

Bericht 59

Auswertung von digitisierten Fernsehbildern
zur Beschreibung bewegter Objekte

Bernd Radig

IfI-HH-B-59/79

März 1978

"The vision problem begins with a large gray-level intensity array, and culminates in a description, that depends on that array, and on the purpose for which it is being viewed. The question is, what has to go on in between."

D. Marr

oder kürzer

"An input picture is read into a computer.
We would like to analyze it."

A. Guzman

"You can't begin writing until you know what's going to happen
- but what happens is dictated by what has already been
written."

T. Stoppard

[Marr 75, p.4] [Guzman 71, p.325] [Newsweek 77]

1. Einleitung

1.1 Zusammenfassung

Eine Folge von Fernsehaufnahmen hält die Veränderungen in einer Szene fest. Das Gesichtsfeld der Kamera bestimmt den räumlichen Ausschnitt, Beginn und Länge der Folge den zeitlichen Ausschnitt der beobachteten Umwelt. Die Aufnahmen können einem Rechner als digitisierte Bilder eingegeben werden. Im Speicher des Rechners liegen die digitisierten Messwerte der Lichtintensität als Matrix von Grauwerten vor, da die Abtastung der Zeilenstruktur einer Schwarzweiss-Aufnahme folgt. Gegenstände in der Szene erzeugen in der Matrix eine örtliche Verteilung von Grauwerten, die wesentlich von der Oberfläche eines Gegenstandes und ihrer Beleuchtung bestimmt ist. Vorliegende Arbeit beschreibt einen Lösungsansatz für das Problem, aus einer Folge von Grauwertmatritzen Aufschluss über die in der Szene vorhandenen Gegenstände, insbesondere ihren Bewegungszustand zu gewinnen.

Die Überlegung, dass neben der Grösse eines einzelnen Grauwertes auch sein Verhältnis zu benachbarten Grauwerten ein Schlüssel für die Analyse der Bilder ist, führte zur Entwicklung eines Algorithmus, der aus einer Grauwertmatrix ein Feld von "Bildvektoren" erzeugt. Die Komponenten eines jeden Vektors beschreiben den mittleren Grauwert und den durch eine lineare Funktion approximierten Verlauf der Messwerte für einen einige Rasterpunkte grossen Ausschnitt der Grauwertmatrix. Die Vektoren werden den Knoten eines Graphen zugeordnet, dessen Kanten Beziehungen zwischen örtlich benachbarten Vektoren darstellen. Eine Hierarchie von Prozessen eliminiert Knoten und Kanten aus dem Graphen, wenn Werte von Vektorkomponenten, Relationen zwischen Vektoren und Merkmale von Teilgraphen vorgebbare Bedingungen nicht erfüllen. Voneinander isolierte Komponenten des Graphen repräsentieren Bereiche eines Bildes, die eine durch die Bedingungen festgelegte Charakteristik der Grauwertverteilung zeigen. Mehrfaches Durchlaufen der Prozesshierarchie mit unterschiedlichen Bedingungen extrahiert aus jedem Bild Bereiche verschiedener Art.

Liegen die Bereiche für eine Bildfolge vor, können Relationen zwischen ihnen berechnet werden. Die Ähnlichkeit von Bereichen aus zeitlich benachbarten Bildern wird durch den Vergleich der die Bereiche beschreibenden Merkmale bestimmt.

Durch transitives Verknüpfen ähnlicher Bereiche werden Bereichsfolgen gebildet. Durch Aufspalten oder Zusammenfassen von Bereichen verbunden mit Modifikationen der Folgen kann eine Verbesserung der Stetigkeit von Bereichsmerkmalen innerhalb einer Folge erreicht werden. Wird ein Gegenstand oder ein Teil von ihm während eines zeitlichen Abschnittes der Beobachtung verdeckt, so wird die von ihm erzeugte Folge von Bereichen unterbrochen. Mit Hilfe des Vergleiches von Merkmalen, die für jede Folge berechnet werden, gelingt es, Bruchstücke zusammenzufügen und damit das in dreidimensionalen Szenen auftretende Problem der Verdeckung anzugehen.

Um Objekte zu beschreiben, werden die Eigenschaften der Folgen oder der in ihnen vertretenen Bereiche benutzt. So lassen sich mit Hilfe der Geschwindigkeit der Bereichsschwerpunkte ruhende von bewegten Gegenständen unterscheiden. Eine systematische Änderung der Fläche weist, wenn eine Verdeckungssituation ausgeschlossen werden kann, auf eine perspektivische Verzerrung hin. Ihre Ursache kann eine Bewegung, die auf den Beobachter hin oder von ihm weg führt, oder eine Drehung des Gegenstandes sein. Ein Hinweis auf eine Drehung kann auch einer Richtungsänderung der in den Bereichen zusammengefassten Vektoren entnommen werden. Erzeugt ein Gegenstand mehr als einen Bereich in jedem Bild, so können Beziehungen zwischen den Bereichsfolgen berechnet und zur Beschreibung des Gegenstandes verwendet werden.

Es ist bei dem hier vorgestellten Ansatz auch möglich, zuerst Beziehungen zwischen Bereichen eines Bildes zu berechnen, bevor die zeitlich geordneten Folgen gebildet werden. Durch diese Möglichkeit kann ein pragmatischer Ausgleich zwischen zwei extremen Verfahrensweisen gesucht werden. Die eine ist, zuerst Korrespondenzen zwischen Ausschnitten der Grauwertmatrizen verschiedener, aufeinander folgender Aufnahmen herzustellen und diese Information bei der Interpretation der Bildfolge zu benutzen. Die andere analysiert jedes Bild für sich und bestimmt den Bewegungsablauf in der Szene über den Vergleich der Einzelbild-Interpretationsergebnisse. In meinem Verfahren ist die Übergangsebene, auf der die symbolische Beschreibung einzelner Bilder zu einer Beschreibung der Bildsequenz verknüpft wird, frei wählbar. Dabei ist zu bedenken, dass die Analyse von Bewegung einen wichtigen Beitrag zum Verständnis der Szene liefert. Daher lohnt es sich nach Beschreibungssymbolen zu suchen, die eine Einbeziehung der Zeitinformation schon auf einer geringen Abstraktionsebene gestatten. Kann das Herstellen von Korrespondenzen zwischen

Teilbereichen aus aufeinander folgenden Aufnahmen zuverlässig geschehen, lässt sich der weitere Analyseprozess auch innerhalb einer einzelnen Aufnahme aus der Folge wirkungsvoll unterstützen.

Die bisher erreichten Ergebnisse ermutigen mich, den im folgenden vorgestellten Ansatz zu verbessern und weiter auszubauen.

1.2 Gliederung

1. Einleitung

1.1 Zusammenfassung

1.2 Gliederung

1.3 Vorwort

2. Aufgabenbereich der digitalen Bildverarbeitung

2.1 Modellbildung in kognitiven Systemen

2.2 Einzelbild Auswertung

2.2.1 Analyse

2.2.2 Verifikation

2.3 Korrespondenz von Bildern

2.3.1 Räumliches Sehen

2.3.2 Dreidimensionale Objektrekonstruktion

2.3.3 Entdecken von Änderungen

2.4 Bildfolgen mit bewegten Objekten

2.4.1 Entdecken von Bewegung

2.4.2 Verfolgen von bewegten Objekten

2.4.3 Erkennen

3. Problem- und Lösungsrahmen

3.1 Beschränkungen

3.2 Ziel

3.3 Lösungsstrategie

3.4 Relationalstruktur

3.5 Weltmodell

4. Grauwerte und Bildvektoren

4.1 Vorverarbeitung

4.1.1 Abtastung

4.1.2 Vorbemerkungen

4.1.3 Definition der Bildvektoren

4.1.4 Erzeugen eines Vektorfeldes

4.2 Gradientenraum

5. Gruppieren von Bildvektoren

5.1 Filtern im Vektorfeld

5.2 Knotenkonsistenz

5.3 Kantenkonsistenz

5.4 Bereichskonsistenz

5.4.1 Graphsuche

5.4.2 Bereichsbeschreibung

5.4.3 Bereichsänderung

6. Bildfolgen mit bewegten Objekten

6.1 Vergleich von Bereichen

6.1.1 Vergleich von Merkmalen

6.1.2 Vergleich zweier Bereiche

6.1.3 Bereichsaddition

6.2 Bereichsfolgen

6.2.1 Transitive Identitätsrelation

6.2.2 Alternative Zerlegung

6.3 Kombination von Folgen

6.3.1 Zeitliche Konkatenation

6.3.2 Räumliche Selektion

7. Ausklang

Literaturverzeichnis

1.3 Vorwort

In den letzten fünf Jahren baute am Institut für Informatik die Forschungsgruppe "Digitale Verarbeitung kontinuierlicher Signale" eine umfangreiche Anlage zur Bearbeitung von Fernsehbildern auf. Die Mitglieder der Gruppe waren in der Zeit R. Bertelsmeier, H. Kemen, H.-H. Nagel, B. Neumann und der Autor selbst. Sie wurden im Labor und im Sekretariat unterstützt von P. Cord, I. Heer und R. Jancke. Ihnen allen sei an dieser Stelle für ihren Beitrag zum Gelingen vorliegender Arbeit gedankt.

Die experimentelle Apparatur umfasst

- zwei Schwarzweiss-Fernsehkameras und vier Monitore,
- einen schnellen Analog-Digital-Wandler, der in 40msec ein Fernsehbild in 573 Zeilen mit 512 Spalten und 256 Graustufen digitisieren kann,
- eine Magnetplatte zur analogen Speicherung von Bildfolgen und Einzelbildern mit einer Kapazität von je 600 Bildern in zwei Kanälen,
- eine Halbleiter-Bildspeicherröhre zur Wiederdarstellung von digitisierten und bearbeiteten Bildern,
- einen Funkbild-Empfänger, umgebaut zu einem Ausgabegerät für digitisierte Grautonbilder und Graphik,
- eine Video-Kreuzschiene zur Verteilung der Analogsignale,
- einen Massenkernspeicher mit einer Kapazität von 3 Bildern einer verringerten Grauwertauflösung von 64 Stufen, jedoch mit einer Zugriffszeit von 100nsec auf ein Byte von 6 Bit Länge,
- ein lokales Rechnernetz mit fünf Prozessrechnern des Typs MINCAL 621 zur Steuerung der Apparatur und Vorverarbeitung der Bilder.

Das Rechnernetz ist über eine PDP-11/20 mit dem Zentralrechner des Instituts, einem DEC-System 10 mit KI-10 Prozessor und 128k Kernspeicher, verbunden.

Drei Aufgabengebiete wurden von der Gruppe bearbeitet.

- Während der Planung und Beschaffung der Apparatur wurden die Schnittstellen zu der erwähnten, speziellen Peripherie und die Verbindung zwischen MINCAL und PDP-11 konstruiert, gebaut und erprobt.
- Die Betriebssysteme der MINCAL Rechner wurden modifiziert und die notwendigen Dienstprogramme zur Unterstützung der Videoperipherie erstellt. Für den Kommunikationsrechner PDP-11 musste ein vollständig neues Betriebssystem entwickelt

werden, das die Möglichkeit bietet, gleichzeitig mehrere physikalische und logische Kommunikationskanäle zwischen den Prozessrechnern untereinander und dem Zentralrechner zu betreiben.

- Für die MINCAL Prozessrechner wurden PASCAL und CONCURRENT PASCAL Compiler zusammen mit der notwendigen Laufzeitunterstützung entwickelt, sowohl um die Formulierung von Anwendungsprogrammen in einer adäquaten Sprache zu ermöglichen, als auch um besser an unsere Anforderungen angepasste Betriebssysteme entwickeln zu können.

Obwohl in den genannten Bereichen die Arbeiten noch nicht abgeschlossen sind und bei einem solch komplexen, wachsenden System weiterhin neue Anforderungen bewältigt werden müssen, ist es doch seit mehr als zwei Jahren schon möglich, Bildfolgen zu speichern, zu digitisieren und zu bearbeiten. Ohne die Grundlage der gemeinsamen Anstrengungen hätte auch die vorliegende Arbeit nicht entstehen können.

Das folgende Kapitel enthält eine Übersicht von Problemen und Lösungsmethoden, die von anderen Autoren bisher im Zusammenhang mit der Bildanalyse aufgegriffen wurden. Der Themenkreis konzentriert sich dabei im Laufe des Kapitels auf das Problem der Beschreibung einer Bildfolge von Szenen mit sich bewegenden Objekten. Es schien mir notwendig, die Verflechtung dieser speziellen Aufgabenstellung mit vielen anderen Bereichen der Bildbeschreibung sichtbar zu machen. Die gebotene Kürze liess eine eingehende Diskussion nicht zu; ich habe versucht, auf einige grundlegende oder exemplarische Arbeiten so hinzuweisen, dass das Assoziationsfeld dieser Arbeit sichtbar wird.

Das dritte Kapitel beschreibt die Zielsetzung dieser Arbeit. Es nennt die Einschränkungen, die einerseits aus technischen Gründen hingenommen wurden, andererseits die Vielfalt der zu analysierenden Bildfolgen einengen. Es stellt den groben Rahmen für die Lösung des Problems vor, eine zuverlässige symbolische Beschreibung eines Bildes anzufertigen, auf deren Grundlage Bilder aus Folgen verglichen werden können. Die formale Struktur der Beschreibung und die Anpassung des Verfahrens an Bildinhalt und Aufgabenstellung werden erläutert.

Das vierte Kapitel enthält die technischen Einzelheiten der Digitisierung eines Bildes und stellt die Algorithmen der Vorverarbeitung vor, die eine Grauwertmatrix in ein Feld von "Bildvektoren" verwandeln. Die Vektoren bilden die Grundlage

für eine Strukturierung der in den Bildern enthaltenen Information. Der Grund für ihre Berechnung ist eine Konsequenz der Überlegung, dass nicht der einzelne Messwert sondern eher sein Bezug zu den Werten in seiner Umgebung wichtig für die Bildbeschreibung ist.

Das fünfte Kapitel beschreibt ein dreistufiges Verfahren, mit dem eine symbolische Beschreibung eines Bildes erzeugt wird. Dabei werden die Eigenschaften der Vektoren und ihre Beziehungen untereinander ausgewertet.

Mögliche Methoden zum Vergleich von Bildern auf Grund der symbolischen Beschreibung stellt das sechste Kapitel vor. Ein Anwendungsbeispiel wird gezeigt, in dem die Spur eines bewegten Objektes in einer Folge von Bildern ermittelt wird.

Den Schluss bildet ein Ausblick auf weitere, vertiefende Untersuchungen des vorgelegten Ansatzes und auf Erweiterungen des Verfahrens, die zur Berücksichtigung komplexer Situationen erforderlich werden können.

2. Aufgabenbereich der digitalen Bildverarbeitung

Ein System zur digitalen Bildverarbeitung erfasst einen Ausschnitt seiner Umwelt über optische Sensoren und wertet die so erhaltene Information mit Hilfe eines Digitalrechners aus. Dabei soll die Akzeptanzbreite der Sensoren nicht auf den sichtbaren Teil des elektromagnetischen Spektrums eingeschränkt sein, sondern auch Mikrowellen (Radar), den Infrarotbereich (Thermogramme), Röntgenaufnahmen, Szintigramme, Elektronenmikroskopie, usw. umfassen. Ganz allgemein wird eine räumliche oder flächige Verteilung von Werten physikalischer Größen in dem beobachteten Ausschnitt gemessen und einem Rechner in digitisierter Form zugeführt. Die Auswertung von Schwarzweiss-Fernsehbildern ist ein typisches Beispiel digitaler Bildverarbeitung, an dem sich allgemeine Probleme in einer Weise studieren lassen, dass ihre Lösungen auch für die anderen genannten Bereiche der Verarbeitung optischer Information von Bedeutung sind. Die Methoden, die hierbei entwickelt werden, lassen sich häufig auch auf die Verarbeitung zweidimensionaler Verteilungen anderer Messwerte übertragen.

In dieser Arbeit wird auf die Auswertung von Bildfolgen eingegangen. Speziell sind hier solche Folgen gemeint, die zusätzlich zur örtlichen die Bestimmung der zeitlichen Verteilung der Messwerte zulassen.

Das Gebiet der Szenenanalyse hat als Untersuchungsgegenstand einzelne Bilder oder Bildfolgen aus der natürlichen Umwelt des Menschen. Programme zu deren Verarbeitung werden oft "(visuelle) Kognitive Systeme" genannt. Die Forderung nach der Realisierbarkeit solcher Systeme erzwingt eine Einschränkung ihrer Anwendung auf einen Ausschnitt der möglichen Vielfalt von Szenen. Ihre Aufgabe ist es, eine symbolische Beschreibung der ihnen vorliegenden Eingabedaten bereitzustellen. Diese Abstraktionen sollen Eigenschaften und Verhaltensweisen der durch die Messwerte erfassten Objekte beschreiben und in einer geeigneten Form einem Interessenten zur Verfügung stellen.

2.1 Modellbildung in kognitiven Systemen

Fragesteller an ein Kognitives System kann z.B. ein Operateur in einem interaktiven System oder das Steuerungsprogramm in einem Robotersystem sein. Um die gewünschte Information zur Verfügung stellen zu können, muss ein kognitives System ein Modell seiner Aussenwelt entwickeln, aus dem die für den Benutzer wichtigen Aussagen entnommen werden können. Es ist um so zuverlässiger, je besser die Aussagen des Modells mit denen der Realität übereinstimmen; Minsky sagt in einem allgemeineren Zusammenhang [Minsky 65]:

"To an observer B, an object A^* is a model of an object A to the extent that B can use A^* to answer questions that interest him about A . . .
 . . . A^* is a good model of A, in B's view, to the extent that A^* 's answers agree with A's, on the whole, over those questions important to B."

In dieser Auffassung des Modellbegriffes ist nicht unterschieden, woher die Informationen kommen, aus denen das Modell der Aussenwelt aufgebaut ist. Beteiligt sind die Ergebnisse von Berechnungen, die mit den sensorischen Eingangsdaten vorgenommen wurden. Weiterhin gehören dazu interaktiv gegebene oder aus einer Datenbasis abgerufene Informationen. Schliesslich hat auch die prototypische Beschreibung der Objekte und Zusammenhänge, die vom Modell erfasst werden sollen, eine wichtige Auswirkung auf das Ergebnis der Modellbildung in einer gegebenen Situation. In der Literatur wird manchmal unter einem Modell nur die prototypische Beschreibung verstanden. Das Ausfüllen dieser Prototypen oder Schablonen wird als Inkarnation ("instance") eines solchen Modells bezeichnet. Meiner Intuition kommt die Definition von Minsky entgegen. Im allgemeinen Sprachgebrauch ist ein Modellbauer jemand, der z.B. nach den Plänen eines Architekten ein verkleinertes Modell eines Hauses herstellt, mit dessen Hilfe Konstruktionsfehler erkannt oder Fragen des ästhetischen Eindrucks geklärt werden können. Das Modell war dann gut, wenn die an ihm studierten Eigenschaften auch am gebauten Haus wiederfindbar sind.
 Einen Prototypen, z.B. eine Form, die in einer Spritzgussmaschine verwendet werden soll, stellt dagegen ein Werkzeugmacher her.

