. Die Gründe dafür sind einerseits ungünstige Perspektiven, so werden z.B. Modellmerkmale für die Seitenwände eines Gebäudes generiert, obwohl das Gebäude fast von vorne betrachtet wird, oder es werden vertikale Linienmerkmale prädiert, obwohl das Haus fast genau am Nadirpunkt liegt. Andererseits können manche Bildmerkmale aufgrund einer zu kleinen Bildauflösung nicht erkannt werden, wie z.B. die Linienmerkmale an den Walmstücken von Krüppelwalmdachhäusern. Dieses Problem ließe sich weitestgehend reduzieren, indem entweder die Bildauflösung bei der Generierung der Ansichten berücksichtigt wird, wie es beispielsweise Eggert et. al. [EBD<sup>+</sup>93] vorschlagen, oder bei der Transformation von Gebäudeansichten zu Constraint-Mengen nur die mit größerer Wahrscheinlichkeit beobachtbaren Merkmale berücksichtigt werden.

Die für die Bewertungsfunktion benötigten bedingten Wahrscheinlichkeiten werden anhand von Trainingsdaten empirisch geschätzt. Hier wäre noch zu untersuchen, welchen Einfluß die Größe der Stichprobe auf die Qualität der Zuordnungen besitzt, und wie gut sich die berechneten Wahrscheinlichkeiten auf andere Testgebiete übertragen lassen. Eine interessante Frage ist dabei, wie stark die Wahrscheinlichkeiten von der Bodenauflösung der Trainingsdaten abhängen, und inwieweit sie auch für die Gebäudeerkennung in Luftbildern mit anderer Bodenauflösung eingesetzt werden können.



# Literaturverzeichnis

- [AHU87] AHO, A., J. E. HOPCROFT und J. D. ULLMAN: *Data Structures and Algorithms*. Addison-Wesley, London, 1987.
- [Bau97] BAUER, W.: *Anforderungen an 3D-Stadtmodelle am Beispiel Stuttgart 21*. In: FRITSCH, D. und D. HOBBIE (Herausgeber): *Proceedings of the Photogrammetric Week '97, Stuttgart*, Seiten 165–173. Wichmann, Heidelberg, 1997.
- [BB98] BILL, R. und C. BOYTSCHIEFF: *Digitale Straßenzüge für die Stadtplaner von morgen*. forschung – Mitteilungen der DFG, 1-2:22–25, 1998.
- [BBS97] BRINGMANN, O., M. F. BUCHROITHNER und H.-H. SEIFERT: *Unschärfe hierarchische Interpretation komplexer analoger Pläne für semantisch-adäquate Visualisierungen*. In: *Nachrichten aus dem Karten- und Vermessungswesen*, Nummer 117 in Reihe I, Seiten 13–29. Bundesamt für Kartographie und Geodäsie, Frankfurt am Main, 1997.
- [BHFS96] BIGNONE, F., O. HENRICSSON, P. FUA und M. STRICKER: *Automatic Extraction of Generic House Roofs from High Resolution Aerial Images*. In: *Computer Vision — Proceedings of ECCV '96*, Nummer 1064 in *Lecture Notes in Computer Science*, Seiten 85–96. Springer-Verlag, Berlin, 1996.
- [BK88] BOYER, K. L. und A. C. KAK: *Structural Stereopsis for 3D-Vision*. IEEE Transactions on Pattern Analysis and Machine Intelligence, 10(2):144–166, März 1988.
- [BKL<sup>+</sup>95a] BRAUN, C., T. H. KOLBE, F. LANG, W. SCHICKLER, V. STEINHAGE, A. B. CREMERS, W. FÖRSTNER und L. PLÜMER: *Modelle für die photogrammetrische Gebäude-rekonstruktion*. Technischer Bericht IAI-TR-95-1, Institut für Informatik III, Universität Bonn, 1995.
- [BKL<sup>+</sup>95b] BRAUN, C., T. H. KOLBE, F. LANG, W. SCHICKLER, V. STEINHAGE, A. B. CREMERS, W. FÖRSTNER und L. PLÜMER: *Models for Photogrammetric Building Reconstruction*. Computer & Graphics, 19(1):109–118, 1995.
- [BMS95] BALTSAVIAS, E., S. MASON und D. STALLMANN: *Use of DTMs/DSMs and Orthoimages to Support Building Extraction*. In: GRÜN, A. et al. [GKA95], Seiten 199–210.
- [Bos97] BOSCH, P.: *Verwendung logischer Programmierung und hierarchischer Constraints bei der Erkennung von Gebäuden in Luftbildern*. Diplomarbeit, Institut für Informatik III, Universität Bonn, 1997.
- [BP95] BEIERLE, C. und L. PLÜMER (Herausgeber): *Logic Programming: Formal Methods and Practical Applications*. North-Holland, Amsterdam, 1995.
- [Bro83] BROOKS, R. A.: *Model-Based Three-Dimensional Interpretations of Two-Dimensional Images*. IEEE Transactions on Pattern Analysis and Machine Intelligence, 5(2):140–150, März 1983.

- [BSMM97] BRONSTEIN, I. N., K. A. SEMENDJAJEW, G. MUSIOL und H. MÜHLIG: *Taschenbuch der Mathematik*. Verlag Harri Deutsch, Frankfurt am Main, 3. Auflage, 1997.
- [BWF95] BRUNN, A., U. WEIDNER und W. FÖRSTNER: *Model-based 2D-Shape Recovery*. In: AL., G. SAGERER ET. (Herausgeber): *Mustererkennung 1995, Proceedings of the 17th DAGM Symposium in Bielefeld*, Seiten 260–268. Springer-Verlag, Berlin, 1995.
- [CH94a] CLARK, J. und D. A. HOLTON: *Graphentheorie: Grundlagen und Anwendungen*. Spektrum Akademischer Verlag, Heidelberg, Berlin, 1994.
- [CH94b] COOK, D. J. und L. B. HOLDER: *Substructure Discovery Using Minimum Description Length and Background Knowledge*. Journal of Artificial Intelligence Research, 1:231–255, 1994.
- [CHRS95] COLLINS, R. T., A. R. HANSON, E. M. RISEMAN und H. SCHULTZ: *Automatic Extraction of Buildings and Terrain from Aerial Images*. In: GRÜN, A. et al. [GKA95], Seiten 169–178.
- [CM84] CLOCKSIN, W. F. und C. S. MELLISH: *Programming in Prolog*. Springer Verlag, Berlin, Heidelberg, New York, Tokyo, 2. Auflage, 1984.
- [DH73] DUDA, R. O. und P. E. HART: *Pattern Classification and Scene Analysis*. Wiley & Sons, New York, 1973.
- [DH97] DERUYVER, A. und Y. HODÉ: *Constraint satisfaction problem with bilevel constraint: Application to interpretation of over segmented images*. Artificial Intelligence, 93(1-2):321–335, 1997.
- [DP88] DECHTER, R. und J. PEARL: *Network-Based Heuristics for Constraint Satisfaction Problems*. Artificial Intelligence, 34:370–425, 1988.
- [DPR92a] DICKINSON, S. J., A. P. PENTLAND und A. ROSENFELD: *3-D Shape Recovery Using Distributed Aspect Matching*. IEEE Transactions on Pattern Analysis and Machine Intelligence, 14(2):174–198, Februar 1992.
- [DPR92b] DICKINSON, S. J., A. P. PENTLAND und A. ROSENFELD: *From Volumes to Views: An Approach to 3-D Object Recognition*. CVGIP: Image Understanding, 55(2):130–154, März 1992.
- [DvH91] DEVILLE, Y. und P. VAN HENTENRYCK: *An Efficient Arc Consistency Algorithm for a Class of CSP Problems*. In: REITER, R. und J. MYOPOULOS (Herausgeber): *Proc. of the 12. Int. Joint Conf. on Artificial Intelligence IJCAI'91 in Sydney, Australia*. Morgan Kaufmann, August 1991.
- [DvHS<sup>+</sup>88] DINCIBAS, M., P. VAN HENTENRYCK, H. SIMONIS, A. AGGOUN, T. GRAF und F. BERTHIER: *The Constraint Logic Programming Language CHIP*. In: ICOT (Herausgeber): *Proceedings of the International Conference on Fifth Generation Computer Systems*, Seiten 693–702, 1988.
- [EBD<sup>+</sup>93] EGGERT, D. W., K. W. BOWYER, C. R. DYER, H. I. CHRISTENSEN und D. B. GOLDFOF: *The Scale Space Aspect Graph*. IEEE Transactions on Pattern Analysis and Machine Intelligence, 15(11):1114–1129, November 1993.
- [EC93] ENGESSER, H. und V. CLAUS (Herausgeber): *Duden Informatik*. Dudenverlag, Mannheim, 2. Auflage, 1993.
- [EC97] ENGLERT, R. und A. B. CREMERS: *Improving reconstruction of man-made objects from sensor images by machine learning*. In: JR. D. MCKEOWN, C. MCGLONE und O. JAMET (Herausgeber): *Proceedings of the SPIE's 11th Int. Symposium on Aerosense, Orlando, Florida, 1997*.