Aufnahmen aus der sichtbaren Umwelt oder natürliche Sprache des Menschen sind auch gerade deshalb bevorzugte Eingangsdaten für ein kognitives System, da die Überprüfung der von ihm entwickelten Modelle für den menschlichen Benutzer einfach ist. Würde man z.B. einem kognitiven System Daten aus Experimenten der Elementarteilchenphysik vorlegen, wäre die Verifikation des Systems mit einem ausserordentlich grossen theoretischen, experimentellen und finanziellen Aufwand verbunden. Neben der einfachen Ergebniskontrolle ist der Wunsch, mit einem Rechner in den gewohnten Kommunikationsmitteln (Schrift, Sprache, Bild) zu verkehren, ein Antrieb für die Entwicklung kognitiver Systeme. Ein weiterer Antrieb für diese Entwicklungen liegt in der praktischen Anwendbarkeit solcher Systeme. Oft sind Objekte nur optischen (berührungslosen) Messungen zugänglich oder werden durch sie am geringsten beeinflusst; dann ist man bei einer automatischen Auswertung tatsächlich gezwungen, das von einem kognitiven System entwickelte Modell über die interessierenden Objekteigenschaften zu befragen.

Vielleicht kann die Entwicklung kognitiver Systeme Anregungen für Fragestellungen geben bei der Untersuchung der Informationsverarbeitung in Lebewesen. Sicherlich lassen sich keine direkten Analogien herstellen, aber Denkmodelle und mathematische Strukturen zur Behandlung der Fragestellung könnten übertragen werden.

Ullman betont den Aspekt der Bewegung [Ullman 77]:

"On one hand, since it (the human visual system) is an already existing system which handles motion perception well, it can be used to guide the research. On the other hand, as the theory is developed, it can be used to illuminate aspects of motion perception by humans."

Kognitive Systeme können ihr Modell der Aussenwelt nicht allein aus den Eingabedaten gewinnen. Warum das so ist, geht aus ihrer Arbeitsweise hervor, die sich etwa folgendermassen charakterisieren lässt:

- Es gibt einen Vorverarbeitungsbereich, in dem die Eingangsdaten in einer systemspezifischen Datenstruktur gespeichert werden.
- Im Ausgabebereich werden die vom Interessenten gestellten Fragen in eine systeminterne Darstellung umgeformt und die Antworten in eine vom Benutzer gewünschte Form gebracht.

- Zwischen beiden liegt ein Transformationsbereich, der sich in Anlehnung an Newell und Ernst [Newell Ernst 65] so beschreiben lässt:

Gegeben sei eine Menge $\{ X \}$ von Objekten und eine Liste $\{ Q \}$ von Operatoren, so dass, falls Q auf X anwendbar ist, ein neues Objekt erzeugt wird. Gegeben sei eine Ausgangssituation XA (hergestellt durch den Vorverarbeitungsbereich) und eine Endsituation XZ (das aus der Fragestellung vom Ausgabebereich erzeugte Ziel). Zu finden ist eine Sequenz von Operationen $S: Q_1, \dots, Q_i$, die XA nach XZ transformiert. Danach müssen drei Arten von Informationen in einem kognitiven System vorhanden sein:

- eine Definition der Objekte X , die in der Ausgangssituation auftreten oder die während des Transformationsprozesses erzeugt werden können,
- eine Definition der Operatoren Q und ihre Anwendbarkeit auf die möglichen Objekte X ,
- und Information zur Einschränkung des Suchraumes beim Finden der Transformationssequenz S , da die kombinatorische Vielfalt eine enumerative Erzeugung von S verbietet.

Insbesondere die Definition der Objekte X gehört mit zum Modell des kognitiven Systems von seiner Umwelt, aber auch die Operatoren sowie die Voraussetzungen und Methoden, die benutzt werden, um den Suchraum einzuschränken, können dazu gehören. Dieses Wissen nennt man manchmal das Langzeitgedächtnis des Systems. Häufig bezeichnet man eine Datenbasis, die solche Informationen enthält, als "Weltmodell". Das Präfix "Welt" soll ausdrücken, dass es sich hierbei nicht um ein bestimmtes Modell einer konkreten Situation, sondern um Wissen über allgemeine Zusammenhänge und Gesetzmässigkeiten handelt. Das Kurzzeitgedächtnis nimmt das Ergebnis der aktuellen Analyse auf, also das Modell für eine bestimmte, beobachtete Situation. Systemen, die eine Folge von Beobachtungen auswerten und miteinander in Beziehung setzen, kann man ein mittelfristiges Gedächtnis zusprechen.

2.2 Einzelbild Auswertung

Viele und früher wohl die meisten Systeme beschäftigen sich mit der Verarbeitung von einzelnen Bildern, d.h. im System selbst werden keine Ergebnisse gespeichert, die die Untersuchung neuer Bilder beeinflussen. Ich gehe nicht ein auf Bildtransformationen, die beispielsweise zum Ziel haben

- den Kontrast zu ändern,
- das Rauschen oder andere Bildstörungen zu vermindern,
- das Datenvolumen zu verringern

und ähnliche Aufgaben (s. z.B. [Rosenfeld Kak 76]), sondern beschränke mich auf das Gebiet der Szenenanalyse.

2.2.1 Analyse

Die Szenenanalyse beschäftigt sich mit Bildern, die als Aufnahmen realer Objekte entstanden sind. Dienten früher Objekte von einfacher geometrischer Form als Vorlagen [Guzman 68], [Waltz 72], [Turner 74], so beschäftigt man sich heute auch mit natürlichen Szenen, damit sind Landschaften [Bajcsy Lieberman 74], [Ohlander 75], [Sakai et al. 76], [Rubin Reddy 77], Büros [Tenenbaum 73], Gesichter [Kelly 70], [Baird Kelly 74a] und andere in der Umwelt des Menschen vorkommende visuelle Eindrücke gemeint.

Das Ergebnis der Analyse ist eine Beschreibung des vorgelegten Bildes durch eine Zuordnung von Bildausschnitten und Namen, die sich auf Objekte aus dem Weltmodell beziehen (Baum, Telefon, Nase). Dabei wird versucht, die ein Objekt beschreibenden Eigenschaften, seine interne Struktur und seine Beziehungen zu anderen Objekten aus dem Bild zu ermitteln [Guzman 71], [Barrow et al. 72]. Baird und Kelly wiesen besonders auf die Notwendigkeit hin, auch funktionale Beziehungen zwischen Objekten zur Interpretation zu benutzen [Baird Kelly 74b].

Fast immer geht dem Zuordnungsprozess eine Zerlegung des Bildes (Segmentation) in Bereiche und Konturen voraus. Die Zahl der hierfür entwickelten Verfahren ist gross, ich verweise auf die jährlich von A. Rosenfeld in der Zeitschrift "Computer Graphics and Image Processing" erscheinende Literaturübersicht, ebenso auf die dort erschienenen Übersichtsartikel von Davis [Davis 73] sowie Riseman und Arbib [Riseman Arbib 77].

Horn und Woodham haben gezeigt, dass man durch Ausnutzen photometrischer Gesetze aus dem Verlauf der Lichtintensität auf der Oberfläche eines Gegenstandes im Prinzip Aussagen über die Orientierung von differentiellen Oberflächenelementen bezogen auf Lichtquelle und Beobachter machen kann [Horn 74,75,77], [Woodham 77]. Die Lage einer Oberfläche im Raum lässt sich auf Grund der lokalen Werte für die Orientierung der Oberflächenelemente bestimmen. Es kann entschieden werden, ob eine Wölbung nach innen oder nach aussen tritt. Die Orientierung eines Oberflächenelementes wird durch seinen in der z-Komponente normierten Normalenvektor $(p, q, -1)$ angegeben. Jeder Orientierung ist damit ein Punkt (p, q) im sogenannten Gradientenraum [Mackworth 73] zugeordnet. Da die Lichtreflexion insbesondere von den Winkeln zwischen Einfallrichtung, Beobachtungsrichtung und Flächennormale abhängt, bietet der Gradientenraum einen problembezogenen und eleganten Formalismus.

Horn hat umfangreiche Untersuchungen darüber angestellt [Horn 77], unter welchen Voraussetzungen näherungsweise ein Zusammenhang zwischen den Grauwerten eines projizierten Bildes und der Lage im Gradientenraum für die sichtbaren Objektoberflächen herstellbar ist. Viele Segmentationsverfahren benutzen intuitive Vorstellungen von diesem Zusammenhang. Die Überlegungen von Horn und Woodham beziehen sich im wesentlichen auf matte, gleichmässig gefärbte Oberflächen. Auf die Analyse natürlicher Szenen lassen sie sich global nicht übertragen. Das Reflexionsverhalten der hier auftretenden Oberflächen folgt selten dem Lambertschen Gesetz, sie können im Extremfalle die Eigenschaften eines Spiegels haben. Ausserdem ist die Beleuchtung der Szene durch direktes Licht, diffuses Streulicht, reflektiertes Licht und Schatten kaum kalkulierbar. Die Voraussetzungen der Theorie von Woodham und Horn sind häufig nur in Teilbereichen eines Bildes erfüllt. So kann der Intensitätsverlauf für zwei in einer Ecke zusammenstossende Oberflächen durchaus der erwarteten Gesetzmässigkeit folgen. Auch die für gewölbte Flächen typische Grauwertverteilung ist manchmal in natürlichen Szenen zu beobachten.

In Bild 4.5, das ich später noch erläutere, sind Richtung und Stärke des Grauwertgradienten für einen Bildausschnitt dargestellt, der ein Fahrzeug auf einer Strasse zeigt. Man sieht deutlich beim konvex gewölbten Wagendach und den Kotflügeln, wie sich um die Reflexionsmaxima Höhenlinien-ähnliche Strukturen herausbilden. Leider gibt es kaum Segmentationsverfahren, die diese Mikrostruktur der Intensitätsverteilung wirkungsvoll ausnutzen und sie in ihrer Ergebnisdatei beschreiben. Einer der wenigen

Autoren, die erkannt haben, dass die Grauwertverteilung in der Umgebung einer Kontur eine wichtige Information darstellt, ist Marr [Marr 75]. Er benutzt verschiedene Operatoren, um Art und Ausdehnung von Grauwertübergängen zu ermitteln und er beschreibt die Art dieser Übergänge durch Begriffe wie 'Stufe' oder 'Dach' und die Ausdehnung als 'scharf', 'breit' oder 'verschwommen'.

Die meisten Segmentationsverfahren stützen sich jedoch nur auf das bloße Vorhandensein von Grauwertänderungen und nicht auf die Systematik ihres Verlaufs. Je nachdem, ob man von Konturen oder Bereichen ausgeht, ergeben sich zwei Segmentationsprinzipien.

Bei der Suche nach Konturen werden Änderungen der Bildfunktion (Grauwertübergänge) oder lokale Schätzungen von statistischen Parametern der Grauwertverteilung benutzt [Yakimovsky Cunningham 76a], [Davis 73], [Nagel 76a], [Martelli 72,76], [Marr 75], [Duda Hart 73], [Rosenfeld Kak 76]. Die von geschlossenen Konturzügen berandeten Bildausschnitte können als Bereiche in die Ergebnisdatei aufgenommen werden [Yakimovsky 75], [Kraasch Zach 78].

Bei der Suche nach Bereichen wird eine optimale Zerlegung des Bildes in Gebiete mit möglichst homogenen Merkmalen angestrebt. Dabei kann man zwei Strategien einschlagen.

Man kann von einer sehr starken Anforderung an die Homogenität der Bereichsmerkmale ausgehen. Das führt zur Entstehung vieler Bereiche, die in einem zweiten Verarbeitungsschritt miteinander verschmolzen werden, bis eine Zerlegung mit den vorgegebenen Eigenschaften entsteht [Brice Fennema 70], [Yakimovsky 73], [Zucker 76a], [Yakimovsky Cunningham 76a], [Ernst et al. 76].

Die andere Methode zerlegt das Bild sukzessive ausgehend vom ganzen Bild, wobei jeweils die Bereiche weiter aufgespalten werden, bei denen die Verteilung der Werte von ausgewählten Merkmalen zu ungleichmäßig ist [Klinger Dyer 76], [Ohlander 75], [Nagin et al. 77], [Price 76]. Auch eine Kombination beider Methoden wird verwendet [Horowitz Pavlidis 74].

Konturen, als die Ränder der entstandenen Bereiche betrachtet, können ebenfalls zur Beschreibung der Segmentation des Bildes dienen.

Foith baut eine Baumstruktur während einer sequentiellen Abtastung für jede Zeile im Bild auf [Foith 77] (s.a. [Ehrich Foith 76]). Sie beschreibt das Intensitätsprofil als Gruppierung von Bergen und Plateaus. Flächige Strukturen entstehen durch die Verknüpfung der Bäume über die Zeilen hinweg.

Die bei der Zerlegung des Bildes entstandene Datei kann man als eine Ebene der symbolischen Beschreibung des Bildinhalts bezeichnen. Andere Ebenen können z.B. durch eine Formanalyse der Konturen [Turner 74], [Shirai 75], [Perkins 76], [Pavlidis 76] oder durch Gruppierung von Bereichen nach topologischen Gesichtspunkten [Foith 77], [Tanimoto Pavlidis 76] gegeben sein.

Die Abstraktion des Eingabebildes durch die Segmentation gestattet es, die Verbindung mit dem Langzeitgedächtnis des Systems herzustellen. Das Resultat der Interpretation ist dann ein aktuelles Modell der Systemumwelt.

In den letzten zehn Jahren ist eine Entwicklung deutlich geworden, die zum Ziel hat, die in Absatz 2.1 erwähnte Kenntnis über die Objekte, die Operatoren und die Transformationssequenz immer stärker zu formalisieren und so zu strukturieren, dass eigenständige, abgrenzbare Bereiche sichtbar werden. Die Serie der "blocks world" Programme angefangen mit dem ersten Versuch von Roberts [Roberts 65] illustriert diesen Prozess. Guzman [Guzman 68] sah für die Behandlung verschiedener Situationen spezielle Prozeduren vor, in denen das Wissen über Konstellationen von Kanten, Eckpunkten und Oberflächen der Quadergebirge so gesammelt war, dass sie auf ganze Klassen von Objekten anwendbar waren. Falk [Falk 70], Huffman [Huffman 71], Clowes [Clowes 71] und Mackworth [Mackworth 73] versuchten, das Wissen über Gesetzmässigkeiten in der Welt der Blöcke zu systematisieren; Waltz [Waltz 72] gelang es mit seinem sogenannten "Filter" Verfahren, die Gesetzmässigkeiten formal elegant zur Analyse zu benutzen. Er demonstrierte den dadurch gewonnenen Fortschritt daran, dass er eine Schatten verursachende Beleuchtung seiner Blockszene zulassen konnte. Turner [Turner 74] erweiterte die Möglichkeiten hin zur Beschreibung von Objekten mit gewölbten Oberflächen und der dort auftretenden Schatten.

Das Verfahren, den zu interpretierenden Objekten alle zulässigen Bedeutungen, unterschieden durch Marken ("label") an den Objekten, zu geben und dann in einem Filterprozess unverträgliche Marken zu eliminieren, hat inzwischen eine weit verbreitete Anwendung gefunden. Halten die vorgefundenen Merkmale eines Objektes nicht die Werte ein, die durch die Belegung mit einer Marke gefordert werden, so ist die Marke nicht mit den Merkmalen verträglich. Ebenso kann eine Beziehung, die zwischen Objekten herrscht, eine bestimmte Kombination von Marken für diese Objekte ausschliessen. Mackworth [Mackworth 77] betrachtet die Objekte als Knoten eines Graphen und die (binären) Relationen zwischen ihnen als

dessen Kanten. Er definiert Knoten-, Kanten- oder Paar- sowie Pfadkonsistenz. Knotenkonsistenz ist erreicht, wenn die Markierung des Knotens mit den lokalen Eigenschaften des repräsentierten Objektes übereinstimmt. Kantenkonsistenz ist erreicht, wenn zu jeder Markierung eines Knotens in allen benachbarten Knoten verträgliche Marken existieren. Pfadkonsistenz ist erreicht, wenn auf allen Wegen, die zwischen zwei beliebigen Knoten möglich sind, nur konsistente Kanten benutzt werden. Freuder schlägt vor [Freuder 76], eine Hierarchie von Zwangsbedingungen aufzubauen, um eine 'globale' Konsistenz zu erreichen.

Eine neuere Anwendung des Filter Prinzips stammt von Tenenbaum und Barrow [Tenenbaum Barrow 76]. Ihr System geht aus von einer ersten Zerlegung eines Bildes in Elementarbereiche und belegt sie mit den ihren Merkmalen verträglichen Marken. Für im Bild benachbarte Bereiche wird die Paarkonsistenz hergestellt. Iterativ werden die beiden benachbarten Bereiche verschmolzen, bei denen die Werte ihrer Merkmale und paarweisen Beziehungen die jeweils grösste Sicherheit bieten, dass beide zu der Projektion des selben realen Objektes gehören. Eine Voraussetzung für das Verschmelzen ist es, dass beide Bereiche einen gemeinsamen Durchschnitt von Marken haben. Bereiche, die definitiv verschiedene Bedeutungen haben, können also nie mehr miteinander verschmolzen werden, auch wenn die aus ihren Merkmalen berechnete Ähnlichkeit sie an die Spitze der Ähnlichkeitsskala schiebt. Auf Grund der Merkmale des neuen Bereiches können eventuell einige Deutungsmöglichkeiten fallengelassen werden und in Konsequenz davon auch Marken von benachbarten Bereichen. Wenn z.B. ein Bereich zu gross wird, als dass er der Sitz eines Stuhles sein könnte, so darf auch die Möglichkeit eliminiert werden, dass ein Bereich angrenzend über ihm die Rückenlehne eines Stuhles ist. Die Iteration endet, wenn alle benachbarten Bereiche verschiedene Bedeutung haben.

Einen vom Prinzip her ähnlichen Ansatz beschreiben Rosenfeld und seine Mitarbeiter [Rosenfeld et al. 76], [Davis Rosenfeld 76]. Sie schlagen vor, den Marken Konfidenzwerte zuzuordnen, die aussagen, für wie wahrscheinlich eine bestimmte Bedeutung zu einem Zeitpunkt gehalten wird. Sie geben Methoden an, mit denen abhängig von der Konfidenzverteilung über den Objektmarken und den Beziehungen zwischen den Objekten die Konfidenz von Marken erhöht oder abgesenkt werden kann. Zucker charakterisiert die Anwendbarkeit des Verfahrens so [Zucker 76b, p.856]:

"Relaxation labelling is an approach to the reduction of local

ambiguities which only requires knowledge about how local symbols can interact. These techniques should be applicable to problems in any domain which are specifiable in this fashion." Im Zusammenhang eines grösseren Systems wird das Verfahren im VISIONS System von Riseman und Mitarbeitern [Arbib Riseman 76], [Hanson Riseman 76], [Williams Lowrance 77] zur Segmentation benutzt [Prager et al. 77].

Das folgende Beispiel für einen anderen Ansatz zeigt, wie eine Formalisierung von Wissen und Kontrollstrukturen zu klaren und effizienten Programmen führt.

1970 beschrieb Kelly ein System, das in der Lage ist, Schwarzweiss-Aufnahmen von menschlichen Gesichtern (en face) zu analysieren und zu vermessen [Kelly 70]. Das System verwendet eine Vielzahl von speziellen Prozeduren, die die Aufgabe haben, nach Merkmalen zu suchen, die für bestimmte Gesichtspartien charakteristisch sind. So wurde z.B. zur Lokalisierung der Augen in den Bildzeilen nach einem Grauwertverlauf gesucht, der ausgehend von einem mittleren Wert für die Haut heller zum Augapfel wird, für die Iris dunkel wird, anschliessend nochmals zu weiss übergeht und zum Schluss wieder bei dem Wert für die Haut ankommt.

1974 griffen Baird und Kelly die Aufgabenstellung noch einmal auf [Baird Kelly 74a]. Nach einer einfachen Zerlegung des Bildes werden die entstandenen Bereiche durch acht Merkmale wie 'klein', 'gross', 'flach', 'hoch', ... und durch den Schwerpunkt beschrieben. Neun Relationen zwischen den Bereichen werden berechnet wie 'total rechts von', 'über', 'dicht neben' und ähnliche. Zur weiteren Analyse verwenden sie ein Produktionensystem. (Eine gute Einführung in Produktionensysteme gibt [Waterman 77]). Eine der vier aufgestellten Regeln lautet:

Gibt es vier Objekte (Bereiche) O_1 , O_2 , O_3 und O_4 , so dass O_1 und O_3 nebeneinander und total über O_2 bzw. O_4 sind, und O_3 und O_4 sind total rechts von O_1 und O_2 dann folgt: O_1 ist die linke Augenbraue, O_2 das linke Auge, O_3 die rechte Augenbraue und O_4 das rechte Auge.

Insgesamt wurden zehn Bilder untersucht. Bei zweien wurden Nasenlöcher und Mund nicht erkannt. Bei allen Bildern stimmten die Abstandsmessungen zwischen den Objektzentren, z.B. den Augenbrauen, mit Handvermessungen an den photographischen Vorlagen und mit den Werten, die das ältere Programm von Kelly ermittelt hatte, sehr gut überein. Der Aufwand zum Erstellen und Testen des Programms, die Länge des Programmcodes und die Laufzeit betragen jedoch nur ein Bruchteil von dem des älteren

Programms.

Eine weitere Strömung, formale Methoden zur Bildanalyse einzusetzen, besteht in der Verwendung linguistischer Techniken. Fu weist jedoch darauf hin [Fu 74], dass der Aufwand für die syntaktische Analyse schnell steigt, wenn die Ausdruckskraft der Sprache erhöht wird. Es ist auch noch kein System bekannt, das Grautonbilder natürlicher Szenen mit Hilfe linguistischer Methoden befriedigend analysieren kann. Fu weist ebenfalls darauf hin, dass zwar kein prinzipieller Unterschied zwischen einem syntaktischen und einem entscheidungstheoretischen Ansatz besteht, aber die syntaktische Analyse auch eine Beschreibung der Struktur und nicht nur eine Klassifikation liefert. Einen entscheidungstheoretischen Ansatz verfolgte Sties in seiner Arbeit über die Beschreibung und Klassifikation von Linienstrukturen [Sties 73].

Bei der Bildbeschreibung ist eine Schwierigkeit, dass eine so natürliche Operation wie die Konkatenation von Symbolen zu einer eindimensionalen Folge im zweidimensionalen Fall nicht mehr gegeben ist. Milgram und Rosenfeld überlegten sich Feld- ("array") Grammatiken [Milgram Rosenfeld 71], um die eindimensionale Satzform zu erweitern. Rosenfeld bemerkt dazu [Rosenfeld 71, p.282]:

"Since a picture does not ordinarily have a natural description as a string of subpictures, phrase-structure grammars as such are not a natural tool for picture generation or analysis. In general a (discrete, for example, digital) picture is an ARRAY of elements having given colors or gray levels; these elements can be regarded as the symbols of a terminal vocabulary. A grammar for a language whose 'sentences' are such pictures would have to have rules 'alpha' --> 'beta' which rewrite arrays as arrays;"

Weiter verbreitet sind die von Pfaltz und Rosenfeld konzipierten WEB-Grammatiken [Pfaltz Rosenfeld 69]. WEBs sind gerichtete, markierte Graphen, die Konkatenation ist ersetzt durch Einbettungsregeln für Teilgraphen. Pfaltz beschreibt ihre Anwendung für die Bildauswertung am Beispiel von Neuronennetzen [Pfaltz 72], für das er auch einen Analysealgorithmus (in FORTRAN) implementiert hat.

Brayer und Fu wenden WEB-Grammatiken auf die Analyse von multispektralen Aufnahmen der Erdoberfläche aus Satelliten an [Brayer Fu 76]. Hier ist besonders die Anwendung von rekursiven Regeln bemerkenswert, um z.B. Linien beliebiger

Länge erzeugen zu können.

Als letztes Beispiel für ein syntaktisches Verfahren erwähne ich noch den deterministischen Parser von Mohr und Haton [Mohr Haton 76]. Das angegebene Beispiel einer zweidimensionalen Grammatik für geradlinige Strichzeichnungen zeigt, dass durch Ausnutzen von Kontext auch gestörte (zu kurze oder zu lange Geraden, fehlende Linien) Muster erkannt werden können. Die Autoren geben an, wie die Produktionen beschaffen sein müssen, damit ein deterministisches Verfahren angewendet werden kann. Ausserdem muss der Analysevorgang von einem beliebigen primitiven Element des Musters einsetzen können, da ja nie feststeht, ob ein bestimmtes Element vorhanden ist.

Auf den ganzen Komplex der stochastischen oder "fuzzy" Grammatiken gehe ich nicht mehr ein, möchte jedoch noch drei Artikel nennen. Zadeh gibt in [Zadeh 72] eine schöne Einführung in die Definition und Semantik von "fuzzy" Sprachen, Huang und Fu beschreiben stochastische Syntaxanalyse für programmierte Grammatiken [Huang Fu 72] und Fu wendet eine stochastische Grammatik auf die Analyse von Chromosomen-Bildern an [Fu 73].

2.2.2 Verifikation

Ging es bei der Analyse darum, die Bedeutung eines vorgelegten Bildes mit Hilfe des Weltmodells herauszufinden, beschäftigt sich die Verifikation mit quantitativen Aussagen über die abgebildeten Objekte. Bolles beschreibt die Aufgaben der Verifikation so [Bolles 76]:

"verification vision (abbreviated VV) . . . uses a model of a scene to locate objects of interest in a picture of the scene. The characteristics that distinguish VV from other types of visual information processing are: (1) the system has a great deal of prior knowledge about the types, placement, and appearance of the objects that form the scene and (2) the goal is to verify and refine the location of one or more objects in the scene."

Die von Bolles beschriebene Ausgangssituation ist auch bei der Bearbeitung von Bildfolgen gegeben, wenn sich die Szene nur langsam ändert und die Analyse der ersten Bilder schon eine

gewisse Sicherheit der Interpretation ergeben hat.

Voraussetzung für die Verifikation sind Modelle für Kamera und Lichtquelle [Hannah 74], [Sobel 74], [Baumgart 74], [Bonde 78], [Parrish Goksel 77], bzw. eine Eichtransformation zwischen der Kamera-Sensorfläche und der Szene. Es muss ein Repertoire an Operatoren zur Extraktion von solchen Merkmalen vorhanden sein, die allein oder in Verbindung miteinander die interessierenden Eigenschaften der Objekte beschreiben. Bolles untersucht in der zitierten Arbeit die Anwendbarkeit und Zuverlässigkeit solcher Operatoren bezüglich der erwünschten Aussage. Er benutzt den Bayes Formalismus, um in einer Trainingsphase gewonnene Wahrscheinlichkeiten für die Signifikanz der Operatoren zur Vorausschätzung ihrer Beiträge zum Erreichen des Verifikationszieles verwenden zu können (s. auch [Duda et al. 76]). Ebenfalls lässt sich damit die Wahl und die Reihenfolge der Operatoren steuern, um mit geringstem Aufwand zu einer sicheren Bestimmung der Objekteigenschaften zu gelangen. Lässt sich eine Eigenschaft mit verschiedenen Kombinationen von Operatoren berechnen, so wird das endgültige Ergebnis nach der Methode der kleinsten Fehlerquadrate aus den einzelnen, gewichteten Resultaten kombiniert, wobei jedoch vorher die Verträglichkeit der Ergebnisse mit dem Szenenmodell überprüft worden ist.

Demuths Verfahren, ein einfaches, prototypisch beschriebenes Objekt zu vermessen [Demuth 76], benutzt ebenfalls die Technik der kleinsten Fehlerquadrate, um die Grauwertmatrix mit der Objektbeschreibung zu vergleichen. Baird beschreibt ein industriell einsetzbares System zur Fertigungskontrolle von integrierten Schaltungen [Baird 76]. Auch darauf trifft die Definition von Bolles zu. Baird zeigt, dass man eine spezialisierende Optimierung der Merkmalsextraktion bei stark eingeschränkten Szenen anwenden kann, so dass eine Aussonderung defekter Teile schritthaltend mit dem Produktionsprozess möglich wird.

2.3 Korrespondenz von Bildern

Die eben beschriebene Verifikation ist auch ein Teilprozess in Systemen, die mehrere Bilder miteinander korrelieren, um die Tiefenstruktur einer Szene aus Stereoaufnahmen zu erschliessen, um aus mehreren Objektansichten ein dreidimensionales Modell zu erstellen oder langfristige Änderungen zu beschreiben.

2.3.1 Räumliches Sehen

Die Tiefenstruktur einer Szene kann nicht bloss durch Bilder von Kameras aus verschiedenen Standpunkten erschlossen werden. Eine alternative Möglichkeit ist das direkte Ausmessen durch einen Laser-Entfernungsmesser [Tenenbaum 73], [Garvey 76], [Duda Nitzan 76], [Ishi Nagata 76], eine andere die spezielle Beleuchtung einer Szene z.B. durch ein Gitter von parallelen Streifen [Popplestone et al. 75] oder durch systematische Ablenkung eines Laserstrahles [Kiessling 76], eine dritte die Ermittlung von Textureigenschaften genügend grosser Flächen, so dass eine bekannte Gesetzmässigkeit zwischen perspektivischer Verzerrung und Änderung der beobachteten Textureigenschaften die Orientierung der Fläche im Raum gestattet [Yegi et al. 75], [Bajcsy Lieberman 76].

Im Zusammenhang mit der Korrelation von Bildern ist hier die Verwendung von mehreren Kameras oder Kamerastandorten interessant. Sieht man von sehr unterschiedlichen Blickwinkeln ab (z.B. orthogonale Projektion [Chang Wang 75]), so gelangt man zur Schmalwinkel Stereovision. Unter der Voraussetzung, dass beide Kameras in ihren optischen und elektrischen Eigenschaften genügend aufeinander abgestimmt sind, lässt sich ein Vergleich schon von sehr wenig abstrahierten Strukturen durchführen [Hannah 74]. Die Annahme ist, dass sich korrespondierende Konturen oder Bereiche durch ihre örtliche Nähe in beiden Bildern zuverlässig finden lassen. Sie stimmt jedoch nicht mehr bei zu grosser Störung der Sensorsignale oder bei ungünstiger Lage von Objekten oder Objektüberflächen. Gimelfarb et al. weisen darauf hin, dass es vorteilhaft sein kann, das Finden von korrespondierenden Bildpunkten mit dem Verstehen des Szeneninhaltes zu koppeln [Gimelfarb et al. 75]. Mori et al. versuchen, eine schnelle und störsichere Korrelation dadurch zu erreichen, dass sie den Bildvergleich zuerst für Bildausschnitte mit gutem Kontrast vornehmen und ihn

dann iterativ und abwechselnd mit korrigierenden Transformationen eines der beiden Bilder auf Nachbarausschnitte ausdehnen [Mori et al. 73]. Im Rahmen des Robot Projektes am Jet Propulsion Laboratory in Pasadena beschäftigten sich Levine et al. mit der Analyse der Tiefenstruktur einer Szene aus Fernseh-Stereoaufnahmen [Levine et al. 73] (vergl. auch [Pingle Thomas 74]).

Verwandt mit dieser Problemstellung ist die Rekonstruktion eines dreidimensionalen Objektes aus verschiedenen Ansichten.

2.3.2 Dreidimensionale Objektrekonstruktion

Burr und Chien benutzen Aufnahmen aus drei Kamerapositionen, so dass sie ein Referenzbild und eine Ansicht aus einem vertikal bzw. horizontal verschobenen Standpunkt zur Verfügung haben, um ein dreidimensionales Modell eines Objektes zu entwickeln [Burr Chien 77]. Zur Stereokorrelation der Bildpaare werden die mittleren Grauwerte aus 9×9 Fenstern der Bildmatrix benutzt, wenn der Intensitätsgradient dort eine Schwelle übersteigt. Aus den Bildern extrahierte Konturen werden durch Geradenstücke angenähert, und aus korrespondierenden Konturen können die entsprechenden Objektkanten im Raum lokalisiert werden. Es wird versucht, die Rekonstruktion mit einem Referenzmodell in Übereinstimmung zu bringen.

Es gibt einige andere Ansätze, um aus Bildern, die von mehreren Blickwinkeln aus aufgenommen sind, ein Objekt zu rekonstruieren.

Underwood und Coates geben ein Verfahren an, ein topologisches dreidimensionales Modell der Anordnung von Oberflächen und Kanten eines eben begrenzten, konvexen Objektes zu gewinnen [Underwood Coates 75].

Nevatia stellt ein Objekt auf einen Drehtisch, nimmt eine Serie von Bildern mit konstanter Winkelschrittweite auf und versucht, korrespondierende lokale Objektmerkmale (Bereiche grosser Grauwert-Varianz) in der Serie zu verfolgen und somit ihre räumliche Ortskurve zu ermitteln [Nevatia 76a]. Da die Objektbewegung nur einen Freiheitsgrad hat, kann der Suchraum stark eingeschränkt werden.

Baker beschreibt ein ähnliches Vorgehen [Baker 76,77]. Seine markanten Punkte sind Änderungen der Krümmung oder Endpunkte von Kreissegmenten, mit denen er durch Gradientenoperatoren gefundene Konturen approximiert hat. Sie dienen als Anker für die Korrelation der Bilder aus der Serie der Aufnahmen des

rotierten Objektes. Ist der Raumvektor solcher Punkte bestimmt, können mit ihrer Hilfe Oberflächenelemente des Objektmodells im Raum orientiert werden.

Neben Autoren, die sich stärker mit den Methoden beschäftigen, die zur rechnerinternen Darstellung komplexer, dreidimensionaler Gegenstände verwendet werden können - Nevatia [Nevatia 74], Braid [Braid 75], Marr [Marr 76] und Nishihara [Marr Nishihara 76] - möchte ich noch die Experimente von Baumgart [Baumgart 74], Shapira [Shapira 74], [Shapira Freeman 77] und Ullman [Ullman 77] erwähnen. Sie alle bieten Anregungen für die Analyse von Bildfolgen mit bewegten Objekten. Einerseits können in natürlichen Szenen Phänomene einer perspektivischen Verzerrung, Schatten und Verdeckung nur unvollkommen erklärt oder erkannt werden, ohne eine Vorstellung von der dreidimensionalen Struktur der Objekte und der Tiefenstruktur der Szene. Andererseits bieten die sich bewegenden Objekte dem Beobachter manchmal verschiedene, langsam ineinander übergehende Ansichten dar, so dass die Erkenntnisse aus den Drehtisch-Experimenten anwendbar sein können.

2.3.3 Entdecken von Änderungen

Als Vorstufe zum Entdecken von Bewegung in Bildfolgen ist das Beschreiben von Änderungen zwischen zwei Bildern zu betrachten. Ich gehe hier nur auf die Arbeit von Price ein, zum einen weil ich die dort angewandte Methodik, die symbolische Beschreibung von Bildern zu vergleichen, für richtig halte, zum anderen weil dort einige Schwierigkeiten geschildert werden, mit denen ich mich ebenfalls auseinandersetzen hatte und die ich zum Teil überwinden konnte [Price 76].

Die Grundlage für die symbolische Beschreibung der Bilder ist eine Zerlegung in Bereiche nach einem modifizierten Verfahren von Ohlander [Ohlander 75]. Es werden Histogramme über Merkmale von Ausschnitten des Bildes (angefangen mit dem ganzen Bild) berechnet. Merkmale, deren Verteilung eine Aufspaltung in zwei Gruppen erkennen lassen, werden benutzt, um die Pixel - das sind "picture elements", deren Merkmale ihre Koordinaten im Raster der Bildabtastung und die damit verbundenen Messwerte für die Szene sind - in Äquivalenzklassen einzuteilen. Der gerade bearbeitete Ausschnitt wird so

geteilt, dass die resultierenden Bereiche möglichst nur Pixel entweder aus der einen oder aus der anderen Äquivalenzklasse enthalten. Das Verfahren schreitet rekursiv weiter fort, bis die Merkmale der resultierenden Bereiche genügend homogen sind. Während Ohlanders erstes Programm noch mehrere Stunden Rechenzeit und interaktive Eingriffe brauchte, um ein Bild zu segmentieren, konnte durch Optimieren und durch ein von Price eingeführtes zweistufiges Verfahren die Rechenzeit für dasselbe Bild auf demselben Rechner auf etwa 20 Minuten gesenkt werden [Price Reddy 77]. Das zweistufige Verfahren arbeitet zuerst auf einem um den Faktor 8×8 reduzierten Bild, die dort gefundene Segmentierung wird in das Original zurückübertragen und dort weiter verfeinert.

Die Bereiche werden durch Merkmale beschrieben, wie ihre Fläche, das Verhältnis vom Quadrat des Umfanges zur Fläche (minimal bei einem Kreis), den Schwerpunkt, die Eigenschaften einer approximierenden Ellipse, die Farb- und Intensitätsverteilung und andere. Als Beziehungen zwischen zwei Bereichen werden ihre Nachbarschaft und ihre relative Lage ('über', 'unter', 'links von' und 'rechts von') berechnet.