- [EG96] ENGLERT, R. und E. GÜLCH: *A One-Eye Stereo System for the Acquisition of Complex 3D-Building Structures*. GEO-INFORMATION-SYSTEMS, Journal for Spatial Information and Decision Making, 9(4):16–21, August 1996.
- [Eng97a] ENGLERT, R.: *Reducing Complexity of 3D Object Reconstruction Due to Symmetry of Model Knowledge*. In: *Proceedings of the Joint ISPRS Commission III/IV Workshop 1997 in Stuttgart, Germany*, Band 32 der Reihe *International Archives of Photogrammetry and Remote Sensing*, 1997.
- [Eng97b] ENGLERT, R.: *Systematic Acquisition of Generic 3D Building Model Knowledge*. In: FÖRSTNER, W. und L. PLÜMER [FP97], Seiten 181–195.
- [Eng98] ENGLERT, R.: *Acquisition of Complex Model Knowledge by Domain Theory-Controlled Generalization*. In: *Proceedings of ISPRS Commission III Symposium, Columbus, Ohio*, 1998.
- [Epp95] EPPSTEIN, D.: *Subgraph isomorphism in planar graphs and related problems*. In: *Proc. of the 6th Symposium on Discrete Algorithms*, Seiten 632–640. Assoc. Comput. Mach. and Soc. Industrial & Applied Mathematics, 1995.
- [FA97a] FRÜHWIRTH, T. und S. ABDENNADHER: *Anwendungen Constraintbasierter Programmierung*. In: *Tagungsband Jahrestagung 1997 der Gesellschaft für Informatik in Aachen*, September 1997.
- [FA97b] FRÜHWIRTH, T. und S. ABDENNADHER: *Constraint-Programmierung*. Springer-Verlag, Berlin, 1997.
- [Fed86] FEDER, M.: *Maximum Entropy as a Special Case of the Minimum Description Length Criterion*. IEEE Transactions on Information Theory, 32(6):847–849, November 1986.
- [FF95] FUCHS, C. und W. FÖRSTNER: *Polymorphic Grouping for Image Segmentation*. In: *Proceedings of the 5th ICCV in Boston*, Seiten 175–182. IEEE Computer Society Press, 1995.
- [FH87] FUA, P. und A. J. HANSON: *Resegmentation Using Generic Shape: Locating General Cultural Objects*. Pattern Recognition Letters, 5:243–252, 1987.
- [FH89a] FUA, P. und A. J. HANSON: *Objective Functions for Feature Discrimination: Applications to Semiautomated and Automated Feature Extraction*. In: *Proceedings of the Image Understanding Workshop 1989*, 1989.
- [FH89b] FUA, P. und A. J. HANSON: *Objective Functions for Feature Discrimination: Theory*. In: *Proceedings of the Image Understanding Workshop 1989*, 1989.
- [FHK<sup>+</sup>92] FRÜHWIRTH, T., A. HEROLD, V. KÜCHENHOFF, T. LE PROVOST, P. LIM, E. MONFROY und M. WALLACE: *Constraint Logic Programming — An Informal Introduction*. In: AL, G. COMYN ET (Herausgeber): *Logic Programming in Action*, Nummer 636 in *Lecture Notes in Computer Science*. Springer, September 1992.
- [FHV<sup>+</sup>97] FRÉRE, D., M. HENDRICKX, J. VANDEKERCKHOVE, T. MOONS und L. VAN GOOL: *On the Reconstruction of Urban House Roofs from Aerial Images*. In: GRÜN, A. et al. [GBH97], Seiten 87–96.
- [FJ91] FLYNN, P. J. und A. K. JAIN: *CAD-Based Computer Vision: From CAD Models to Relational Graphs*. IEEE Transactions on Pattern Analysis and Machine Intelligence, 13(2):114–132, Februar 1991.
- [FKL97] FISCHER, A., T. H. KOLBE und F. LANG: *Integration of 2D and 3D Reasoning for Building Reconstruction Using a Generic Hierarchical Model*. In: FÖRSTNER, W. und L. PLÜMER [FP97], Seiten 159–180.

- [FKL<sup>+</sup>98] FISCHER, A., T. H. KOLBE, F. LANG, A. B. CREMERS, W. FÖRSTNER, L. PLÜMER und V. STEINHAGE: *Extracting Buildings from Aerial Images using Hierarchical Aggregation in 2D and 3D*. Computer Vision & Image Understanding, 72(2), 1998.
- [FLF94] FUCHS, C., F. LANG und W. FÖRSTNER: *On the Noise and Scale Behaviour of Relational Descriptions*. In: EBNER, HEIPKE und EDER (Herausgeber): *ISPRS*, Band 30, Seiten 256–267. SPIE, 1994.
- [För89] FÖRSTNER, W.: *Image Analysis Techniques for Digital Photogrammetry*. In: *Proceedings of the 42th Photogrammetric Week*, Band 13 der Reihe *Series of the Institute of Photogrammetry*, Seiten 205–221. University Stuttgart, September 1989.
- [För94] FÖRSTNER, W.: *A Framework for Low Level Feature Extraction*. In: EKLUNDH, J.-O. (Herausgeber): *Computer Vision, ECCV '94, Vol. II*, Nummer 801 in *Lecture Notes in Computer Science*, Seiten 383–394. Springer-Verlag, 1994.
- [FP97] FÖRSTNER, W. und L. PLÜMER (Herausgeber): *Semantic Modeling for the Acquisition of Topographic Information from Images and Maps*. Birkhäuser Verlag, Basel, Switzerland, 1997.
- [Fre89] FREUDER, E. C.: *Partial Constraint Satisfaction*. In: SRIDHARAN, N. (Herausgeber): *Proc. of the 11. Int. Joint Conf. on Artificial Intelligence IJCAI'89 in Detroit, MI, USA*. Morgan Kaufmann, August 1989.
- [FS97a] FISCHER, A. und V. STEINHAGE: *Ein modellbasiertes Konzept zur städteplanerischen Kartierung durch digitale Bildanalyse*. In: *Int. Kolloquium über Anwendungen der Informatik und der Mathematik in Architektur und Bauwesen IKM'97*, Weimar, 1997.
- [FS97b] FISCHER, A. und V. STEINHAGE: *Solid Modeling for Building Extraction from Aerial Images*. In: SKALA, V. (Herausgeber): *5th Int. Conf. in Central Europe on Computer Graphics and Visualization 97, WSCG'97*, Seiten 114–123, 1997.
- [FSB89] FOX, M. S., N. SADEH und C. BAYKAN: *Constrained Heuristic Search*. In: SRIDHARAN, N. (Herausgeber): *Proc. of the 11. Int. Joint Conf. on Artificial Intelligence IJCAI'89 in Detroit, MI, USA*. Morgan Kaufmann, August 1989.
- [Fuc97] FUCHS, C.: *Extraktion polymorpher Bildstrukturen und ihre topologische und geometrische Gruppierung*. Doktorarbeit, Institut für Photogrammetrie, Universität Bonn, 1997.
- [FvDFH90] FOLEY, J. D., A. VAN DAM, S. K. FEINER und J. F. HUGHES: *Computer Graphics: Principles and Practice*. Addison-Wesley, 1990.
- [FW92] FREUDER, E. C. und R. WALLACE: *Partial Constraint Satisfaction*. Artificial Intelligence, 58:21–70, 1992.
- [FW96] FREUDER, E. C. und R. J. WALLACE: *Partial Constraint Satisfaction*. In: JAMPPEL, M. et al. [JFM96], Seiten 63–110.
- [GBH97] GRÜN, A., E. P. BALTSAVIAS und O. HENRICSSON (Herausgeber): *Automatic Extraction of Man-Made Objects from Aerial and Space Images (II)*. Birkhäuser, Basel, 1997.
- [GCS91] GIGUS, Z., J. CANNY und R. SEIDEL: *Efficiently Computing and Representing the Aspect Graphs of Polyhedral Objects*. IEEE Transactions on Pattern Analysis and Machine Intelligence, 13(6):542–551, 1991.
- [GKA95] GRÜN, A., O. KÜBLER und P. AGOURIS (Herausgeber): *Automatic Extraction of Man-Made Objects from Aerial and Space Images*. Birkhäuser, Basel, 1995.

- [GMLR98] GÜLCH, E., H. MÜLLER, T. LÄBE und L. RAGIA: *On the performance of semi-automatic building extraction*. In: *Proceedings of ISPRS Commission III Symposium, Columbus, Ohio, 1998*.
- [Grü97] GRÜN, A.: *Automation in Building Reconstruction*. In: FRITSCH, D. und D. HOBBIE (Herausgeber): *Proceedings of the Photogrammetric Week '97, Stuttgart*, Seiten 175–186. Wichmann, Heidelberg, 1997.
- [Gül92] GÜLCH, E.: *A Knowledge Based Approach to Reconstruct Buildings in Digital Aerial Imagery*. In: FRITZ, L. W. und J. R. LUCAS (Herausgeber): *International Archives of Photogrammetry and Remote Sensing*, Band 29, B2, Seiten 410–417, 1992.
- [Gül97] GÜLCH, E.: *Application of Semi-Automatic Building Acquisition*. In: GRÜN, A. et al. [GBH97], Seiten 129–138.
- [Gül98] GÜLCH, E.: *Halbautomatische 3D-Gebäudeerfassung*. In: *Nachrichten aus dem Karten- und Vermessungswesen*, Nummer 118 in *Reihe I*, Seiten 39–54. Bundesamt für Kartographie und Geodäsie, Frankfurt am Main, 1998.
- [GW84] GEORGEFF, M. P. und C. S. WALLACE: *A General Selection Criterion for Inductive Inference*. In: *Proceedings of ECAI'84 — Advances in Artificial Intelligence*, Seiten 473–482. Elsevier Science, 1984.
- [HB97] HENRICSSON, O. und E. P. BALTSAVIAS: *3-D Building Reconstruction with ARUBA: A Qualitative and Quantitative Evaluation*. In: GRÜN, A. et al. [GBH97], Seiten 139–148.
- [HE80] HARALICK, R. M. und G. L. ELLIOTT: *Increasing Tree Search Efficiency for Constraint Satisfaction Problems*. *Artificial Intelligence*, 14:263–313, 1980.
- [Hen95] HENRICSSON, O.: *Inferring Homogeneous Regions from Rich Image Attributes*. In: GRÜN, A. et al. [GKA95], Seiten 13–22.
- [Hen96] HENRICSSON, O.: *Analysis of Image Structures using Color Attributes and Similarity Relations*. Doktorarbeit, Institut für Geodäsie und Photogrammetrie, ETH Zürich, 1996.
- [HK87] HERMAN, M. und T. KANADE: *The 3D MOSAIC Scene Understanding System: Incremental Reconstruction of 3D Scenes from Complex Images*. In: FISCHLER, M. A. und O. FIRSCHEIN (Herausgeber): *Readings in Computer Vision: Issues, Problems, Principles, and Paradigms*, Seiten 471–482. Morgan Kaufmann Publishers, Los Altos, California, 1987.
- [HN88] HUERTAS, A. und R. NEVATIA: *Detecting Buildings in Aerial Images*. *Computer Vision, Graphics, and Image Processing CVGIP*, 41:131–152, 1988.
- [HN98] HUERTAS, A. und R. NEVATIA: *Detecting Changes in Aerial Views of Man-Made Structures*. In: *Proceedings of the 6th Int. Conference on Computer Vision, ICCV'98 in Bombay, India*. IEEE Computer Society Press Los Alamitos, CA, USA, 1998.
- [HS79] HARALICK, R. M. und L. G. SHAPIRO: *The Consistent Labeling Problem: Part I*. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1:173–184, 1979.
- [HS93a] HARALICK, R. M. und L. G. SHAPIRO: *Computer and Robot Vision*, Band II. Addison-Wesley Publishing Company, 1993.
- [HS93b] HARALICK, R. M. und L. G. SHAPIRO: *Computer and Robot Vision*, Band I. Addison-Wesley Publishing Company, 1993.