Korrespondierende Bereiche werden in beiden Bildern auf Grund des Übereinstimmens ihrer Merkmale und Relationen gesucht. Dabei wird das Gewicht einzelner Merkmale verändert je nach der Erwartung, wie stark sich die Merkmale zwischen den Bildern geändert haben werden. Erfasst z.B. das zweite Bild einen leicht anderen Ausschnitt als das erste, so sollte der Vergleich der Schwerpunktkoordinaten weniger gewichtet werden; ein veränderter Aufnahmeabstand entwertet den Flächenvergleich. Sind korrespondierende Bereiche gefunden, kann der Vergleich ihrer Merkmale nun benutzt werden, um Unterschiede zwischen den Bildern zu beschreiben. Dabei sollten globale Änderungen wie die eben angeführten durch geeignete Transformationen der Bereichsmerkmale kompensiert werden. Das kann zuverlässig geschehen, wenn die Transformationsparameter mit Hilfe der Merkmale von vielen, richtig zugeordneten Bereichen berechnet werden können. Die Unterschiede zwischen zwei Bereichen beschreiben eben nur die Unterschiede zwischen Bereichen und nicht die der Objektabbildungen oder sogar der Objekte selbst. Daher ist auch das Verfahren von Price eher geeignet, die Güte des Segmentationsalgorithmus zu messen, als semantische Änderungen zwischen zwei Bildern festzustellen. Die deprimierende Erfahrung, dass bisher mancher noch so hoffnungsvoll ersonnene Segmentationsalgorithmus nicht in der Lage ist, Bilder von natürlichen Szenen, die eventuell noch bei sich ändernden Aufnahmebedingungen entstanden sind, zuverlässig zu beschreiben, hat wohl jeder gemacht, der sich mit dem

Vergleich solcher Bilder beschäftigte. Insbesondere das Problem, dass die Abbildung eines Objektes durch eine unterschiedliche Anzahl von Symbolen (Bereichen) beschrieben wird, ist schwer zu lösen. Ich verweise auf die folgenden Bemerkungen zu den Ansätzen von Martin und Aggarwal sowie Bertelsmeier und Radig. Die bedeutungsarme Detaillierung tritt immer stärker in Erscheinung, je mehr man versucht, den komplementären Fehler, nämlich die Beschreibung von mehreren Objektabbildungen durch ein einziges Symbol, zu vermeiden.

2.4 Bildfolgen mit bewegten Objekten

2.4.1 Entdecken von Bewegung

Wenn daher im folgenden von Objekten die Rede ist, sind damit keine Implikationen über ihre physikalische Natur verbunden. Sie sind eher durch die Eigenschaft definiert, auf der Bildmatrix räumlich zusammenhängende Bereiche mit einigermaßen homogenem Grauwert zu erzeugen. Zur Abgrenzung von Effekten, die durch das Rauschen der Messwerte verursacht werden, werden heuristische Annahmen über die Grösse des Kontrastes, die Grösse der Bereiche und die Kontinuität ihrer Merkmale über mehrere Bilder hinweg gemacht.

Um bewegte Objekte in (fast) Echtzeit verfolgen zu können, werden häufig noch spezielle Elektronik Systeme eingesetzt [Uno et al. 76], [Limb Murphy 75], [Onoe et al. 73] oder sehr starke Heuristiken für ein spezielles Problem benutzt [Chien Jones 75], z.B. in der Überwachung des Strassenverkehrs [Onoe Ohba 76], [Rüenaufner 76]. Ein interaktives System, das vom Operateur bezeichnete Ausschnitte, die je ein Auto enthalten, über eine Serie von Filmbildern mit Hilfe einer Korrelationstechnik verfolgt, wurde von Wolferts beschrieben [Wolferts 73,74].

Leese et al. geben zwei Verfahren an, um die Bewegung von Wolkenformationen aus Satellitenbildern zu erschliessen [Leese et al. 70]. In beiden wird jedes Bild in Felder von 64*64 Rasterpunkten unterteilt. Die Felder werden durch die Belegung mit einer Binärmatrix klassifiziert, die durch Vergleich der Grauwerte mit einem vorgegebenen Intervall entsteht. Können in den verglichenen Bildern in ihren

Merkmale korrespondierende Felder gefunden werden, so wird der Bewegungsvektor aus den Schwerpunktskoordinaten der Felder berechnet. Beschreiben die Merkmale eines Feldes übereinanderliegende Wolkenbänke mit unterschiedlicher Bewegung, so wird das Verfahren unzuverlässig. Weitere Autoren, die sich mit Wolkenbewegung befassen, sind Smith und Phillips [Smith Phillips 72] sowie Hall, Endlich et al. [Hall et al. 72].

Potter beschreibt zwei Methoden, die ebenfalls ohne Vorkenntnis der im Bild vorhandenen Objktanordnung auskommen. In seiner ersten Arbeit [Potter 75a] überzieht er die Bilder mit einem rechtwinkligen Gitter und bestimmt von jedem Kreuzungspunkt entlang der vier Gitterlinien den Abstand zur nächsten Diskontinuität der Bildfunktion. Für ein Bildpaar stehen dann für jeden Gitterpunkt je vier Abstandsmasse zur Verfügung, aus denen er einen Bewegungsvektor jedem Gitterpunkt zuordnet. Die Punkte werden klassifiziert als zu einem bewegten Objekt, zum Hintergrund oder zu einem Differenzbereich gehörend.

In seiner späteren Arbeit [Potter 75b] verwendet er kreuzförmige Masken, die von einem Gitterpunkt ausgehend in vier Richtungen bis zur nächsten Diskontinuität vorstossen. Mit einem heuristischen Verfahren versucht er ausgehend von gleichen Gitterpunkten die in beiden Bildern korrespondierenden Masken zu finden. Gelingt das, kann wiederum jedem Gitterpunkt ein Geschwindigkeitsvektor zugeordnet werden. Das Ziel ist eine Separierung der bewegten Objekte vom Hintergrund durch Gruppieren von Gitterpunkten mit demselben Bewegungsvektor. Schwierigkeiten, mit denen das Verfahren nicht fertig wird, sind unzuverlässige Bestimmung der Diskontinuitäten, Rotation der Objekte, gegenseitige Verdeckung von Objekten und Geschwindigkeitsänderungen, wenn sie zum Stillstand führen.

Ein einfaches Korrelationsexperiment wurde von Budy und Dreschler beschrieben [Budy Dreschler 76]. Aus dem ersten Bild einer Folge wird manuell ein Ausschnitt als Maske bestimmt, der genau ein bewegtes Objekt enthält. Ein etwas grösserer, umfassender Ausschnitt des zweiten Bildes dient als Suchrahmen für die Kreuzkorrelation mit der Maske aus dem ersten Bild. Die neue Position der Maske wird aus dem Korrelationsmaximum bestimmt. Für die folgenden Bilder wird entweder die Maske mit den Grauwerten aus dem ersten Bild weiterbenutzt oder sie wird jedesmal neu mit den Grauwerten aus dem letzten Bild gefüllt. Bei Drehung der Objekte und Eintreten anderer Objekte in den Ausschnitt versagt das Verfahren.

Nagel berichtet über einen Ansatz, der im Gegensatz zu dem von Potter mit segmentierten Bildern arbeitet [Nagel 76b]. Er zerlegt das Bild durch ein modifiziertes Verfahren [Nagel 76a] von Yakimovsky [Yakimovsky 75] in von geschlossenen Konturen berandete Bereiche und versucht, durch Vergleich von Bereichsmerkmalen zwischen zwei Bildserien, in denen sich ein Objekt in einem Bildausschnitt um mehr als seine Ausdehnung in Bewegungsrichtung verschoben hat, die Bereiche entweder dem stationären Teil oder der von der Objektbewegung herrührenden Bildkomponente zuzuordnen. Die grösste zusammenhängende Aggregation von benachbarten Bereichen, die zu der zweiten Gruppe gehören, wird als erste Schätzung der Objektbeschreibung benützt. Sie wird in einem Kreuzkorrelationsverfahren über zeitlich benachbarte Bilder verfeinert. Gleichzeitig wird der Geschwindigkeitsvektor berechnet.

Auch dieses Verfahren funktioniert nur dann, wenn genau ein Objekt relativ ungestört und ohne sich zu drehen durch den vorgegebenen Bildausschnitt wandert.

Ein Verfahren von Dreschler kann etwas weniger stark eingeschränkte Bildfolgen verarbeiten [Dreschler 77]. Ein fester Ausschnitt, der ein bewegtes Objekt in einer Serie von Bildern enthält, wird mit einem Referenzbild verglichen, in dem sich das Objekt nicht in dem Ausschnitt befindet. Übersteigt die Grauwertdifferenz gleicher Rasterpunkte einen vorgegebenen Schwellenwert, so werden diese Punkte in den Bildern der Serie markiert. In jedem Bild werden benachbarte, markierte Punkte zu Bereichen zusammengefasst. Für genügend grosse Bereiche werden der mittlere Grauwert und die Streuung, die Länge des Umfanges, die Zahl der den Bereich berandenden geschlossenen Kurven, die Schwerpunktskoordinaten und sein Kontrast gegen den Hintergrund als Merkmalsvektor berechnet. Alle Bereiche der Serie werden in einem "minimum spanning tree clustering" Verfahren gruppiert. Unter der Voraussetzung, dass das sich bewegendes Objekt in verschiedenen Bildern nicht durch eine unterschiedliche Anzahl von Bereichen repräsentiert wird, kann eine richtige Zuordnung der Bereiche, die von der Bewegung des Objektes herrühren, über alle Aufnahmen der Serie erfolgen. Dabei toleriert das Verfahren die Änderung einzelner Eigenschaften der Bereiche, z.B. die der Grösse. Es ist aber durch seine einfache Differenzbildung anfällig gegenüber Situationen, in denen das Objekt Gebiete des Hintergrundes passiert, die einen mit Teilen von ihm ähnlichen Grauwert haben. Das Problem der Verdeckung bleibt ungelöst und eine Drehung des Objektes wird unter der Voraussetzung nicht zu grosser Änderungen von Aufnahme zu Aufnahme toleriert, aber nicht erkannt.

Ein Problem solcher Häufungsgebiet-Analysen ist der starke Einfluss der Definition der Metrik im Merkmalsraum. Das Skalieren der einzelnen Achsen ist in ihrer Konsequenz nicht leicht zu durchschauen bzw. intuitiv verständlich.

Ein zweites Problem ist die Gleichbehandlung der Merkmale über die Orts- und Zeitkoordinaten. Nehmen wir als Beispiel den Kontrast eines Bereiches gegen den Hintergrund. Wenn sich das Objekt vor einem einigermaßen homogenen Hintergrund bewegt, ist der Kontrast sicherlich eine Eigenschaft, die zu einer Häufung der Merkmalsvektoren des Bereichs im Merkmalsraum beiträgt. Bewegt sich das Objekt jedoch vor einem schachbrettartigen Hintergrund, so ist es fraglich, ob die Konstanz der anderen Merkmale eine Aufspaltung der Wolke von Punkten im Merkmalsraum immer verhindern kann.

Ein drittes Problem ist die fehlende Unterscheidung zwischen systematischen und zufälligen Variationen von Merkmalen. Gerade weil der minimale spannende Baum in dem von Dreschler vorgeschlagenen Verfahren erst aufgebaut wird, wenn alle Beobachtungen ausgewertet sind, könnte ein systematischer Verlauf der Werte von Bereichsmerkmalen ausnutzbar sein. Eine Möglichkeit hierzu wäre die Häufungsanalyse nicht nur der Merkmale selbst, sondern auch von Merkmalsdifferenzen zeitlich benachbarter Bereiche, wobei die Menge der untersuchten Paare natürlich heuristisch eingeschränkt werden müsste.

Eine Verbesserung der Entscheidung, ob zwei Intensitätsmessungen am gleichen Ort in verschiedenen Bildern von derselben Objektoberfläche stammen, wird von Jain, Militzer und Nagel vorgestellt [Jain et al. 77]. Das Verfahren beruht auf der Idee, das von Yakimovsky vorgeschlagene Likelihood Verhältnis [Yakimovsky 75] zur Entscheidung, ob zwei örtlich benachbarte Bereiche aus derselben Normalverteilung der Messwerte stammen, als Kriterium zu benutzen, ob zwei zeitlich benachbarte Bereiche dieselbe Objektoberfläche beschreiben. Als Bereich werden rechteckige Ausschnitte von 4×6 Rasterpunkten benutzt, die das digitisierte Bild von 512×573 Punkten ohne gegenseitige Überlappung überdecken. Innerhalb einer Bildserie werden das zweite und die folgenden Bilder mit dem ersten verglichen. Für jedes Rechteck wird gezählt, wie oft entschieden wurde, dass die in ihm enthaltenen Grauwerte aus verschiedenen Verteilungen stammen.

Hier wird nun eine interessante Beobachtung gemacht, die an einem einfachen Beispiel erläutert werden soll. Angenommen, es bewege sich ein homogen graues Rechteck der Größe $x=20 * y=6$ Rasterpunkte in Zeilenrichtung vor einem homogenen, kontrastierenden Hintergrund mit einer Geschwindigkeit von 4 Rasterpunkten pro Bild. In der interessierenden Zeile ergeben

sich folgende Zählerstände:

2. Bild ...0010000100...
3. ...00210002100...
4. ...003210032100...
5. ...0043210432100...
6. ...00543215432100...
7. ...006543255432100...
8. ...0076543555432100...
9. ...00876545555432100...

Die Folgerungen aus dieser Beobachtung ziehen Jain und Nagel in ihrem Bericht [Jain Nagel 77]. Bezeichnet man das Feld von 128×96 Zählern als Differenzbild, so lassen sich in ihm sogenannte "aktive Bereiche" von Zählern bilden, deren Inhalt grösser als Null ist und die über eine Vierernachbarschaft zusammenhängen. Vorausgesetzt, dass die Likelihood Entscheidung richtig getroffen wurde, haben die aktiven Bereiche folgende Eigenschaften (bei einer einfachen Bewegung der Objekte):

- Sie dehnen sich in der Bewegungsrichtung der für ihr Entstehen verantwortlichen Objektprojektion (das ist die Projektion des Objektes auf die Bildebene der Kamera) aus.
- Die Zählerinhalte entlang einer Zeile oder Spalte bilden monotone Teilfolgen mit einer Länge grösser als Eins.
- Genügend grosse und nicht zu stark texturierte Objekte erzeugen mehrere aktive Bereiche (einen für jedes Grauwertplateau im Objekt), die spätestens dann zu einem verschmolzen sind, wenn sich die Objektprojektion um ihre Ausdehnung entlang ihrer Bewegungsrichtung verschoben hat.
- Ist das der Fall, kann man sich den aktiven Bereich als bestehend aus drei Teilbereichen denken:
 - 1) Der Teilbereich, der den Ort repräsentiert, an dem die Objektprojektion im ersten Bild war. In ihm nimmt der Zählerinhalt mit jedem weiteren verglichenen Bild der Folge zu.
 - 2) Der Teilbereich, der den augenblicklichen Ort beschreibt, auch hier erhöht sich der Zählerstand.
 - 3) Dazwischen liegt ein Teilbereich, der inzwischen wieder von der Objektprojektion freigegeben wurde, hier bleibt der Zählerinhalt konstant (falls kein anderes Objekt eintritt); dieser Teilbereich vergrössert sich mit jedem weiteren Bild.

Für jede Spalte und jede Zeile in einem aktiven Bereich wird ein Merkmalsvektor gebildet, der die Monotonie der aus den Zählerinhalten gebildeten Folge, die Auf- oder Abstiegstendenz,

den Anteil von Plateaus und andere Eigenschaften beschreibt. Aus den Merkmalsvektoren für die Einzelfolgen wird einer für den ganzen aktiven Bereich berechnet. Hierbei wird der Formalismus der "fuzzy" Logik [Zadeh 65,68,72] eingesetzt. Zusammen mit einem sogenannten Differenzbild zweiter Ordnung, in dem die Änderungen zwischen zwei zeitlich benachbarten Differenzbildern eingetragen sind, wird fortlaufend über eine Schätzung Buch geführt, welche aktiven Bereiche von einem bewegten Objekt generiert wurden und wann sich die Objektprojektion um mehr als ihre Ausdehnung in Bewegungsrichtung verschoben haben wird. Darauf aufbauend kann für jedes Bild angegeben werden, in welchem koordinatenparallelem Rechteck sich die Projektion eines bewegten Objektes gerade befindet. Beispiele für Strassenszenen werden gezeigt, in denen für zwei Objekte (Auto und Fussgänger) zutreffende Angaben gemacht werden konnten. Ein Fortschritt in diesem Ansatz ist, dass für eine Folge von Aufnahmen mit bewegten Objekten in natürlicher Umgebung ein Plan (im Sinne von Kelly [Kelly 71]) bereitgestellt werden kann [Nagel 77], der es erlaubt, eine nachfolgende Analyse auf die Bildbereiche zu konzentrieren, in denen ein bewegtes Objekt vermutet werden kann.

2.4.2 Verfolgen von bewegten Objekten

Während es in den vorher beschriebenen Beispielen darum ging, Bereiche in der Szenenfolge zu entdecken, in denen Bewegung stattfindet, wenden sich Aggarwal, Chow, Duda und Martin einem anderen Problem zu. Sie versuchen, starre Objekte in ihrem Bewegungsablauf zu verfolgen, auch wenn die Objekte sich drehen und verdecken. In allen Experimenten werden künstlich erzeugte oder auf Binärbilder reduzierte Aufnahmen benutzt, in denen die Objekte wie Schablonen auf kontrastierendem Hintergrund erscheinen.

Aggarwal und Duda lassen als Bereiche nur polygonal begrenzte Objekte zu, die allerdings Öffnungen im Inneren haben dürfen [Aggarwal Duda 75]. Jedes Bild wird beschrieben durch die sichtbaren Vertizes (Endpunkte von Geraden). Falls sich Polygone überlappen, entstehen falsche Vertizes. Sie können z.B. bei einer Relativbewegung der Polygone dadurch entdeckt werden, dass sich der Winkel zwischen den einlaufenden Geraden ändert oder dass sich die Länge dieser Geraden ändert. Vorausgesetzt, dass von Bild zu Bild nicht mehr als ein Vertex

verschwindet oder erscheint, können falsche von echten Vertizes unterschieden werden. Modelle für die Polygone werden aus den echten Vertizes gebildet und verfeinert. Es kann sogar eine Beschreibung eines Polygons erstellt werden, wenn zwar alle Vertizes einmal in der Bildsequenz sichtbar sind, nie aber alle gleichzeitig.