- [HU90] HOPCROFT, J. E. und J. D. ULLMAN: *Einführung in die Automatentheorie, Formale Sprachen und Komplexitätstheorie*. Addison-Wesley, Bonn, 1990.
- [HZ83] HUMMEL, R. A. und S. W. ZUCKER: *On the Foundations of Relaxation Labeling Processes*. IEEE Transactions on Pattern Analysis and Machine Intelligence, 5(3):267–287, Mai 1983.
- [Ing71] INGELS, F. M.: *Information and Coding Theory*. Intext Educational Publishers, San Francisco, Toronto, London, 1971.
- [Jam96] JAMPPEL, M. B.: *Over-Constrained Systems in CLP and CSP*. Doktorarbeit, Dep. of Computer Science, City University London, UK, September 1996.
- [JFM96] JAMPPEL, M., E. FREUDER und M. MAHER (Herausgeber): *Over-Constrained Systems*. Nummer 1106 in *Lecture Notes in Computer Science*. Springer-Verlag, Berlin, Germany, 1996.
- [JHR97] JAYNES, C., A. HANSON und E. RISEMAN: *Model-Based Surface Recovery of Buildings in Optical and Range Images*. In: FÖRSTNER, W. und L. PLÜMER [FP97], Seiten 211–227.
- [JM94] JAFFAR, J. und M. J. MAHER: *Constraint Logic Programming: A Survey*. Journal of Logic Programming, 19/20:503–581, 1994.
- [KGV83] KIRKPATRICK, S., C. D. GELATT und M. P. VECCHI: *Optimization by simulated annealing*. Science, 220(4598):671–680, 1983.
- [Koc97] KOCH, K. R.: *Parameterschätzung und Hypothesentests*. Dümmler Verlag, Bonn, 3. Auflage, 1997.
- [Kol98] KOLBE, T. H.: *Constraints for Object Recognition in Aerial Images — Handling of Unobserved Features*. In: MAHER, M. und J.-F. PUGET (Herausgeber): *Principles and Practice of Constraint Programming – Proceedings of CP’98*, Nummer 1520 in *Lecture Notes in Computer Science*. Springer-Verlag, Berlin, 1998.
- [KPC96] KOLBE, T. H., L. PLÜMER und A. B. CREMERS: *Using Constraints for the Identification of Buildings in Aerial Images*. In: *Proceedings of the 2. Int. Conf. on Practical Applications of Constraint Technology PACT’96 in London*, Seiten 143–154. The Practical Application Company Ltd., 1996.
- [KPC00] KOLBE, T. H., L. PLÜMER und A. B. CREMERS: *Identification of Buildings in Aerial Images Using Constraint Relaxation and Variable Elimination*. IEEE Intelligent Systems, 15(1):33–39, 2000.
- [Kra94] KRAUS, K.: *Photogrammetrie*, Band 1. Dümmler Verlag, Bonn, 5. Auflage, 1994.
- [Kul97] KULSCHEWSKI, K.: *Building Recognition with Bayesian Networks*. In: FÖRSTNER, W. und L. PLÜMER [FP97], Seiten 196–210.
- [Kum92] KUMAR, V.: *Algorithms for Constraint Satisfaction Problems: A Survey*. AI Magazine, 13(1):32–44, 1992.
- [KvD76] KOENDERINK, J. J. und A. J. VAN DOORN: *Visual Perception of Rigidity of Solid Shape*. Journal of Math. Biology, 3:79–85, 1976.
- [KvD79] KOENDERINK, J. J. und A. J. VAN DOORN: *The Internal Representation of Solid Shape with Respect to Vision*. Biological Cybernetics, 32:211–216, 1979.
- [KWT88] KASS, M., A. WITKIN und D. TERZOPOULOS: *Snakes: Active Contour Models*. International Journal of Computer Vision, 1(4):321–331, 1988.

- [Lan99] LANG, F.: *Geometrische und semantische Rekonstruktion von Gebäuden durch Ableitung von 3D-Gebäudeecken*. Dissertation. Reihe Informatik. Shaker Verlag, Aachen, 1999.
- [LF96a] LANG, F. und W. FÖRSTNER: *3D-City Modeling with a Digital One-Eye-Stereo System*. In: *International Archives of Photogrammetry and Remote Sensing*, Band 31, B4, Seiten 261–266, 1996.
- [LF96b] LANG, F. und W. FÖRSTNER: *Surface Reconstruction of Man-Made Objects using Polymorphic Mid-Level Features and Generic Scene Knowledge*. In: *Proceedings of the 18th ISPRS Congress, Wien*, 1996.
- [LG98] LÄBE, T. und E. GÜLCH: *Robust Techniques for Estimating Parameters of 3D Building Primitives*. In: *Proceedings of ISPRS Commission II Symposium, Cambridge, UK, July 1998*, 1998.
- [LH95] LAWSON, C. L. und R. J. HANSON: *Solving Least Squares Problems*. Classics in Applied Mathematics. SIAM, Philadelphia, 1995.
- [LHN95] LIN, C., A. HUERTAS und R. NEVATIA: *Detection of Buildings from Monocular Images*. In: GRÜN, A. et al. [GKA95], Seiten 125–134.
- [Llo87] LLOYD, J. W.: *Foundations of Logic Programming*. Springer-Verlag, Berlin, Heidelberg, New York, 2. Auflage, 1987.
- [LS93] LANG, F. und W. SCHICKLER: *Semiautomatische 3D-Gebäudeerfassung aus digitalen Bildern*. Zeitschrift für Photogrammetrie und Fernerkundung, 5:193–200, 1993.
- [Mac77] MACKWORTH, A. K.: *Consistency in Networks of Relations*. Artificial Intelligence, 8:99–118, 1977.
- [Mac92] MACKWORTH, A. K.: *The logic of constraint satisfaction*. Artificial Intelligence, 58:3–20, 1992.
- [Mac98] MACRONE, M.: *Heureka! Das archimedische Prinzip und 80 weitere Versuche, die Welt zu erklären*. Deutscher Taschenbuch Verlag, München, 1998.
- [May98] MAYER, H.: *Automatische Objektextraktion aus digitalen Luftbildern*. Nummer 494 in *Reihe C*. Deutsche Geodätische Kommission, 1998. Habilitationsschrift.
- [MBS94] MASON, S., E. BALTSAVIAS und D. STALLMANN: *High Precision Photometric Data Set for Building Reconstruction and Terrain Modeling*. Technischer Bericht Institute for Geodesy and Photogrammetry, ETH Zürich, Switzerland, 1994.
- [Mes89] MESEGUER, P.: *Constraint Satisfaction Problems: An Overview*. AICOM, 2(1):3–17, März 1989.
- [MF90] MITTAL, S. und B. FALKENHAINER: *Dynamic Constraint Satisfaction Problems*. In: *Proc. of AAAI 1990 in Boston*, Seiten 25–31, 1990.
- [MM88] MOHR, R. und G. MASINI: *Good Old Discrete Relaxation*. In: *Proceedings of EC-CAI'88 in München*, Seiten 651–656, 1988.
- [MN88] MOHAN, R. und R. NEVATIA: *Perceptual Grouping for the Detection and Description of Structures in Aerial Images*. In: *Proceedings of the Image Understanding Workshop 1988*, 1988.
- [Moh89] MOHAN, R.: *Perceptual Organization for Computer Vision*. Doktorarbeit, Institute for Robotics and Intelligent Systems, University of Southern California, 1989.

- [MS92] MEIER, M. und J. SCHIMPF: *An Architecture for Prolog Extensions*. In: *Proc. of the 3rd Int. Workshop on Extensions of Logic Programming, Bologna, 1992*.
- [MS94] MCGLONE, J. C. und J. A. SHUFELT: *Projective and Object Space Geometry for Monocular Building Extraction*. In: *Proceedings of the Conference on Computer Vision and Pattern Recognition*, Seiten 54–61, Los Alamitos, CA, USA, 1994. IEEE Computer Society Press.
- [Nil82] NILSSON, N.: *Principles of Artificial Intelligence*. Springer-Verlag, Berlin, 1982.
- [NLH97] NEVATIA, R., C. LIN und A. HUERTAS: *A System for Building Detection from Aerial Images*. In: GRÜN, A. et al. [GBH97], Seiten 77–86.
- [NP82] NEVATIA, R. und K. E. PRICE: *Locating Structures in Aerial Images*. IEEE Transactions on Pattern Analysis and Machine Intelligence, 4(5):476–484, September 1982.
- [Nud83] NUDEL, B.: *Consistent-Labeling Problems and their Algorithms: Expected-Complexities and Theory-Based Heuristics*. Artificial Intelligence, 21(2):135–178, 1983.
- [OAD<sup>+</sup>92] (ORGANIZER), K. BOWYER, N. AHUJA, C. DYER, O. FAUGERAS, K. IKEUCHI, R. JAIN, J. MUNDY und A. PENTLAND: *Workshop Panel Report: Why Aspect Graphs are not (yet) Practical for Computer Vision*. CVGIP: Image Understanding, 55(2):212–218, März 1992.
- [Pap97] PAPULA, L.: *Mathematik für Ingenieure und Naturwissenschaftler*, Band 3. Vieweg Verlag, 2. Auflage, 1997.
- [Pea88] PEARL, J.: *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann, San Mateo, California, 2. Auflage, 1988.
- [PM95] PRESTWICH, S. und S. MUDAMBI: *Improved Branch and Bound in Constraint Logic Programming*. Technischer Bericht ECRC-95-19, ECRC München, 1995.
- [PS85] PREPARATA, F. P. und M. I. SHAMOS: *Computational Geometry*. Springer-Verlag, Berlin, 1985.
- [PTVF92] PRESS, W. H., S. A. TEUKOLSKY, W. T. VETTERLING und B. P. FLANNERY: *Numerical Recipes in C*. Cambridge University Press, Cambridge, 2. Auflage, 1992.
- [QR89] QUINLAN, J. R. und R. L. RIVEST: *Inferring Decision Trees Using the Minimum Description Length Principle*. Information and Computation, 80:227–248, 1989.
- [Rad93] RADIG, B. (Herausgeber): *Verarbeiten und Verstehen von Bildern*. Handbuch der Informatik. Oldenbourg Verlag, München, 1993.
- [Rei91] REIMER, U.: *Einführung in die Wissensrepräsentation*. B. G. Teubner, Stuttgart, 1991.
- [Ris83] RISSANEN, J.: *A Universal Prior for Integers and Estimation by Minimum Description Length*. The Annals of Statistics, 11(2):416–431, 1983.
- [Ris87] RISSANEN, J.: *Minimum-Description-Length Principle*. Encyclopedia of Statistical Sciences, 5:523–527, 1987.
- [RN95] RUSSELL, S. J. und P. NORVIG: *Artificial Intelligence — A Modern Approach*. Prentice-Hall, Englewood Cliffs, New Jersey, 1995.
- [RS94] REINHARDT, F. und H. SOEDER: *dtv-Atlas zur Mathematik*, Band 1: Grundlagen Algebra und Geometrie. Deutscher Taschenbuch Verlag, München, 10. Auflage, 1994.

- [SB93] SCHUTTE, K. und G. BOERSEMA: *Hypothesizing a 3-D scene from a segmented aerial photograph*. In: *Proceedings of the 2nd Conf. on Optical 3D Measurement Techniques in Zurich*, Seiten 452–459, Karlsruhe, 1993. Wichmann Verlag.
- [SB97] SIEBE, E. und U. BÜNING: *Application of Digital Photogrammetric Products for Cellular Radio Network Planning*. In: FRITSCH, D. und D. HOBBIE (Herausgeber): *Proceedings of the Photogrammetric Week '97, Stuttgart*, Seiten 159–164. Wichmann, Heidelberg, 1997.
- [Sch96] SCHUTTE, K.: *Recognition of buildings from aerial images*. In: BERNSSEN, J. A. C., J. J. GERBRANDS, A. A. HOEVE, A. W. M. SMEULDERS, M. A. VIERGEVER und A. M. VOSSEPOEL (Herausgeber): *3rd Quinquennial Review 1991-1996 Dutch Society for Pattern Recognition and Image Processing*, Seiten 211–225. NVPBHV, Delft, 1996.
- [SFH92] SUETENS, P., P. FUA und A. J. HANSON: *Computational Strategies for Object Recognition*. ACM Computing Surveys, 24(1):5–61, März 1992.
- [SH85] SHAPIRO, L. G. und R. M. HARALICK: *A Metric for Comparing Relational Descriptions*. IEEE Transactions on Pattern Analysis and Machine Intelligence, 7(1):90–94, Januar 1985.
- [SH87] SHAPIRO, L. G. und R. M. HARALICK: *Relational Matching*. Applied Optics, 26(10):1845–1851, Mai 1987.
- [Sha48] SHANNON, C. E.: *A Mathematical Theory of Communication*. Bell System Journal, 27, 1948.
- [Shu96] SHUFELT, J. A.: *Exploiting Photogrammetric Methods for Building Extraction in Aerial Images*. In: *Proceedings of the ISPRS Symposium, Comm. III in Vienna 1996, Austria*, Band 31 der Reihe *International Archives of Photogrammetry and Remote Sensing*, 1996.
- [Spi93] SPIES, M.: *Unsicheres Wissen*. Spektrum Akademischer Verlag, Heidelberg, Oxford, 1993.
- [SS94] STERLING, L. und E. SHAPIRO: *The Art of Prolog*. MIT Press, Cambridge, Massachusetts, 2. Auflage, 1994.
- [SSF96] SESTER, M., W. SCHNEIDER und D. FRITSCH: *Results of the Test on Image Understanding of ISPRS Working Group III/3*. In: *Proceedings of the ISPRS'96 in Vienna, Austria*, Band 31 der Reihe *International Archives of Photogrammetry and Remote Sensing*, 1996.
- [SSH97a] SPREEUWERS, L., K. SCHUTTE und Z. HOUKES: *A model driven approach to extract buildings from multi-view aerial imagery*. In: GRÜN, A. et al. [GBH97], Seiten 109–118.
- [SSH97b] SPREEUWERS, L., K. SCHUTTE und Z. HOUKES: *Modelbased Recognition and Parameter Estimation of Buildings from Multi-View Aerial Imagery Using Multi-Segmentation*. In: *Proceedings of the Joint ISPRS Commission III/IV Workshop 1997 in Stuttgart, Germany*, Band 32 der Reihe *International Archives of Photogrammetry and Remote Sensing*, 1997.
- [Ste97] STEINHAGE, V.: *On the Integration of Object Modeling and Image Modeling in Automated Building Extraction from Aerial Images*. In: GRÜN, A. et al. [GBH97], Seiten 139–148.
- [Ste98] STEINHAGE, V.: *Zur automatischen Gebäuderekonstruktion aus Luftbildern*. Habilitationsschrift, Institut für Informatik, Universität Bonn, 1998.