Chow und Aggarwal verzichten zwar auf die polygonale Berandung, müssen dafür aber ihre Bildfolgen durch andere Voraussetzungen einschränken [Chow Aggarwal 77]. Sie führen ein "prädiktives Modell" ein. Die Bereiche, die am Anfang einem, bei späterer Überdeckung mehreren Objekten entsprechen können, werden durch einfache Merkmale beschrieben (Fläche, Schwerpunkt, Hauptachsen, ...); mit ihrer Hilfe geschieht die Zuordnung entsprechender Bereiche in der Bildfolge. In der Anfangsphase der Bildsequenz sind die Objekte vollständig sichtbar, das prädiktive Modell wird aufgebaut und verfeinert und insbesondere die Schätzung der Bewegung abgesichert. Verschmelzen zwei Objektbereiche miteinander, so wird das prädiktive Modell für den Gesamtbereich erzeugt. Bewegen sich beide Objekte gleichförmig weiter, so kann das prädiktive Modell für jedes Bild berechnet werden und stimmt mit dem im Bild vorgefundenen Bereich überein.

Um nicht mehr auf die Anfangsphase angewiesen zu sein, mussten Martin und Aggarwal die ganzheitliche Beschreibung der Bereiche aufgeben [Martin Aggarwal 77].

Die geschlossenen Konturen, die die Bereiche umgeben, werden durch Stücke von Geraden und Kreisbögen angenähert. Die dabei benutzte Methode ist schon von einigen Autoren (Turner [Turner 74], Shirai [Shirai 75], McKee [McKee Aggarwal 76], Radig [Bertelsmeier Radig 77a]) erfolgreich angewendet worden und geht wahrscheinlich auf eine Idee von Barrow zurück [Barrow Popplestone 71]. Trägt man den Tangentialwinkel einer Kurve, der aus den im Raster erlaubten, diskreten Konturstücken angenähert werden muss, gegen die Bogenlänge der Kurve auf, so stellen sich Geradenstücke als Parallelen zur Längen-Achse dar, Kreise als Geraden, die umso steiler verlaufen, je geringer der Radius der Kurve ist [Barrow Popplestone 71, Appendix]. Diese Funktion kann man mittels der Methode der kleinsten Fehlerquadrate durch Geradenstücke annähern, deren Rücktransformation in das Konturbild eine Kennzeichnung von Konturstücken durch Radius und Bogenlänge ermöglicht.

Der Bildvergleich geht in zwei Stufen vor sich. Erst werden ähnliche Bereiche gesucht, dann gemeinsame Sequenzen von Konturstücken. Das Bilden einer Sequenz beginnt bei den zwei ähnlichsten Konturstücken, die in beiden Bildern an

korrespondierenden Bereichen anliegen. Daran werden in beiden Richtungen weitere Stücke angehängt, bis das Unterschreiten einer vorgegebenen Ähnlichkeitsschwelle das Verlängern der Sequenz beendet. Durch einen Vergleich von Längen und Orientierung beider Sequenzen wird entschieden, ob sie eine beiden Bildern gemeinsame Eigenschaft beschreiben. Bewegungsvektoren - zugelassen sind Translation und Rotation - werden für korrespondierende Sequenzen berechnet und bilden die Grundlage für die Hypothese, dass Sequenzen mit gleichen Bewegungsvektoren zu einem Objekt gehören. Solche Sequenzen werden zu Modellen von Objekten zusammengestellt.

Das System kann ein Objektmodell aufspalten, wenn sich herausstellt, dass zwei überlagerte Objekte sich zu trennen beginnen. Es wird verwirrt, wenn es das erste Element einer Sequenz falsch zugeordnet hat. Das passiert, wie die Autoren bemerken, wenn ein Objekt oder überlagerte Objekte mehrere ähnlich geformte Randstücke besitzen. Eine Abhilfe wäre, mehrere Zuordnungshypothesen und ihre Konsequenzen parallel zu verfolgen, wobei zu schlecht bewertete Objektmodelle im Laufe der Zeit fallengelassen werden könnten. Lowerre benutzt solch eine Methode, die besten N Hypothesen zu verfolgen, in seinem HARPY-System zur Spracherkennung [Lowerre 76]. Sie werden durch neue Informationen, die entlang der Zeitachse auftauchen, neu bewertet und umgeordnet.

Eine Ausweitbarkeit des Ansatzes von Martin und Aggarwal auf natürliche Szenen ist kaum möglich (und von den Autoren auch nicht angekündigt). Denn ein Problem ist es, die Bilder so zu segmentieren, dass dieselben Oberflächenteile der Objekte ähnliche Bereiche im zerlegten Bild erzeugen, so dass die erste Stufe des Vergleiches sicher durchgeführt werden könnte. Gelänge das, wäre schon ein Hauptproblem des Bildvergleiches gelöst. Ein anderes Problem liegt in der Methode, die zur Konturbeschreibung führt. Bei der Anpassung von Geraden an die Funktion des Tangentialwinkels über der Bogenlänge müssen Schwellen für den höchsten zugelassenen Fehler benutzt werden. Die Konturbeschreibung, wie sie vom Segmentationsprozess abgeliefert wird, variiert durch das der Intensitätsmessung überlagerte Rauschen und durch wechselnden Kontrast beim Passieren des Objektes über einen nicht homogenen Hintergrund. Häufig führt das dazu, dass die Konturen von Bild zu Bild in zwar global ähnliche, aber lokal verschieden viele und unterschiedlich lange Stücke zerlegt werden. Auch das noch zu berücksichtigen, würde die Kombinatorik der Zuordnung weiter erhöhen.

2.4.3 Erkennen

In diesem Abschnitt werden Ansätze vorgestellt, die versuchen, das Geschehen in einer Szene mit bewegten Objekten zu beschreiben. Dazu müssen Objekte identifiziert und ihre Interaktion mit anderen Objekten in der Szene analysiert werden.

Badler stellt ein System vor, das von einer idealen Beschreibung (Strichzeichnung) eines jeden Bildes in einer zeitlich geordneten Folge ausgeht [Badler 75]. Er führt "primitive events" ein, die Bewegungsabläufe - wie sie mit einfachen Verben (z.B. springen, anhalten, ruhen) assoziiert sind - mitsamt ihrer Dauer beschreiben. Die chronologische Ordnung der primitiven Ereignisse zusammen mit einer Beschreibung der Objekte ergibt den Handlungsablauf in der Szene. Von ihm kann das System eine verbale Beschreibung ausgeben. Eine kompakte interne Darstellung der Szenenfolge, die auch vom Interpretationsziel her naheliegt, erreicht Badler dadurch, dass er von Bild zu Bild eintretende Änderungen (s.a. Winston's "comparison notes" [Winston 75]) beschreibt.

Tsotsos baut auf Badlers Arbeit auf [Tsotsos 76]. Er versucht, erkennbare Schwächen in dessen Konzept zu beheben und die Voraussetzungen für eine Implementation eines eingeschränkten Systems zu schaffen.

Beide Arbeiten zusammen sind sicherlich ein grosser Schritt vorwärts bei der semantischen Ausdeutung von Bildfolgen mit bewegten Objekten. Was jedoch noch fehlt, und dazu liefert die hiermit vorgelegte Arbeit einen Beitrag, ist eine Brücke zu schlagen von einer Sequenz von Grauwert-Matrizen zu einer symbolischen Beschreibung der Bilder, so dass Badlers Anregungen aufgegriffen werden können.

Hogg macht einen Ansatz, ausgehend von Bild-Digitisierungen einer Büroszene einen Handlungsablauf zu entdecken und zu beschreiben [Hogg 76]. Er verwendet Bilddifferenzen, um Bereiche zu entdecken, in denen eine Änderung gegenüber einem Referenzbild (Hintergrund) aufgetreten ist. Die Bereiche werden durch die sie einschliessenden, koordinatenparallelen Rechtecke beschrieben, zu kleine Bereiche werden unterdrückt. Eine einfache Metrik ermittelt Abstand und Grössenunterschied zwischen zwei Rechtecken. Objekte bestehen aus einem einzelnen Rechteck oder einem Paar.

Die Korrespondenz zwischen Objekten in aufeinanderfolgenden Bildern wird durch Bewerten der Übereinstimmung in Grösse und

Form hergestellt, wobei die endgültige Zuordnung in der Reihenfolge absteigender Ähnlichkeit vorgenommen wird. Die Kombination von Rechtecken zu Objekten wird geleitet durch eine Vorgabe von bevorzugten Erscheinungsformen der real in der Szene vorkommenden Dinge. Aus der Positionsänderung der Objekte werden einfache Bewegungszustände abgeleitet, z.B. bewegen nach links oder rechts, anhalten, ruhen. Ein zweiter Typ von Bewegung kann aus einer Formänderung erkannt werden. Hogg gibt als Beispiel die Form 'schmales, hohes Rechteck' für einen stehenden und 'niedriges, breites Rechteck' für einen liegenden Menschen an. Ein Wechsel zwischen beiden Formen interpretiert er als Aufstehen bzw. Hinlegen. In dieser Phase kann er natürlich nur noch die Positionsinformation und nicht mehr die Formkonstanz als Korrespondenzmerkmal verwenden. Die Möglichkeit, ein Objekt als Paar von Rechtecken darzustellen, ist ein erster Versuch, mit Störungen fertig zu werden, die durch teilweise Verdeckung oder ungenügenden Kontrast zum Hintergrund entstehen. Hoggs Arbeit ist ein erster Ansatz, Handlung ausgehend von realen Szenen unter Benutzung möglichst einfacher Konzepte auf mehreren Abstraktionsebenen zu beschreiben.

Eine ganze Hierarchie von Abstraktionsebenen wird in einem von Bertelsmeier und Radig konzipierten System benutzt [Bertelsmeier Radig 77a,b]. Es reflektiert die Annahme, dass Eigenschaften realer Objekte, die sich direkt in der Grauwertmatrix der Szene manifestieren, nicht immer mit einer solchen Zuverlässigkeit zu ermitteln sind, um als Grundlage für einen Vergleich von Bildern einer Folge dienen zu können. Hauptursachen für die Störung der Grauwertrepräsentation von realen Objekten sind

- das Amplituden- und Rasterrauschen der Aufnahmeapparatur;
- Beleuchtungsänderungen von Teilen der Szene z.B. durch Wolkenschatten oder Umschalten von künstlichen Lichtquellen;
- die sich ändernden Reflexionsbedingungen für Oberflächen von bewegten Objekten;
- die teilweise Verdeckung von Objekten;
- die Änderungen durch eine perspektivische Projektion einer dreidimensionalen Szene in die zweidimensionale Bildebene.

Deswegen wurde entschieden, Informationen über die Strukturierung eines Bildes zu akkumulieren und erst, wenn eine genügende Sicherheit über die Zuordnung eines Bildausschnittes zu einem realen Objekt erreicht ist, einen Vergleich von Bildern an Hand der Objektbeschreibung vorzunehmen.

Die Organisationsform der Bildbeschreibung ist eine hierarchische Relationalstruktur (vgl. [Barow et al. 72]), deren Basis an einfache, durch eine Vorverarbeitungsstufe erzeugte Elemente geknüpft ist, während ihre Spitzen in einer Relationalstruktur verankert sind, die die momentane Konfiguration von realen Objekten in der Szene beschreibt. In der Vorverarbeitungsphase wird eine Relationalstruktur aus Punkten (Koordinatenpaaren), Konturstücken und Bereichen erzeugt, ähnlich der RSE-Struktur im VISIONS System [Hanson Riseman 76], [Arbib Riseman 76]. Relationen wie 'Endpunkt' oder 'Anfangspunkt' fixieren die Konturen im Koordinatensystem. Eine 'Rand'-Relation ordnet jedem Bereich die ihn berandenden Konturen zu. Das Segmentationsverfahren ist in [Kraasch Zach 78] beschrieben.

Die Konturen sind als eine Folge von Schritten in die vier möglichen Richtungen entlang der positiven und negativen Koordinatenachsen kodiert. Sie werden durch Geradenstücke und Kreisbögen angenähert. Zwischen Bereichen gilt die Nachbarschaftsrelation, wenn sie gemeinsame Konturstücke als Rand haben. Ein Bereich ist in einem anderen enthalten, wenn beide kein gemeinsames Konturstück haben und der eine ganz im Inneren des anderen liegt.

Die so gewonnene symbolische Beschreibung - die "Skizze" - eines Bildes wird zu Objekten gruppiert. Sie sind durch Prototypen definiert, in denen angegeben ist, welche Typen von Skizzenelementen verwendet werden und welche Relationen zwischen ihnen gelten müssen. Solch ein Objekt direkt in der Skizze zu suchen, verbietet die kombinatorische Vielfalt. Deswegen werden Objekte durch Zusammensetzen aus einfacheren Objekten aufgebaut, die wiederum aus noch einfacheren Objekten aufgebaut sein können, solange bis die Ebene der Skizze erreicht ist.

Ein Interpretationsprozess für ein bestimmtes Objekt sucht nun zunächst in der Skizze nach Elementen, die Bausteine für die Strukturen in der nächst höheren Ebene sind. Dafür gibt es meistens viele Kandidaten, aber es müssen nur wenige Relationen überprüft werden. Aus den akzeptierten Skizzenelementen werden dann die Strukturen der nächst höheren Ebene konstruiert, wobei jetzt nur noch die neu hinzukommenden Relationen zwischen ihnen überprüft werden müssen. Das Verfahren schreitet bis zur obersten Ebene fort.

Ein Beispiel: Aus vier Geraden werden Rechtecke gebildet. Aus den Rechtecken werden helle und dunkle Quadrate aussortiert. Paare solcher Quadrate werden zu Reihen von acht Feldern gruppiert. Acht parallele Reihen bilden ein Schachbrett. Bei den Quadraten z.B. muss nicht mehr der rechte Winkel zwischen den Seiten verifiziert werden, das ist schon

auf der Ebene der Rechtecke geschehen, sondern es ist lediglich eine Überprüfung auf gleiche Länge der Seiten und auf Einhalten des Grauwertes des eingeschlossenen Bereiches notwendig.

Während es für syntaktische Verfahren schwierig ist, gestörte Strukturen zu analysieren, ist in diesem System die Fähigkeit dazu eingebaut. Man kann die Suche nach Strukturen in der Skizze, die denen im Prototyp entsprechen, als ein Graphisomorphie-Problem auffassen [Ambler et al. 75]. Definiert man nun einen Hilfsgraphen, dessen Knoten aus den hypothetischen Zuordnungen zwischen Skizzenstruktur und der Struktur des Prototypen auf einer bestimmten Ebene der Hierarchie bestehen und dessen Kanten ausdrücken, dass die im Prototyp vorgeschriebenen Relationen zwischen den Strukturen auch in der Skizze gelten, dann bilden die maximalen Cliques im Zuordnungsgraphen die bestmöglichen Inkarnationen des Prototypen in der Skizze. Cliques sind Teilgraphen, in denen jeder Knoten mit jedem anderen verbunden ist. Turner hat ein ähnliches Verfahren benutzt [Turner 74] und ebenfalls zur Suche nach Cliques eine Modifikation des Algorithmus von Bron und Kerbosch verwendet [Bron Kerbosch 73], der zur Zeit das effektivste bekannte Verfahren ist [Mulligan 77].

Da keine heuristische Einschränkung des Suchraums vorgenommen wird, werden alle Zuordnungen gefunden. Zu schlechte können über vorgebbare Schwellenwerte ausgesondert werden. Eine ausführliche Beschreibung des Interpretationsprozesses findet sich in [Bertelsmeier].

Die in der Skizze gefundenen Inkarnationen von Objekten werden auf Konsistenz mit einem vorgegebenen Szenenmodell geprüft, das am Anfang einer Bildserie noch sehr allgemeine Erwartungen über den Inhalt der Szene enthält. Es gibt erlaubte Konfigurationen von Objekten vor. Gesicherte Erkenntnisse aus den Interpretationen vergangener Bilder modifizieren das Szenenmodell. In ihm werden Listen derjenigen Objekteigenschaften und Objektbeziehungen geführt, deren zeitliches Verhalten interessiert. Nach den ersten Bildern einer Folge kann das Szenenmodell als prädiktives Modell dienen (vgl. [Chow Aggarwal 77]). Über das Szenenmodell können auch semantische Beziehungen zwischen den Objekten als Interpretationshilfen eingeführt werden, z.B. die Erfahrung, dass Objekte, die auf einem Schachbrett stehen, vorzugsweise Schachfiguren sind. Eine Auswertung der über die zeitliche Entwicklung geführten Listen kann zur Bildung von Bewegungskonzepten führen und so einen Anschluss an die Arbeiten von Badler [Badler 75] und Tsotsos [Tsotsos 76] ermöglichen.

Erste Ergebnisse liegen für die Interpretation einzelner, einfacher Bilder vor [Bertelsmeier]. Die Erfahrungen mit dem System lassen einige Schwächen erkennbar werden.

Der Interpretationsprozess, der aus den Skizzenelementen Objekte aufbaut, ist stark formalisiert, er erfordert eine syntaktisch strenge Formulierung der Prototypen (Bertelsmeier stellt zu ihrer Formulierung eine Syntax und einen Übersetzer bereit). Es werden alle im aktuellen Szenenmodell vorhandenen Objekte gesucht, wobei alle möglichen Zuordnungen von Skizzenelementen zu Objekten parallel verfolgt werden. Da aber gestörte Objekte erkannt werden sollen, muss eine Konfidenz der Zuordnung auf jeder Ebene der Hierarchie berechnet werden, um zu schlechte Hypothesen aufgeben zu können. Der Wert für die Konfidenz gründet sich nur auf die Merkmale und Relationen von Skizzenelementen, die aus den darunterliegenden Ebenen aufgestiegen sind. Wünschenswert wäre, vor dem Fallenlassen der Hypothese berücksichtigen zu können, ob es überhaupt bessere Zuordnungshypothesen für die beteiligten Skizzenelemente gibt und ob nicht durch den auf höheren Ebenen grösseren Kontext die fragliche Hypothese wieder gestützt wird.

Überleben z.B. bei den Feldern eines Schachbrettes zu wenig Hypothesen für Quadrate, weil die Schachfiguren die Feldgrenzen verdecken, so wird eventuell die Hypothese für das Schachbrett verworfen.

Je niedriger aber die Konfidenzschwelle gelegt wird, um so mehr Zuordnungen werden erzeugt und um so stärker wird die Bildanalyse damit befrachtet, die Konsistenz der Objekthypothesen mit dem Szenenmodell zu überprüfen. Das Szenenmodell wiederum muss dann immer stärkere Zwangsbedingungen enthalten, um noch eine genügende Filterwirkung zu besitzen.