- [Sto92] STOKES, J.: *Parsing Segmented Images*. In: *Proceedings of the XVII ISPRS Congress in Washington D. C.*, Band 29, 1992.
- [Tel97] TELEKI, L.: *Dreidimensionale Qualitative Gebäuderekonstruktion*. Doktorarbeit, Institut für Photogrammetrie, Universität Bonn, Oktober 1997.
- [vH87] HENTENRYCK, P. VAN: *A Theoretical Framework for Consistency Techniques in Logic Programming*. In: *Proceedings of IJCAI'87 Milan, Italy*, 1987.
- [vH89] HENTENRYCK, P. VAN: *Constraint Satisfaction in Logic Programming*. Logic Programming Series. MIT Press, Cambridge, MA, 1989.
- [vHD86] HENTENRYCK, P. VAN und M. DINCIBAS: *Domains in Logic Programming*. In: *AAAI'86: Proceedings National Conference on Artificial Intelligence in Philadelphia*, Seiten 759–765, 1986.
- [vHD87] HENTENRYCK, P. VAN und M. DINCIBAS: *Forward Checking in Logic Programming*. In: *ICLP'87: Proceedings 4th International Conference on Logic Programming in Melbourne*, Seiten 229–256, 1987.
- [vHSD92] HENTENRYCK, P. VAN, H. SIMONIS und M. DINCIBAS: *Constraint satisfaction using constraint logic programming*. *Artificial Intelligence*, 58:113–159, 1992.
- [Vos92] VOSSELMAN, G.: *Relational Matching*. Nummer 628 in *Lecture Notes in Computer Science*. Springer-Verlag, Berlin, Heidelberg, New York, Tokyo, 1992.
- [Wal75] WALTZ, D. L.: *Understanding Line Drawings of Scenes with Shadows*. In: WINSTON, P. H. (Herausgeber): *Psychology of Computer Vision*, Seiten 19–91. McGraw-Hill, New York, 1975.
- [Wal96] WALTER, V.: *Zuordnung von raumbezogenen Daten – am Beispiel der Datenmodelle ATKIS und GDF*. Doktorarbeit, Institut für Photogrammetrie, Universität Stuttgart, Oktober 1996.
- [WB89] WILSON, M. und A. BORNING: *Extending Hierarchical Constraint Logic Programming: Nonmonotonicity and Inter-Hierarchy Comparison*. In: *Proceedings of the 1989 North American Conference on Logic Programming in Cleveland, Ohio*, 1989.
- [Wei95] WEIDNER, U.: *MDL-basierte Formrekonstruktion zur Gebäudeextraktion*. In: *Ta- gungsband zur Wissenschaftlichen Jahrestagung 1995 der DGPF*, 1995.
- [Wei97a] WEIDNER, U.: *Digital Surface Models for Building Extraction*. In: GRÜN, A. et al. [GBH97], Seiten 193–202.
- [Wei97b] WEIDNER, U.: *Gebäudeerfassung aus Digitalen Oberflächenmodellen*. Nummer 474 in *Reihe C*. Deutsche Geodätische Kommission, 1997. Dissertation.
- [WF95] WEIDNER, U. und W. FÖRSTNER: *Towards Automatic Building Extraction from High Resolution Digital Elevation Models*. *ISPRS Journal*, 50(4):38–49, 1995.
- [WHK90] WALKER, E. L., M. HERMAN und T. KANADE: *A Framework for Representing and Reasoning about Three-Dimensional Objects for Vision*. In: CHEN, S. (Herausgeber): *Advances in Spatial Reasoning*, Band 2, Kapitel 6, Seiten 219–247. Ablex Publishing, Norwood, New Jersey, 1990.
- [Wil93] WILSON, M. A.: *Hierarchical Constraint Logic Programming*. Doktorarbeit, Dept. of Computer Science, University of Washington, Mai 1993.
- [WNS97] WALLACE, M., S. NOVELLO und J. SCHIMPF: *ECLiPSe: A Platform for Constraint Logic Programming*. Technischer Bericht IC-Parc, Imperial College London, UK, August 1997.

# Abbildungsverzeichnis

1.1	Bislang verwendete Gebäudemodelle zur automatischen Gebäudeextraktion. . . . .	3
2.1	Verschiedene Arten der 3D-Objekt- und 2D-Bildmodellierung. . . . .	11
2.2	Das Gesamtmodell, bestehend aus einer vierstufigen Bestandteilshierarchie. . . . .	19
2.3	Modellierung des Merkmalsaggregats <i>Ecke</i> . . . . .	20
2.4	Beispiele für Gebäudeteile: drei Endstücke und zwei Verbinder. . . . .	20
2.5	Parametrisierung der Gebäudeteile am Beispiel eines Satteldachterminals. . . . .	21
2.6	Beispiele für Gebäude — aggregiert aus einfachen Konnektoren und Terminalen. . . . .	21
2.7	Aspektgraph eines Walmdachhauses und Dekomposition eines Aspekts. . . . .	22
2.8	Ausschnitt aus einem Luftbild und Ergebnis der Merkmalsextraktion. . . . .	24
2.9	Bestimmung relevanter Bildbereiche anhand eines Digitalen Oberflächenmodells. . . . .	25
2.10	3D-Rekonstruktion von Eckpunkten durch Mehrbildanalyse. . . . .	26
2.11	Konstruktion einer Gebäudehypothese aus Gebäudeteilen. . . . .	27
2.12	Mögliche Gebäudehypothesen und ihre Freiheitsgrade für gegebene 3D-Ecken. . . . .	29
2.13	Beispiel für eine nichtlineare Parametrisierung eines Gebäudeeckpunktes. . . . .	32
3.1	Beispiel für einen Subgraph-Isomorphismus. . . . .	38
3.2	Merkmalsnachbarschaftsrelation und -Constraint. . . . .	45
3.3	Linienparallelitätsrelation und -Constraint. . . . .	46
3.4	Kollinearitätsrelation und -Constraint. . . . .	47
3.5	Same_Side_Lines-Relation und -Constraint. . . . .	48
3.6	Generierung von Same_Side_Lines-Constraints für konkave Flächen. . . . .	49
3.7	Bestimmung der Constraint-Schwellwerte anhand von Trainingsdaten. . . . .	51
3.8	Häufigkeitsverteilungen der Abweichungen für Parallelität und Kollinearität. . . . .	52
3.9	Häufigkeitsverteilung der Linienwinkeldifferenzen zwischen Modell und Bild. . . . .	52
3.10	Fünf zueinander parallele Linien einer Gebäudeansicht bilden einen $K_5$ -Graph. . . . .	53
3.11	CLP-Klausel zur Repräsentation der Vorderansicht eines Satteldachhauses. . . . .	56
3.12	Erfolgreiche Zuordnung von Bild- zu Modellmerkmalen an einem Beispiel. . . . .	57
3.13	Zwei von 972 möglichen Zuordnungen von Bild- zu Modellmerkmalen. . . . .	59
3.14	Bestimmung der Giebelpunktkoordinaten anhand zugeordneter Bildmerkmale. . . . .	63
3.15	Bestimmung zu gruppierender Linienfragmente. . . . .	64

4.1	Luftbildausschnitt und extrahierte Bildmerkmale mit typischen Störungen. . . . .	67
4.2	Beobachtbare Störungen in realen Bildern. . . . .	70
4.3	Methoden zur Behandlung überbeschränkter Constraint-Probleme. . . . .	72
4.4	Erweiterte CLP-Klausel zur Repräsentation eines Satteldachhauses. . . . .	80
4.5	Suche nach einem Gebäude in vier Bildern mit unterschiedlichen Perspektiven. . .	82
4.6	Beispiel für ein einfaches Konfigurationsproblem. . . . .	83
5.1	Beispiel für eine aufgrund fehlender Gewichtung fehlerhafte Zuordnung. . . . .	85
5.2	Beispiel für zwei verschiedene inexakte relationale Zuordnungen. . . . .	86
5.3	Abbildung des Zuordnungsproblems auf ein Kommunikationsproblem. . . . .	88
5.4	Ergebnis der gewichteten Zuordnung von Satteldachhypothese und Bildmerkmalen. .	97
6.1	Zuordnung dreier verschiedener Gebäudeansichten zu demselben Bildausschnitt. . .	100
6.2	Beispielgraph und seine Repräsentation als Adjazenzmatrix und -liste. . . . .	106
6.3	Codierungslängen eines Graphen für die verschiedenen Codierschemata. . . . .	108
6.4	Partitionierung der Adjazenzmatrix des Merkmalsnachbarschaftsgraphen. . . . .	110
6.5	Synthetisches Bild eines Gebäudes und Ergebnis der Merkmalsextraktion. . . . .	117
6.6	Identifikation des wahrscheinlichsten Modells für ein L-förmiges Gebäude. . . . .	122
6.7	Identifikation der wahrscheinlichsten Ansicht am Beispiel eines Walmdachhauses. .	123

# Tabellenverzeichnis

3.1	Komplexität des Gebäudemodells und extrahierte Bildmerkmale für ein Beispiel. . .	57
3.2	Meßergebnisse von sechs verschiedenen Suchvorgängen. . . . .	58
4.1	Auswirkungen der Fehlerarten auf die Erfüllbarkeit der einzelnen Constraints. . . .	69
4.2	Bestimmung der besten Zuordnung unter Berücksichtigung von Unbeobachtbarkeit und Unschärfe. . . . .	81
5.1	Beispiel für die Angabe der a priori und bedingten Wahrscheinlichkeiten. . . . .	91
5.2	Die aus Tabelle 5.1 berechnete gemeinsame Information. . . . .	91
5.3	Eigenschaften und Kardinalitäten der verwendeten Relationen. . . . .	92
5.4	Anhand von 28 Gebäudeansichten geschätzte bedingte Wahrscheinlichkeiten. . . .	94
6.1	Codierungslängen der Ansichten von vier unterschiedlich komplexen Haustypen. . .	115
6.2	Codierungslängen der fünf beobachtbaren Ansichten eines Satteldachhauses. . . .	116
6.3	Codierungslängen der Gebäudeansichten von sechs einfachen Haustypen. . . . .	116
6.4	Zuordnung und Bewertung der verschiedenen Ansichten eines Satteldachhauses. . .	118
6.5	Zuordnung und Bewertung von sechs einfachen Haustypen in gleicher Perspektive.	119
6.6	Klassifikationsergebnisse für unterschiedlich komplexe Gebäudeansichten. . . . .	120
6.7	Auswirkungen längerer Codierungen auf die Klassifikationsergebnisse. . . . .	121



# Index

- A priori Wahrscheinlichkeit
  - Bildrelation, 90
  - Modellrelation, 90, 99
- Adjazenzliste, 106, 111
- Adjazenzmatrix, 106
- Äußere Orientierung, 30
- Ansichtenhierarchie, 28
- Aspekt, 12, 21
- Aspektgraph, 12, 27, 43
- Aspekthierarchie, 12, 21, 27
- ATKIS, 61
- Ausgleichungsproblem
  - Lösungsverfahren, 33
  - nichtlineares, 33
- Backtracking, 39, 41, 42
- Baumtiefensuche
  - mit Backtracking, 39, 41
  - mit Forward-Checking, 39
  - mit Look-Ahead, 39
- Bayes-Netz, 68
- Bayessches Theorem, 99
- bedingte Information, 88
- Beobachtbarkeit, 9
- Berechnungszustand, 42, 76
- Beschreibungslänge, 101, 105
- Bestandteilshierarchie, 12
- Bildmerkmale, 12, 19, 23, 59, 69, 90
- Bildmodell, 12, 21
  - ideales, 60
- Bildstörungen, 68
- Bitstring, 107, 111
- Bottom-Up-Strategie, 13, 17
- Branch and Bound, 78
- Caching, 55, 58
- CAD-Modell, 43
- CHIP, 55
- Clique, 109, 112
- Codierung
  - Codewort, 102
  - Huffman-, *siehe* Huffman-Code
  - optimale, 102, 110
  - Permutationsreihe, 107
  - Präfixcode, *siehe* Präfixeigenschaft
- Codierungslänge, *siehe* Beschreibungslänge
  - Kollinearität, 112
  - Linienparallelität, 112
  - Merkmalsnachbarschaft, 110
  - mittlere, 102
  - Same\_Side\_Lines, 112
  - Verschiedenheit, 113
- Constraint, 36
  - Bilevel-, 60
  - geometrischer, 44
  - Kollinearitäts-, 47
  - komplementäre Cons.-Menge, 95, 96
  - Linienparallelitäts-, 45, 54
  - Linienwinkel-, 50
  - Merkmalsnachbarschafts-, 44, 71
  - Relaxation von, 71, 72, 74
  - Same\_Side\_Lines-, 48
  - Schwellwerte, 51
  - topologischer, 44
  - Verschiedenheits-, 50
- Constraint Logic Programming, 41, 54
  - Alldistinct-Constraint, 50, 79
  - Atmost-Constraint, 79
  - CLP(FD), 43, 54, 113
  - Constraint-Speicher, 42
  - First-Fail-Heuristik, 43
  - Folgerung, 42, 77
  - Forward-Checking, 43, 55, 58
  - HCLP, 72
  - Look-Ahead, 43, 55, 58
  - Widerspruch, 42, 77
- Constraint Solving, 36
  - Consistent Labeling Problem, 37
  - Constraint Satisfaction Problem, 36
    - globale Lösungsverfahren, 38
    - Konsistenztechniken, 40
    - Lösung eines CSP, 36, 42, 64, 67
    - lokale Lösungsverfahren, 40
  - Over-Constrained System (OCS), *siehe* überbeschränktes Constraint-System
  - Propagierung, 40
- Digitales Oberflächenmodell (DOM), 25, 58
- Dynamic CSP, 73, 82