Die Prototypen sind syntaktisch vorgegeben. Daher müssen sie so formuliert werden, dass sie gegen erwartete Störungen unempfindlich sind. Eine Art von erwarteter Störung ist die gegenseitige Verdeckung von Objekten. Sie wird im augenblicklichen System nur soweit toleriert, bis die Konfidenzschwelle für die Objekthypothese unterschritten ist. Danach geht das Objekt verloren. Ein Ausweg wäre für den Fall, dass das Objekt erst einmal vollständig sichtbar war, bei Eintreten der Verdeckung einen temporären Prototypen für das Objekt abzuleiten, der die unsichtbaren Teile nicht enthält (vergl. [Chow Aggarwal 77]).

Ähnliche Massnahmen könnten für den Fall getroffen werden, dass sich ein Objekt auf Grund von perspektivischer Verzerrung verändert. Um beim Beispiel des Schachbrettes zu bleiben, sei angenommen, dass es sich aus einer Achsensymmetrie zum Beobachter herausdreht. Dabei verformen sich die Rechtecke der

Felder zu Rhomben. Dann wird von einem bestimmten Drehwinkel an die Toleranz für den bei Quadraten geforderten rechten Winkel zwischen den Seiten überschritten; einer wesentlich auf quadratischen Erscheinungsformen beruhenden Schachbrett-Hypothese wird die Grundlage entzogen. Hier gilt ein analoger Einwand wie für die Häufungsanalyse-Verfahren. Wenn Toleranzen oder Entfernungen im Merkmalsraum nicht durch zufällige sondern durch systematische Variation überschritten werden, sollte ein Zuordnungssystem das erkennen und durch Anpassung des prädiktiven Teiles im Szenenmodell darauf reagieren.

Eine gewisse Stabilität gegenüber Grauwertschwankungen infolge von Reflexions- und Beleuchtungsänderungen, Schatten und Rauschen sollte durch einen starken Bezug auf die Formbeschreibung von Objekten erreicht werden. Leider hat sich die Bereitstellung der Formbeschreibung als schwieriges Problem herausgestellt. Meine Erfahrungen bei der Implementation der dazu notwendigen Algorithmen [Bertelsmeier Radig 77a] sind mit ein Grund für die in den nächsten Kapiteln beschriebenen Untersuchungen zur Umkodierung und Segmentierung von Grauwertbildern.

Zur Erläuterung muss ich noch einmal auf die mögliche partielle Zuordnung zwischen Skizze und Objektprototyp zurückkommen. Toleriert wird unter anderem das Fehlen von Skizzenelementen, z.B. eines Geradenstückes als Seite für ein Rechteck, nicht jedoch das Vorhandensein überzähliger Elemente, z.B. wenn eine Seite aus einem stumpfen Winkel zweier Geraden gebildet wird (Fig. 2.1a). Dann werden die beiden Rechtecke gefunden, die aus den drei ganzen Seiten und je einem Schenkel des Winkels bestehen (Fig. 2.1b,c).

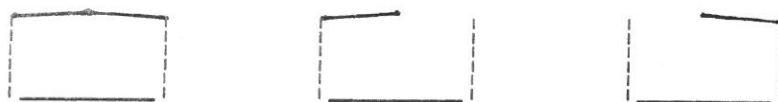


Fig. 2.1

a)

b)

c)

Das System erkennt die Situation nicht, da Zuordnungswidersprüche erst im Szenenmodell aufgelöst werden, und kann auch den Skizzengenerator [Bertelsmeier Radig 77a] nicht veranlassen, eine alternative Beschreibung des stumpfen Winkels durch eine Gerade zu versuchen. Der Grund für die unzuweckmässige Konturapproximation liegt darin, dass die Algorithmen für die Formbeschreibung nicht mit dem Teil des Systems wechselwirken, der die Konturen aus dem Bild

extrahiert. Die endgültige Festlegung des Konturverlaufes wird nicht beeinflusst durch den Kontext einer Formbeschreibung, geschweige denn durch den Kontext einer Einbettung in eine Objektinkarnation.

Ein Problem für die Formbeschreibung liegt auch in der ungünstigen Darstellung der Konturen durch den verwendeten Segmentationsalgorithmus [Kraasch Zach 78]; er beschreibt Konturstücke als eine Folge von horizontalen und vertikalen Kantenelementen. Voraussetzung für die Formfindung ist eine Beschreibung als eine Folge von Richtungsvektoren aus einer mehr oder minder feinen Quantelung des Winkelbereiches. Um die Vektoren zu berechnen, ist eine Mittelung über Kantenelemente notwendig. Turner benutzt z.B. Stücke der Länge von sieben Elementen [Turner 74]. Der Nachteil einer festen Länge ist ein Abrunden der Ecken. In ersten Versuchen benutzte ich als einem Kantenelement zugeordneten Richtungsvektor die Richtung der längsten Geraden, die alle benachbarten Treppenstufen zweimal schneidet (Fig. 2.2).

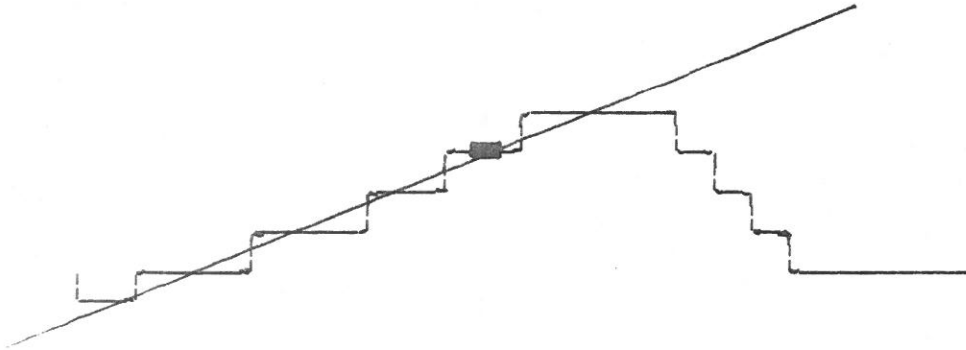


Fig. 2.2.

In der jetzigen Version ist das zu aufwendige Verfahren ersetzt durch eines, das eine syntaktische Analyse der benachbarten Kantenelemente benutzt, um herauszufinden, mit welcher Länge eine Mittelung vorgenommen werden kann. Anregungen gaben Arbeiten über eine linguistische Beschreibung von Geraden auf einem Gitter [Broner 74], [Rothstein Weiman 76], [Gaafar 77]. Die Analyse benutzt einen Entscheidungsbaum. Fig. 2.3a) zeigt einige akzeptierte, 2.3b) einige zurückgewiesene Folgen. Jede Folge der Länge drei wird akzeptiert.



Fig. 2.3a



Fig. 2.3b

All diese Anstrengungen sind nötig, um die Annäherung einer stetigen Kontur (des realen Objektes) durch eine Darstellung auf einem quadratischen Raster wieder rückgängig zu machen. Deswegen habe ich bei dem neuen Ansatz darauf geachtet, nicht in einem Verarbeitungsschritt Informationen zu verlieren, die später mühsam wieder restauriert werden müssen. In diesem Fall war wichtige Richtungsinformation durch das Segmentationsverfahren verloren gegangen. Die Richtungsinformation hätte wahrscheinlich günstiger durch Anwenden eines geeigneten Operators auf die Grauwertmatrix (z.B. [Hueckel 71]) als durch das beschriebene Verfahren gewonnen werden können.

Ein ähnliches Problem ergibt sich aus der Entkoppelung von Segmentation und Interpretation für die Zerlegung des Bildes in Bereiche. Da Zahl und Anordnung von Skizzenelementen, also auch von Bereichen, in den Objektprototypen fest vorgegeben sind, führt die Aufspaltung der Bereichsbeschreibung eines realen Objektes in eine abweichende Anzahl von Bereichen zu einer schlechten Bewertung der Zuordnung. Da auch die Formbeschreibung des Randes der zugeordneten Bereiche nicht mehr stimmt, sinkt auch dafür die Konfidenz. Aus diesem Grund musste in unserem früheren Ansatz ein "Superfleck" als Skizzenelement eingeführt werden, der eine Kollektion von benachbarten Bereichen vertritt. Wie jedoch die Kombination von Bereichen wirksam durch eine Heuristik unterstützt werden kann, ist noch unklar.

Der gewählte syntaktische Ansatz zur Beschreibung der Objekte und der mangelnde Informationsfluss zwischen den Stufen des Systems erlauben nur die Analyse sehr eingeschränkter Vorlagen. Ich habe die Erfahrungen so eingehend geschildert, weil sie mit ein Anlass waren, ein neues System zu konzipieren, das von den Grauwertmatrizen der Bildfolge bis zur Szenenbeschreibung eine einheitliche und durchlässige Struktur benutzt, und bei dem Erkenntnisse auf einem höheren Niveau Entscheidungen der darunterliegenden Ebenen revidieren können. Natürlich sollte

man es vermeiden, noch nicht genügend abgesicherte Entscheidungen zu treffen ("principle of least commitment", [Marr 75]); bei implementierten Systemen verhindern jedoch oftmals Beschränkungen von Speicherplatz und Laufzeit das Aufbewahren und Verfolgen vieler Hypothesen.

3. Problem- und Lösungsrahmen

Das Ziel der vorliegenden Untersuchungen besteht darin, ein Verfahren zu entwickeln und auf einer Datenverarbeitungsanlage zu erproben, das aus der Beobachtung einer Szene stammende Daten auswertet. Die Daten sind unpräparierte Digitisierungen von Fernsehaufnahmen. Die Vorgänge in der Szene werden in einer Bildfolge festgehalten. Das im Rechner ablaufende Programm soll eine Beschreibung der Vorgänge ausgeben. Um das Ziel auch nur in die Nähe einer Realisierbarkeit zu rücken, muss die Bedeutung der im vorigen Satz verwandten Begriffe sehr eingeschränkt werden.

3.1 Beschränkungen

Die Randbedingungen für das Programm sind noch am leichtesten anzugeben. Die am Fachbereich für Informatik vorhandenen Rechner diktierten Speichervolumen und Peripherie sowie das Menü der Programmiersprachen. Die Laufzeit des Programmes sollte nicht mehr als einige Minuten pro Bild betragen; eine Serie von 10 bis 30 Bildern überschreitet sonst schnell das zur Verfügung stehende Zeitkontingent.

Bei der Auslegung der Digitisierung gilt natürlich auch das Diktat des Vorhandenen, aber hier war doch der Planungsspielraum grösser. Zur Beobachtung der Szene wurde lediglich eine Schwarzweiss-Fernsehkamera verwendet. Zusätzliche Farbinformation würde es zwar erleichtern, eine komplexere Szene auszuwerten, trägt aber wenig zur prinzipiellen Lösung des Problems bei. Ähnlich sind die Verhältnisse beim Hinzuziehen von Messwerten, die durch Entfernungsmesser oder eine zweite Kamera gewonnen werden können. Dabei tritt noch das Problem auf, Daten aus verschiedenen Quellen miteinander in Beziehung zu setzen. Ich bin der Meinung, dass die Analyse von Grautonbildern genügend tiefen Einblick in die prinzipielle Schwierigkeit der Aufgabenstellung gestattet. Gleichzeitig sollte man aber im Auge behalten, dass für konkrete Anwendungsfälle die Auswahl der Sensoren pragmatisch zu erfolgen hat. Die Verwendung einer Fernsehkamera bietet den Vorteil des einfachen Experimentierens, da die Grauwerte unmittelbar in einer zur Digitisierung geeigneten Form bereitstehen. Die erreichbare Präzision eines Filmabtasters ist nicht vonnöten;

es wäre auch die Zwischenstufe der Filmentwicklung notwendig, und die Konversionsdauer ist relativ hoch.

Ist die Art der Daten als Messung der Lichtintensität mittels einer Fernsehkamera festgelegt, bleibt noch ihr Volumen zu bestimmen. Die erreichbare Ortsauflösung ist durch die Zeilenzahl und die Bandbreite des Videosignals nach oben begrenzt, die Genauigkeit der einzelnen Messwerte durch den Störabstand der Signalquelle. Wir haben uns bei der Auslegung der Apparatur an der oberen Grenze orientiert und tasten 300000 Punkte mit 256 Graustufen ab (s. Kap. 4.1). Die zeitliche Auflösung ist durch die Fernsehnorm mit 25 Bildern pro Sekunde gegeben. Da ein Bild noch nicht in 40msec auf einen digitalen Hintergrundspeicher ausgelagert werden kann, wird zur Zwischenspeicherung eine Analog-Magnetplatte mit der Fähigkeit zur Einzelbildwiedergabe eingesetzt.

Damit seien die Randbedingungen für das Programm und die ihm zur Analyse vorgelegten Daten ausreichend beschrieben; es bleibt die Frage, welche Szenen wir beobachten und welche Vorgänge wir zulassen wollen. Beides steht in engem Zusammenhang, denn die Auswahl der Akteure bestimmt die Art der zu erwartenden Aktionen. Für diese Arbeit hatte ich mir zum Ziel gesetzt, Ausschnitte aus der normalen Umwelt des Menschen aufzunehmen. Die beiden einfachsten Möglichkeiten waren, die Kamera im Laboratorium aufzustellen und die dort umhergehenden Personen zu beobachten (Bild 3.1) oder die Kamera aus dem Laborfenster auf den Strassenverkehr zu richten (Bild 3.2). Dabei wurden einige Einschränkungen beachtet (vergl. [Nagel 77]).

- Die Ausleuchtung der Szene war so beschaffen, dass ein Videosignal mit gutem Signal/Rausch-Verhältnis zustande kommen konnte.
- Die Beleuchtung sollte diffus sein, um kräftige, harte Schatten zu vermeiden. Zwar wurde das Problem der Schatten in der Literatur schon behandelt [Waltz 72], [Turner 74], [Horn 74,75,77], die Ergebnisse lassen sich jedoch nicht leicht auf natürliche Szenen übertragen.
- Eine schnelle Änderung der Lichtverhältnisse war ebenfalls nicht zugelassen, insbesondere wurden während der Beobachtungszeit keine künstlichen Lichtquellen ein- oder ausgeschaltet.

Bei der gewählten Aufnahmetechnik handelt es sich um die optische Projektion einer dreidimensionalen Szene in ein zweidimensionales Bild auf der Sensorfläche der Kamera. Um hier eine zu grosse Komplexität zu vermeiden, mussten weitere

Einschränkungen hingenommen werden.

- Die Gegenstände dürfen ihre Grösse (in der Projektion) nicht rapide ändern.
- Es ist zwar zugelassen, dass sich Objekte verdecken, aber sie sollten undurchsichtig sein und keine spiegelnden Teile enthalten.
- Die sich bewegenden Gegenstände sollten dem Beobachter im wesentlichen dieselbe oder eine sich nur langsam ändernde Ansicht zeigen, um ein explizites Verwenden von dreidimensionalen Modellen für einzelne Objekte vermeiden zu können.

An die Gegenstände selbst werden noch folgende Anforderungen gestellt:

- Sie müssen so gross oder so nah sein, dass ihr Abbild nicht an der Auflösungsgrenze des Rasters verschwindet.
- Andererseits sollte ein bewegtes Objekt nicht das Bildfenster vollständig ausfüllen.

Die Abtrennung von bewegten Objekten relativ zu einem stationären Hintergrund wird erleichtert,

- wenn die Kameraposition während des Beobachtungszeitraumes fest bleibt.

Eine durch die Aufnahmetechnik vorgegebene Bedingung ist,

- dass sich das Abbild eines Objektes von Bild zu Bild um nicht mehr als einige Prozent der Bilddiagonalen verschiebt, sonst werden seine Konturen zu unscharf.

3.2 Ziel

Die Semantik einer Szene und die in ihr ablaufende Handlung zu beschreiben - etwa wie es den Vorstellungen von Badler entspricht [Badler 75] - setzt voraus, ein Modell der Gegenstände in der Szene aus den Messwerten zu rekonstruieren. Mit Modell meine ich hier wieder die Inkarnation eines prototypisch beschriebenen Objektes.

Gegenstände sind durch die Strukturierung ihres Aufbaus und durch ihre Beziehungen zueinander definiert; besonders die gegenseitigen Beziehungen verdeutlichen ihre Funktion und Aktion. Die Handlung in einer Szene drückt sich darin aus, dass sich Beziehungen zwischen Gegenständen ändern. Mit Gegenstand

meine ich nicht nur semantisch selbständige Einheiten wie Baum oder Fussgänger, sondern auch Teile davon wie Blatt oder Unterarm. Auch Konzepte wie Licht oder Schwerkraft können an Beziehungen beteiligt sein. Die Beziehung zwischen einem grünen Blatt und dem Licht ist eben, dass das Blatt einen Teil des Spektrums besonders gut reflektiert. Der Schattenwurf ist ein wichtiges, nichtgegenständliches Konzept in natürlichen Szenen.

Verallgemeinert besteht eine Szene aus einer Kollektion von Objekten, die miteinander in Beziehung treten und dadurch einen Teil ihrer Bedeutung gewinnen und dann auch gewöhnlich mit einem Namen belegt werden, der ihrer Bedeutung in der Szene entspricht.

Akin und Reddy führten ein sehr schönes Experiment durch [Akin Reddy 77], in dem ein Rechner protokolliert, wie Versuchspersonen bei der Analyse von Bildern versuchen, durch Herausfinden von Eigenschaften einzelner Bildteile sowie von Beziehungen zwischen Bildteilen die Bedeutung der abgebildeten Objekte zu erkennen.

Zwei Dialogpartner befinden sich in getrennten Räumen und können über Fernschreiber und Rechner miteinander kommunizieren. Der eine - der Sensor - hat ein Bild vor sich liegen und beantwortet Fragen des zweiten - des Analysators - über das Bild. Fragen, Antworten und Kommentare des Fragenden werden vom Rechner automatisch protokolliert. Normalerweise versucht der Analysator zuerst, sich einen groben Überblick über den Aufbau des Bildes zu verschaffen, um ganze Gruppen von Hypothesen ausschalten zu können. Er fragt z.B., ob das Bild in seinem oberen Teil blau ist, um die Vermutung einer Landschaftsaufnahme zu prüfen. Die Hypothese einer Büroszene könnte die Frage nach sich ziehen, ob viele geometrisch regelmäßige Strukturen im Bild sind. Hat der Analysator eine befriedigende Sicherheit bezüglich des Bildthemas gewonnen, versucht er mit Hilfe von Fragen nach Farben, Formen und räumlicher Lage von Bildteilen einen Aufschluss über vorhandene Gegenstände zu gewinnen. Dabei gibt er sich durchaus mit unpräzisen Angaben des Sensors zufrieden (der das Bild nicht quantitativ auswertet); der Analysator versucht eher aus mehreren Beziehungen die notwendige Sicherheit der Deutung zu gewinnen. Bei der Auswahl der Fragen lässt er sich davon leiten, wie sich ein Charakteristikum des von ihm vermuteten Objektes wohl im Bild darstellen könnte.