- Ecke, 19
  - 3D-Rekonstruktion, 25
  - Koordinatenschätzung, 62
- ECLIPSE, 55, 78
- Entropie, 88, 102
- Fakt, 41, 54
- Fehlwert, 75, 77, 82, 86, 89
- Formparameter, 30
- Forward-Checking, 39, 43
- Fragmentierung, 59, 69
- Fremdverdeckung, 68
- Fuzzy-Logik, 68
- GDF, 61
- Gebäude, 21
- Gebäudeansicht, 28, 55
- Gebäudehypothese
  - Formparameter, 30
  - Generierung, 26
  - Parameterschätzung, 30, 62
  - Transformation in Constraints, 55, 79
  - Verifikation, 28
- Gebäudemodell
  - allgemeines Polyedermodell, 15
  - generisches, parametrisiertes, 19
  - parametrisierte Volumenprimitive, 16
  - prismatisches Polyedermodell, 14
- Gebäudeteil, 20, 26
- gemeinsame Information, 88, 91
- General OCS (GOCS), 73
- Generate & Test, 38
- Generieren-und-Verifizieren-Strategie, 13, 18, 22
- Generisches, parametrisiertes Modell, 19
- Gesamtstrategie, 22
- Graph
  - Adjazenzliste, *siehe* Adjazenzliste
  - Adjazenzmatrix, *siehe* Adjazenzmatrix
  - Clique, *siehe* Clique
  - Codierung, 106
  - Permutationsreihe, 107
  - plättbarer, 53
  - planarer, 53
- Gruppierung, 59
  - Einbettung in CSP, 64
  - geometrische, 24, 62
  - topologische, 24
- HCLP, 72
- Heuristiken, 40, 96
  - A\*-Suche, 97
  - Best-First, 97
  - First-Fail, 40, 58, 80, 97
- Hierarchische Modellierung, 12
- Huffman-Code, 102
- Inferenzregel, 77, 81
  - Forward-Checking, 43, 55, 76
  - Look-Ahead, 43, 55, 76
- Information
  - bedingte, 88
  - eines Symbols, 87, 101
  - gemeinsame, 88, 91, 104
  - Redundanz, *siehe* Redundanz
- Informationstheorie, 87
  - Entropie, *siehe* Entropie
- Innere Orientierung, 31
- Klausel, 41, 55
- Kommunikationsproblem, 87
- Komposition, 37
- Konfigurationsproblem, 83
- Konklusion, 41
- Konnektor, 20
- Konsistenz
  - Kanten-, 40
  - Knoten-, 40
- Konsistenztechniken, 40
- Konstruktive Festkörpergeometrie (CSG), 11, 21
- Koordinatensystem
  - Modell-, 30
  - Welt-, 30
- Kraftsche Ungleichung, 102
- Kuratowski, Satz von, 53
- Logische Programmierung, 41
  - Anfrage, 41
  - Fakt, 41, 54
  - Klausel, 41, 55
  - Konklusion, 41
  - Prämisse, 41
  - Regel, 41, 54
- Look-Ahead, 39, 43
- MaxCSP, 72, 82
- Maximum A Posteriori-Strategie, 99, 101, 104
- Maximum Entropy-Strategie, 101
- Maximum Likelihood-Schätzung, 89, 101
- Mehrbildzuordnung, 23
- Merkmalsaggregat, 19, 21
- Merkmalsextraktion, 12, 23, 68
- Merkmalsnachbarschaftsgraph (MNG), 24
- Methode der kleinsten Fehlerquadrate, 33
- Minimaler Spannbaum, 109, 112
- Minimum Description Length-Prinzip, 101
- Modellmerkmale, 90

- Unterscheidung Dach  $\leftrightarrow$  Wand, 93, 114
- Neuronales Netz, 43, 68
- NP-Vollständigkeit, 37
- Objektmodell, 11, 21
- Ockham's Razor, 101
- Optimale Codierung, 102
- Over-Fitting, 105
- Parallelprojektion, 46
- Parametrisierte Volumenprimitive, 16
- Partial Constraint Solving (PCSP), 73
- Permutationsreihe, 107, 112
- Polyedermodell
  - allgemeines, 15
  - prismatisches, 14
- Präfixeigenschaft, 102, 113
- Prämisse, 41
- Prinzip der minimalen Beschreibungslänge, *siehe* Minimum Description Length-Prinzip
- Prolog, 41, 54
- Quantisierung, 68
- Randflächendarstellung (B-Rep), 11, 21, 45
- Rauschen, 68
- Redundanz, 110
- Regel, 41, 54
- Regions of Interest (ROI), 25
- Relation
  - Eigenschaften, 109
  - Kollinearitäts-, 47, 92, 112
  - Linienparallelitäts-, 45, 92, 112
  - Merkmalsnachbarschafts-, 44, 92, 110
  - Same\_Side\_Lines-, 48, 92, 112
  - totale, 94, 95
  - Unterscheidung Dach  $\leftrightarrow$  Wand, 93, 114
  - Verschiedenheits-, 50, 92, 113
- Relational Matching, 29, 60
  - inexakt, 86
- Relationalstruktur, 37, 43, 86, 89, 105
- Relaxation
  - kontinuierliche, 68
- Rotationsmatrix, 31
- Schwellwerte, 51, 71
- Segmentierung, 23, 59, 60
- Semantische Modellierung (DFG-Projekt), 18
- Skalierungsfaktor, 32
- Störmerkmal, 69
- Statusvariable, 75, 76, 95
- Strategie
  - Bottom-Up, 13, 17
  - Generieren und Verifizieren, 13, 18, 22
  - Gesamt-, 22
  - Top-Down, 13
- Subgraph-Isomorphismus, 37, 53, 86
- Suchbaum, 39, 58
- Suchraumeinschränkung, 36, 56, 58, 60, 79, 96
- symbolische Bildbeschreibung, 12
- symbolische Bildmerkmale, 12, 23
- Terminal, 20, 27
- Terminalkonfiguration, 76
- Top-Down-Strategie, 13
- Training
  - bedingte Wahrscheinlichkeiten, 93
  - Schwellwerte, 51
  - Trainingsdaten, 93
- Transitionssystem, 76
- Translationsmatrix, 31
- überbeschränktes Constraint-System, 71, 82
- Übergangsrelation, 76
- Übertragungskanal, 87
- Übertragungswahrscheinlichkeit, 88, 91, 93, 99
- Umringpolygon, 22, 112
- Unbeobachtbarkeit, 69, 71, 72, 74, 75
- Unschärfe, 46, 51, 68, 69, 71
- Unsicherheit, 63, 69, 71, 89
- Variableneliminierung, 71, 73, 75
- Verschmelzen von Merkmalen, 68, 71
- Wahrscheinlichkeit
  - a priori, 92
  - bedingte, 93, 99
- Wildcard, *siehe* Fehlwert
- Zentralprojektion, 31, 46
- Zuordnung
  - inexakte relationale, 86, 89, 100
  - relationale, 29, 60