Das Ziel meines Ansatzes ist es, Voraussetzungen zu schaffen, die eine Automatisierung dieses Analyseprozesses bei Bildfolgen erleichtern. Dabei beschreibt das eben skizzierte Experiment nicht die Methodik sondern eher die gewünschte Leistung des Programmes. Ich habe mich in dieser Arbeit darauf konzentriert, zuverlässige und vielseitige, elementare Beschreibungsmittel bereitzustellen, die jedoch in eine formale Struktur eingebettet sind, um so eine kontinuierliche Erweiterung auf abstraktere Beschreibungen zuzulassen (vergl. die "general purpose models" [Zucker et al. 75]). Eine wesentliche Abstraktion beim Analysieren von Bildfolgen ist das Suchen von Korrespondenzen zwischen Bildteilen. Weiss man in einer Serie von Aufnahmen, welche Konfiguration von Grauwerten durch ein und dieselben realen, ruhenden oder bewegten Gegenstände erzeugt wurden, so hat man den entscheidenden Schlüssel für die Beschreibung der Handlung in der Bildfolge gefunden. Deswegen habe ich meine Beschreibungsmittel daraufhin geprüft, ob sie das Finden von Korrespondenzen unterstützen.

Die hier gewählte Vorgehensweise, Korrespondenzen zwischen Objekten auf einer mittleren Abstraktionsebene herzustellen, d.h. weder Grauwertmatrizen noch Interpretationsergebnisse einzelner Bilder zu vergleichen, wird bestärkt durch die Ergebnisse der Arbeit von Ullman [Ullman 77]. Er hat nachgewiesen, dass beim Erkennen von bewegten Gegenständen der menschliche Gesichtssinn den Vergleich relativ einfacher Strukturen benutzt, um einen Gegenstand zu verfolgen und dass dieser Prozess einer semantischen Interpretation des Gegenstandes vorgeschaltet ist. Ullman hat gezeigt, dass man unter Beachtung ähnlicher Voraussetzungen, wie ich sie im vorigen Abschnitt angegeben habe, die dreidimensionale Struktur einer Szene mit bewegten Objekten erschliessen kann, ohne die Bedeutung der Szene kennen zu müssen.

Die Auswahl der Szenen, die im folgenden als Beispiel dienen, spiegelt nicht die Absicht einer bestimmten Anwendung wider; sie dienen vielmehr zur Illustration der von mir entworfenen Operatoren auf Grauwertbildern. Vorliegende Arbeit endet dort, wo eine konkrete Spezialisierung die Verwendung von darauf abgestimmten Beschreibungsmitteln erfordern würde.

3.3 Lösungsstrategie

Die einzige Frage, die der automatische Analysator an seinen Sensor stellen kann, ist:

"Welche Lichtintensität z wurde zur Zeit t an dem Ort (x,y) des projizierten Bildes gemessen?"

Diese Messwerte sind die Ausgangsobjekte, aus denen er alle anderen Beziehungen ableiten muss, die Inhalte und Handlung der Szene beschreiben. Welche Relationen das sind, hängt von dem erwarteten Inhalt der Szene ab und von der Kenntnis, wie die Beziehungen, die die realen Gegenstände definieren, in den sensorischen Daten repräsentiert sind.

Der Vorgang der automatischen Analyse spielt sich etwa so ab: Zwischen den Ausgangsobjekten werden Relationen berechnet. Dadurch sind neue Objekte definiert, zwischen denen wiederum Relationen berechnet werden können, die wiederum Objekte definieren. Die Korrespondenz zwischen diesen Relationen und Objekten zu den realen Beziehungen und Gegenständen wird über ein "Weltmodell" hergestellt. In ihm ist damit festgelegt, welchen Szeneninhalte mit welchem Handlungsablauf der Analysator beschreiben kann. Ausserdem können die im Weltmodell bekannten Gesetzmässigkeiten für Gegenstände und Aktionen in der realen Szene den Analyseprozess steuern, so dass nur sinnvolle Relationen berechnet und damit sinnvolle Objekte generiert werden.

Ein Vergleich mit der in Kapitel 2.1 geschilderten Arbeitsweise eines kognitiven Systems ordnet die Messwerte den Objekten der Ausgangssituation zu. Die Transformationen nehmen die Prozeduren vor, die die Relationen berechnen und dadurch neue Objekte erzeugen. Das Weltmodell gestaltet die Transformationssequenz.

Eine ähnliche Anschauung des Analysevorganges findet sich im Konzept der "recognition cones" von Uhr [Uhr 73b,75b], [Uhr Douglass 77]. Uhr entwickelt die Vorstellung einer Pyramide von übereinanderliegenden Ebenen von Gedächtniselementen. In die unterste Schicht mit der höchsten Ortsauflösung werden die sensorischen Daten eingegeben. Zwischen den Ebenen sind Transformationen definiert, von denen eine jede Daten aus Zellen der unter ihnen liegenden Schicht verarbeitet und ihr Ergebnis in einer Zelle der darüberliegenden Schicht ablegt. An der Spitze der Pyramide findet sich eine symbolische Beschreibung der ganzen Szene. Uhr hat den Zellen ein

Erinnerungsvermögen gegeben, ihr Inhalt bleibt bei der Analyse eines neuen Bildes (abgeschwächt) erhalten. Der Mechanismus soll ihm ermöglichen, die Ergebnisse aus einer zeitlich geordneten Folge miteinander in Beziehung zu setzen [Uhr 73a,75b,c].

Die Idee ist in meinem System dadurch konkretisiert, dass es erlaubt ist, objektdefinierende Relationen zu berechnen, deren Träger-elemente aus verschiedenen "Zeitscheiben" stammen dürfen. Ist z.B. ein Gegenstand während der ganzen Beobachtungszeit teilweise verdeckt, sind jedoch alle Teile irgendwann einmal sichtbar, so kann das Modell des Gegenstandes als Inkarnation eines Prototypen aus Elementen zusammengesetzt werden, die verschiedenen Bildern entstammen. Ein einfaches Beispiel für eine objektbeschreibende Relation, die sowohl Relationen aus einem Bild als auch solche aus verschiedenen Bildern der Folge benutzt, ist die Berechnung des über die Beobachtungszeit gemittelten Schwerpunktes eines Gegenstandes. Er sei in den Grauwertmatrizen repräsentiert durch Bereiche, die seiner Oberfläche zugeordnet wurden. In jedem Bild kann man den gemeinsamen Schwerpunkt dieser Bereiche berechnen; die Resultate werden über die Folge gemittelt. Denselben Wert erhält man jedoch auch, wenn einfach der mittlere Schwerpunkt aller, einem Gegenstand zugeordneten Bereiche gebildet wird, gleichgültig aus welchem Bild sie stammen und ob die Repräsentation des Gegenstandes in allen Bildern vorhanden ist.

Der Ablauf des Analysevorganges kann durch das Weltmodell nur unvollständig festgelegt werden. Zwei Gründe sind dafür massgebend. Der eine ist, dass die Messwerte statistisch gestört, mit Fehlern behaftet sind. Der zweite ist, dass Näherungen verwendet werden müssen, um die Beziehungen in der realen Welt mit den Messwerten in Verbindung zu bringen, sei es, dass beim Systementwurf nicht genau bekannt ist, wie sich die Beziehungen in den sensorischen Daten ausdrücken, sei es, dass es nicht möglich ist, in einer Datenverarbeitungsanlage in endlicher Zeit alle Beziehungen zu berechnen, die zum Beispiel ein Mensch bewusst oder unbewusst benutzt, um seine Umwelt zu interpretieren. In dem Experiment von Akin und Reddy gelang es manchen Versuchspersonen nicht, das ihrem Partner vorliegende Bild korrekt zu beschreiben.

Unser automatischer Analysator muss mit fehlerhafter und unvollständiger Information in allen Stufen der Interpretation auskommen. Uhr hält in seinem Artikel [Uhr 75a] ein Plädoyer für solche Systeme. Die Theorie des Umganges mit vagem Wissen wurde in diesem Jahrzehnt entwickelt [Zadeh 65,68,72] und hat eine stürmische Entwicklung genommen, es existieren schon über

tausend Artikel zu diesem Gebiet. Eine PLANNER-ähnliche Sprache [Hewitt 69,71,72] zur Manipulation und Deduktion solchen Wissens wurde von LeFaivre [LeFaivre 74,76] geschaffen, sie basiert auf LISP.

Falsche Schlussfolgerungen aus gestörter Information kann der Analysator nur dadurch vermeiden, dass er die Konsistenz der Schlüsse auf Widersprüche zu der im Weltmodell vorhandenen Kenntnis der Gesetzmässigkeiten und Bedingungen über die reale Szene prüft. Ein Beispiel, auf das ich später noch zurückkomme, mag das verdeutlichen.

Wenn in einem Bild der Oberfläche eines Gegenstandes ein bestimmter, durch einen Segmentationsalgorithmus berechneter Ausschnitt zugeordnet ist und in dem darauffolgenden Bild der derselben Oberfläche zugeordnete Ausschnitt nur noch halb so gross ist, dann ist das ein offensichtlicher Widerspruch zu der im Abschnitt 3.1 angegebenen Einschränkung, dass sich die Ansicht eines Gegenstandes nur langsam ändern soll.

Der Analysator sollte in der Lage sein, einen solchen Widerspruch zu entdecken und aufzuklären. Ein mögliches Vorgehen beschreibe ich in Kapitel 6.1. Es gibt merkwürdigerweise keine Systeme, die Bilder gleichzeitig mit verschiedenen Segmentationsverfahren bearbeiten (vergl. [Riseman Arbib 77, p.272,3]), obwohl es dafür eine grosse Zahl von geeigneten Algorithmen gibt, die unterschiedliche Schwächen und Vorzüge haben. Dabei ist es doch nur von Vorteil, wenn man alternative Wege benutzt, um eine Aussage über die Szene zu machen für den Fall, dass Resultate widersprüchlich oder zu unsicher werden. Das kombinierte Ergebnis zweier guter Segmentationsverfahren ist vielleicht besser und billiger zu erhalten als das eines sehr guten, aber in Grenzfällen immer noch fehlerhaften Algorithmus. Ein Fortschritt bezüglich der Berechnung von Objekteigenschaften in der Arbeit von Bolles (s. Kap. 2.2.2, [Bolles 76]) ist es, dass er mehrere Verfahren kritisch anwenden kann, um ein Ergebnis zu erzielen. Ich gehe in Kapitel 6.1 auf alternative Segmentationen ein.

Die Frage taucht auf: Wann Aussagen sicher genug sind, um einen Widerspruch mit dem Weltmodell feststellen beziehungsweise ausschliessen zu können? Generell lässt sich darauf keine Antwort geben, aber man kann die Frage umformulieren, um einen Ansatz für eine bessere Einsicht zu bekommen. Wann ist es sinnvoll, eine Aussage genau zu berechnen oder einen Widerspruch festzustellen? Noch anders formuliert: Wie hoch sind die Kosten, d.h. Rechenzeit, Speicherplatz und andere, um eine Aussage zu berechnen und was nützt die Aussage im Hinblick auf ein vorgegebenes

Interpretationsziel? Nur vom Zweck der Szenenbeschreibung kann abgeleitet werden, mit welcher Genauigkeit welche Beziehungen berechnet werden sollen, ob also die Erhöhung der Konfidenz für eine Relation von 95% auf 98% sinnvoll ist oder nicht. Und nur aus der Kenntnis der Implementation ist bestimmbar, wie hoch der Berechnungs- und Zugriffsaufwand für eine bestimmte Relation ist. Mit beiden Faktoren zusammen lässt sich lenken, welche Aussagen in welcher Reihenfolge berechnet werden, nämlich die zuerst, die den grössten Nutzen bei geringsten Kosten versprechen.

Ein System, das allein den Kostenfaktor benutzt, um den Aufwand zur Berechnung von Aussagen zu begrenzen, wurde von Yakimovsky und Cunningham entwickelt [Yakimovsky Cunningham 76b]. Shortliffe schlägt als zukünftige Verbesserung seines MYCIN-Systems unter anderem vor, die Abarbeitung der im System enthaltenen Regeln dynamisch zu ordnen, gelenkt durch Kosten und Nutzen [Shortliffe 76]. MYCIN ist ein umfangreiches Produktionen-System zur Unterstützung der medizinischen Diagnose und Therapie bei bakteriellen Infektionen. Es führt einen Dialog mit dem Benutzer über Symptome und Befunde des Krankheitsbildes und weist Behandlungsmöglichkeiten nach (s.a. [Davis et al. 77]). Shortliffe schlägt vor, wenn eine Wahl möglich ist, zuerst die Regeln anzuwenden, die wenige Bedingungen in ihrer Prämisse enthalten und die Bedingungen in der Reihenfolge zu erforschen, dass zuerst solche benutzt werden, deren Wahrheitswert schon im System bekannt ist. Falls schon eine von denen nicht erfüllt ist, kann die Untersuchung der restlichen, konjunktiv verknüpften Bedingungen unterbleiben. Ausserdem will er aktuell anwendbare Regeln danach sortieren, welche Evidenz sie, falls die Prämisse erfüllt ist, dem gerade untersuchten Faktum hinzufügen. (Ich kann mir vorstellen, dass die zur Beschreibung von programmierten stochastischen Grammatiken entwickelten Formalismen, eventuell mit dynamischer Berechnung der mit den Regeln assoziierten Wahrscheinlichkeiten, einen eleganten Lösungsansatz ermöglichen könnten [Huang Fu 72]).

Über eine Veränderung des Kosten / Nutzen Verhältnisses lässt sich sozusagen die Aufmerksamkeit des Systems auf wichtige Aufgaben konzentrieren. Ist zum Beispiel das Ziel der Analyse, nur die Trajektorien von bewegten Objekten zu beschreiben, und ist das System in der Lage, Ausschnitte in den Bildern der Folge zu markieren, in denen nur ruhende Gegenstände abgebildet sind, etwa durch den Algorithmus von Jain und Nagel [Jain Nagel 77], so können diese Bildausschnitte mit dem Nutzen Null initialisiert werden. Falls sich im Laufe der Untersuchung der

Nutzen nicht erhöht, werden die Bildteile nicht untersucht. Ein Anlass, den Nutzen zu erhöhen, wäre die Berechnung einer Relation, die einen ruhenden Gegenstand einschliesst. Soll überprüft werden, ob sich das bewegte Objekt 'Fussgänger' auf der Strasse befindet, wird der Nutzen für die Bildteile in der Umgebung des Fussgängers erhöht.

Die Kosten zur Berechnung einer Relation können sich ebenfalls im Laufe der Analyse ändern. Die Trägerelemente der Relation sind ja wiederum Relationen, die natürlich berechnet sein müssen, bevor die Relation berechnet werden kann. Je mehr von den Teilrelationen schon berechnet sind, umso geringer werden die Kosten. Ist die Relation einmal berechnet und wird sie aufbewahrt, so bleibt als endgültiger Kostenfaktor bei einer erneuten Berechnung nur der Anteil für den Speicherzugriff (vergl. das Konzept der "memo functions" [Michie 68]). Der Wert für den Kostenfaktor kann aus Messungen des Zeitbedarfs für die eine Relation berechnende Prozedur und dem benötigten Speicherplatz für die Ablage des Resultats ermittelt werden. Zum Schluss sei noch darauf hingewiesen, dass sich die Kosten- und Nutzen-Faktoren auch verwenden lassen, um das Auslagern von berechneten Relationen auf Hintergrundspeicher zu lenken.

3.4 Relationalstruktur

Ausgehend von der Anschauung, dass Gegenstände ihre Bedeutung zu einem Teil aus ihren Beziehungen zur Umgebung gewinnen und sich die Aktion von Gegenständen in der Szene durch die Änderung dieser Beziehungen ausdrückt, wurde als formaler Rahmen zur Beschreibung der Datenstruktur eine Relationalstruktur gewählt.

a) Relation

Eine N-stellige Relation R ist eine Teilmenge des karthesischen Produktes von Mengen

$$T_i, \quad i = 1, N; \text{ den Trägern}$$

$$R \subset T_1 \times T_2 \times \dots \times T_N$$

Ein Element der Relation lässt sich als N-stelliges Tupel schreiben:

$$r = (t_1, t_2, \dots, t_N) \quad t_i \in T_i$$

Eine Relation lässt sich durch Aufzählen ihrer Tupel

$$R = \{ r_1, r_2, \dots, r_q \} \quad |R| = q$$

oder durch eine charakteristische Funktion angeben

$$f_R(t_1, t_2, \dots, t_N) = \begin{cases} 1 \Rightarrow (t_1, t_2, \dots, t_N) \in R \\ 0 \Rightarrow (t_1, t_2, \dots, t_N) \notin R \end{cases}$$

Will man unsicheres Wissen mit Hilfe von Relationen darstellen, so kann man der charakteristischen Funktion einen kontinuierlichen Wertebereich geben, der dann für ein Tupel die Konfidenz beschreibt, mit der es zu R gehört. Die charakteristische Funktion kann z.B. eine aus statistischen Messungen gewonnene Wahrscheinlichkeit für das Auftreten des Argument-Tupels beschreiben. Wenn es sich als zweckmässig erweist, kann der Konfidenzwert zusammen mit dem Tupel angegeben werden; eine "unscharfe" ("fuzzy") Relation besteht aus der Menge der Paare:

$$R = \{ (r, k)_1, (r, k)_2, \dots, (r, k)_N \}$$

$$(r, k) = ((t_1, t_2, \dots, t_N), k)$$

für jedes Element $r \in R$.

Die Darstellung ist analog zu der in "FUZZY" [LeFaivre 74,76].

Wahlster [Wahlster 77] hat einige deutsche Bezeichnungen für "fuzzy set" gesammelt: 'F-Menge', 'gefaserte Menge', 'unscharfe Menge', 'gefranste Menge' und 'verschwommene Menge'. Ich werde im folgenden das Präfix 'F' verwenden, um auf die "fuzzy" Natur der Mengen, Relationen oder Operationen ausdrücklich hinzuweisen.

b) Hyperrelationen

Gegeben sei eine Basismenge B und eine Menge von Relationen H . Gilt für jede Relation R aus H , dass ihre Träger aus B oder aus H stammen, so heisst R eine Hyperrelation und H eine Hyperrelationalstruktur. Sei TR die Menge der Träger und $T_i R \in TR$ ein Träger der Relation R , dann gilt

$$\bigwedge_{R \in H} T_i R \in B \text{ oder } T_i R \in H, \text{ d.h. } TR \subset (H \cup B)$$

Zur Kennzeichnung, dass eine Relation $R_i \in H$ Träger einer Relation $R_j \in H$ ist, schreiben wir

$$R_i < R_j$$

Explizit sei $i = j$ erlaubt, dann ist die Relation selbst einer ihrer Träger. Dieser formale Trick erlaubt es, Relationen zu definieren, deren Stellenzahl nicht genau angegeben sein muss, etwa

KONTUR C KONTUR x KANTENELEMENT

Um die Berechenbarkeit der Relationen zu sichern, wollen wir " $<$ " zu einer Ordnungsrelation auf H machen. Wir fordern die

- Transitivität: $R_j < R_i \text{ und } R_i < R_k \Rightarrow R_j < R_k$
- Reflexivität: $R_i < R_i \Leftrightarrow R_i \in TR_i$
- Identivität: $R_i < R_j \text{ und } R_j < R_i \Rightarrow i = j$

Das Ordnungsdiagramm von H ist ein gerichteter Graph, der keine Zyklen enthält bis auf solche der Länge eins. Wir können H auch als hierarchische Hyperrelationalstruktur betrachten. In der Hierarchie lassen sich drei Klassen von Relationen finden,

- Basisrelationen: $HB = \{ R_i \mid \bigwedge_{j=1, N} T_j R_i \in B \}$
- Zielrelationen: $HZ = \{ R_i \mid \bigvee_{j \neq i} R_i < R_j \}$
- Interrelationen: $HI = \{ R_i \mid \bigvee_{j, k} R_j < R_i < R_k \}$

Die Basisrelationen enthalten keine anderen Relationen als Träger, die Zielrelationen sind in keiner anderen Relation enthalten.

In dieser Anwendung wird nur eine Basisrelation benutzt, nämlich die Menge der Grauwerte:

$$g = (x, y, z, t), \quad x \in [0, 511], \quad y \in [0, 572], \quad z \in [0, 255], \quad t \in [0, \dots)$$

Ihre charakteristische Funktion wird durch die Fernsehkamera und den nachgeschalteten Analog-Digital-Wandler berechnet (s. Kap. 4.1.1).

3.5 Weltmodell

Die Kenntnis der Korrespondenz zwischen den Relationen, die der Analysator berechnet, und den Beziehungen zwischen den Gegenständen in der realen Szene ist in zwei verschiedenen Arten im System vorhanden. Ein Teil des Wissens steckt in Prozeduren, die die Relationen berechnen. Sie wissen, welche untergeordneten Relationen sie benötigen und auf welche Weise deren Werte zur Berechnung der neuen Relation zu verknüpfen sind. Die Ergebnisse werden in einer globalen Datenbasis abgelegt (vgl. das "blackboard" Prinzip in den HEARSAY-Systemen [Lesser Erman 77]), die zum Teil im Arbeitsspeicher, zum Teil in externen Dateien gehalten wird. Die relationenberechnenden Prozeduren sind als Prozesse formuliert, deren Abhängigkeitsbeziehungen der "<-Relation entsprechen. Die Prozesse residieren in verschiedenen Programmen und Rechnern. Die Grauwertmatrix wird in einem Kleinrechner erzeugt, ihre Umkodierung in ein Vektorfeld (s. Kap. 4.1.3) erfolgt in einem Prozess, der zur Zeit als Programm im Zentralrechner läuft, später ebenfalls auf einen Kleinrechner übertragen wird. Die restlichen Prozesse sind auf zwei Programme ebenfalls im Zentralrechner verteilt. Ich gehe in den folgenden Kapiteln darauf ein.

Ein zweiter Teil des Weltmodells wird durch eine Parameterdatei gebildet. Die Parameter werden benutzt, um spezielles Wissen über eine Szene verfügbar zu machen. Sie steuern den Ablauf der Prozeduren. So legt einer der Parameter fest, durch welchen Grauwertbereich der Begriff 'dunkel' repräsentiert wird. Andere Parameter beschreiben Kontrast zwischen oder Homogenität von Grauwerten, oder Kollinearität von Bildvektoren. Alle Relationen sind weitgehend parametrisiert, damit das Weltmodell an unterschiedliche Szenen angepasst werden kann. Dabei wurde darauf geachtet, dass die Parameter eine einfache Beziehung zu den Merkmalen der Gegenstände in der realen Szene haben. Bei der Beschreibung der einzelnen Relationen in den folgenden Kapiteln gehe ich näher auf die Bedeutung der Parameter ein und gebe Beispiele an, die die Auswirkung ihrer Variation veranschaulichen.

4. Grauwerte und Bildvektoren

4.1 Vorverarbeitung

4.1.1 Abtastung

Die zu messende Grösse bei der Verarbeitung eines Schwarzweiss-Fernsehbildes ist die Bestrahlungsstärke auf der lichtempfindlichen Schicht der Kameraröhre (Vidikon oder Plumbikon in der hier verwendeten Anlage). Denkt man sich über die abgetastete Fläche, auf der das reelle Bild der Szene durch das Objektiv erzeugt wird, ein kartesisches Koordinatensystem gelegt, so lässt sich die Bestrahlungsstärke B als Funktion von x und y angeben.

$$B = f(x,y)$$

Bei der Digitisierung wird die Amplitude des Videosignals alle 100 nsec gemessen, dadurch wird dem Koordinatensystem ein Raster unterlegt. Per Vereinbarung liegt der Nullpunkt des Rasters in der linken, oberen Ecke des Bildes. Argument und Wertebereich der gemessenen Bildfunktion sind positive ganze Zahlen.

$$z_i = g(x_i, y_i), \quad z_i \in [0, 255], \quad x_i \in [0, 511], \quad y_i \in [0, 572]$$

Der an einem Punkt (x_i, y_i) gemessene Wert z_i ist im Idealfalle proportional zum Integral der Bestrahlungsstärke über eine Umgebung des Punktes auf dem Röhrentarget. Die Umgebung hängt ab von der Geometrie des Abtaststrahles, von der Abtastgeschwindigkeit in x - bzw. y -Richtung, von den Zeitkonstanten in den Signalverstärkern und anderen Faktoren. Ein Bilddurchlauf dauert 40 msec, wobei etwa 10 msec für die Übertragung der Synchronisationspulse verwendet werden. Die Grenzfrequenz des Videosignals liegt unter 5 MHz. In unserem experimentellen Aufbau werden $573 \cdot 512 = 293376$ Punkte pro Bild abgetastet, das entspricht einer Abtastfrequenz von 10 MHz. Die Abtastdichte ist in horizontaler Richtung geringer als in vertikaler, denn das Seitenverhältnis des Fernsehbildes ist $4/3$, das des Rasters $512/573$. Der Grund hierfür liegt zum einen am Festhalten an der Standard-(CCIR)-Fernsehnorm, durch die die Zeilenanzahl bestimmt ist, zum anderen wäre es unverhältnismässig aufwendig gewesen, einen

Analog-Digital-Konverter mit einer Wandlerfrequenz von mehr als 10 MHz einzusetzen, um das Raster in der Waagerechten zu verfeinern.

Der Messwert z_i weicht im wesentlichen aus drei Gründen von der Bildfunktion $g(x=x_i, y=y_i)$ ab, durch

- das Amplitudenrauschen des Abtaststrahles und der Verstärkerkette von der Kameraröhre bis zum Eingang des Analog-Digital-Wandlers,
- das Digitisierungsrauschen des Wandlers,
- das Rasterrauschen, hervorgerufen durch die Ablenkelektronik der Fernsehkamera. (Verringerbar durch Einsatz von Kameras mit lichtempfindlichen Halbleiterstrukturen, z.B. charge coupled device).

Falls eine Magnetplatte (uns steht der Typ Ampex MD 400 zur Verfügung) zur Zwischenspeicherung von Bildfolgen eingesetzt wird, erhöhen sich Amplituden- und Rasterrauschen. Für das zweite sind die Gleichlaufschwankungen des Plattenantriebs verantwortlich. Zum ersten tragen das Rauschen der Aufnahme- und Wiedergabeverstärker, das Rauschen des Modulators und Demodulators (das Videosignal wird einer Trägerfrequenz aufmoduliert) und das Rauschen der Magnetschicht bei.

Die von den Herstellern angegebenen technischen Daten der beteiligten Geräte sowie Messungen unserer Forschungsgruppe lassen bei guter Justierung der Anlage einen Amplituden-Rauschabstand von mindestens 34dB ohne bzw. 27dB mit Platte erwarten. Das Kurzzeit-Rasterrauschen liegt bei etwa einem Rasterpunkt. Gute Justierung bedeutet hier, dass die Kamera voll angesteuert ist, die Schwarz- und Weisspegel an den Eingangsbereich des Analog-Digital-Wandlers angepasst sind, die Gradationsentzerrung abgeschaltet ist und alle Geräte konstante Betriebsbedingungen erreicht haben. Der Rauschabstand ist gemittelt über alle Grauwerte, die Nichtlinearität des Wiedergabe-Gerätes und des menschlichen Auges sind unberücksichtigt.

Neben den oben genannten treten noch andere statistische Fehler auf, deren Einfluss jedoch geringer ist. Ausserdem gibt es systematische Fehler, wie z.B. die Temperaturabhängigkeit aller Stufen, in denen das Videosignal analog verarbeitet wird, oder die Schwankungen der Eigenschaften der lichtempfindlichen Schicht in der Aufnahmeröhre in Abhängigkeit vom Ort. Eine

ausführliche Erörterung der Eignung einer Fernsehkamera als Messinstrument gibt Rüenaufner [Rüenaufner 76] im Zusammenhang mit seiner Untersuchung ihrer Einsatzmöglichkeit als Bewegungsdetektor im Strassenverkehr.

Es soll jedoch noch auf zwei störende Effekte hingewiesen werden, die bei der Aufnahme von bewegten Objekten auftreten. Das optische Bild wird in der lichtempfindlichen Schicht in ein Ladungsbild umgewandelt, das bei der Abtastung wieder gelöscht wird. Das geschieht jedoch nicht vollständig, so dass es bei der Aufnahme kontrastierender, bewegter Objekte zu Nachzieherscheinungen kommt.

Da das Bild im Zeilensprungverfahren abgetastet wird, haben zwei örtlich aufeinanderfolgende Fernsehzeilen einen zeitlichen Abstand von 20 msec. In der Zeit kann sich das Objekt um mehrere Rasterpunkte bewegt haben. Da ein Umbau der gesamten Fernsehperipherie auf eine eigene Norm nicht in Frage kommt, kann der Effekt nur verringert werden, indem entweder die Halbbilder getrennt bearbeitet werden oder die Szene durch Blitzlicht beleuchtet wird, so dass sich das Ladungsbild innerhalb einiger Mikrosekunden aufbaut. Beide Auswege wurden bei den hier vorgestellten Experimenten nicht benutzt, vielmehr wurde gefordert, dass die Verfahren gegen alle bisher beschriebenen Störungen unempfindlich sein sollen.

4.1.2 Vorbemerkungen

Die im folgenden beschriebene Vorverarbeitung eines digitalen Bildes dient drei Zielen. Sie soll eine Umkodierung liefern, die gegen die oben beschriebenen Störeinflüsse unempfindlich ist, die eine spätere Abstraktion, d.h. symbolische Beschreibung des Bildinhaltes erleichtert und die das Datenvolumen reduziert. Die Güte einer solchen Reduktion lässt sich daran messen, inwieweit sich die Bildfunktion $z = f(x,y)$ aus dem reduzierten Bild rekonstruieren lässt. Sie ist zwar unbekannt, da nur die gestörten Messungen $z_i = g(x_i, y_i)$ vorliegen, jedoch lässt sich zumindest ein optischer Vergleich zwischen den Wiederdarstellungen des digitisierten Originals $g(x_i, y_i)$ und einer aus den vorverarbeiteten Daten rekonstruierten Bildfunktion durchführen.

Die Methode des optischen Vergleichs wurde übrigens in einem der ersten interaktiven Systeme zur Auswertung von

Wettersatelliten-Bildern angewendet [Arking 67]. Dort wurden zur Abtrennung der hellen Wolken von der Erdoberfläche Binärbilder durch Vergleich der Grauwerte mit einer variablen Schwelle erzeugt. Die Schwelle, bei der das Binärbild nach Beurteilung durch einen Operateur am besten mit dem Original übereinstimmte, wurde benutzt, um den Bedeckungsgrad der Erdoberfläche durch Wolken quantitativ zu bestimmen.

Eine einfache Methode, das Rauschen zu unterdrücken, besteht darin, ein Bild mehrfach zu digitisieren und die Ergebnisse zu mitteln. Ihre Anwendung lohnt sich jedoch nur bei Standbildern, da bei analog zwischengespeicherten Bildern aus Bildfolgen das Kamerarauschen sowie das Rauschen des Aufnahmeverstärkers und der Magnetschicht nicht mehr ausgemittelt werden kann.

Alternativ dazu kann man über eine Anzahl örtlich benachbarter Messwerte z_i mitteln. Damit kann auch eine Verminderung des Datenvolumens einhergehen je nach Grösse und Überlappung der Nachbarschaftsbereiche. Frühere Versuche des Autors, eine einfache, räumliche Integration des Bildes einer Segmentation zugrunde zu legen, die später noch von Albrecht, Fischer und Platz [Albrecht et al. 75] sowie von Budy und Dreschler [Budy Dreschler 76] aufgegriffen wurden, verliefen unbefriedigend, auch bei Verwendung von bimodalen Mittelwerten, die ein Verschleifen von Grauwertübergängen vermindern (formale Definition s. [Budy Dreschler 76] p. 6).

Aus der Physik sind Methoden, gestörte Messungen durch analytische Funktionen anzunähern, wohl bekannt und der einfachste Fall ist die Annahme eines linearen Zusammenhangs. Das im folgenden beschriebene Verfahren geht von der Annahme aus, dass sich die Bildfunktion tatsächlich für kleine Bereiche durch eine lineare Funktion des Ortes beschreiben lässt. Eine Approximation durch andere Funktionen, z.B. Polynome höherer Ordnung ist jedoch nicht durchgeführt worden. Hier wären kubische Spline-Funktionen einsetzbar, um eine stetige Anpassung an die Bildfunktion zu erreichen. Leider sind erst kürzlich effiziente Algorithmen entwickelt worden, um lokale Approximationen an eindimensionale [Ellis McLain 77] und zweidimensionale [Dierckx 77] Verteilungen zu berechnen. Lokal bedeutet, dass nur die Funktionswerte einer engen Umgebung des anzupassenden Ausschnittes benutzt werden müssen.

4.1.3 Definition der Bildvektoren

Es sei angenommen, dass die Verteilung der Beleuchtungsstärke auf der Kamera-Sensorfläche für einen kleinen Ausschnitt eine lineare Funktion des Ortes ist.

$$z(x,y) = U*x + V*y + W$$

Weiterhin sei der Messwert z_i um den tatsächlichen Wert $z(x_i, y_i)$ normalverteilt mit der Standardabweichung s_i . Dann ist die Wahrscheinlichkeit P_i dafür, an einem Ort (x_i, y_i) den Wert z_i zu beobachten, gegeben durch

$$P_i = 1/(s_i * \sqrt{(2*\pi)}) * e^{(-1/2*((z_i - z(x_i, y_i))/s_i)^2)}$$

Die Wahrscheinlichkeit, gleichzeitig bestimmte Messwerte von N Punkten (für die noch ein linearer Zusammenhang gelten soll) zu beobachten, ist

$$P(U, V, W) = \prod_i (P_i) \text{ mit } i = 1, \dots, N$$

Für jede Schätzung

$$z'(x,y) = A*x + B*y + C$$

des wahren Verlaufs ergibt sich ebenso

$$P(A, B, C) = \prod_i (1/(s_i * \sqrt{(2*\pi)}) * e^{(-1/2*((z_i - z'(x_i, y_i))/s_i)^2)}$$

Die Schätzung der Koeffizienten nach der "maximum likelihood" Methode besteht darin, anzunehmen, dass die wahre Verteilung am besten durch eine solche Wahl der Koeffizienten beschrieben wird, die die Wahrscheinlichkeit $P(A, B, C)$ maximiert. Das ist gleichbedeutend damit, dass die Summe

$$\sum_i ((1/s_i^2) * (z_i - A*x_i - B*y_i - C)^2)$$

ihr Minimum annimmt.

Die partiellen Ableitungen nach den Koeffizienten liefern das Gleichungssystem

$$\sum_i (x_i/s_i^2) * (z_i - A*x_i - B*y_i - C) = 0$$

$$\sum_i (y_i/s_i^2) * (z_i - A*x_i - B*y_i - C) = 0$$

$$\sum_i (1/s_i^2) * (z_i - A * x_i - B * y_i - C) = 0$$

Transformiert man den Ursprung des Koordinatensystems in den Mittelpunkt des Ausschnitts gebildet durch die N verwendeten Rasterpunkte, so dass sich die Werte für x_i , y_i und s_i (s. unten) symmetrisch um den Nullpunkt gruppieren, vereinfachen sich die Gleichungen zu

$$\sum_i (x_i/s_i^2) * (z_i - A * x_i) = 0$$

$$\sum_i (y_i/s_i^2) * (z_i - B * y_i) = 0$$

$$\sum_i (1/s_i^2) * (z_i - C) = 0$$

mit den Lösungen

$$C = \frac{\sum_i (z_i/s_i^2)}{\sum_i (1/s_i^2)}$$

$$B = \frac{\sum_i (z_i * y_i/s_i^2)}{\sum_i (y_i^2/s_i^2)}$$

$$A = \frac{\sum_i (z_i * x_i/s_i^2)}{\sum_i (x_i^2/s_i^2)}$$

Die Anzahl der benutzten Rasterpunkte ist nach unten begrenzt durch die Forderung, dass eine Ebene sechs benachbarte Rasterpunkte in zwei Spalten und drei Zeilen repräsentieren soll. Dann ist das Seitenverhältnis der neuen Rasterung $(512/573) * (3/2) = 4/3$ dasselbe wie beim Fernsehbild und die Auflösung beträgt $256 * 191 \approx 50000$ Ebenen, die sich jede über einem quadratischen Bereich der Sensorfläche erheben. Nach oben ist die Zahl begrenzt durch die Voraussetzung, dass die Bildfunktion im betrachteten Bereich linear sein soll. Es wurde ein Fenster von $4 * 5$ Rasterpunkten benutzt, das sequentiell um zwei Schritte in x-Richtung und am Zeilenende um drei Schritte in y-Richtung verschoben wird.

Die Messunsicherheit wurde für alle Grauwerte z_i des Bildes als gleich angenommen, fällt also als konstanter Normierungsfaktor aus den Gleichungen für die Koeffizienten heraus. Jedoch wurde der Faktor $1/s_i^2$ als Gewicht w_i für die im Ausschnitt verwendeten Rasterpunkte beibehalten, um den Beitrag der inneren $2 * 3$ Punkte des $4 * 5$ Fensters am Ergebnis der Anpassung zu erhöhen. Um die unterschiedliche Abtastdichte in x- und y-Richtung zu kompensieren, wurde eine elliptische Gewichtsfunktion gewählt

$$w_i = 1/\sqrt{(x_i^2 + exc^2 * y_i^2)}$$

Ausserdem werden die Koeffizienten auf ein Feld der Seitenlänge eins unnormiert.

$$A = \frac{\sum_i (z_i * x_i * w_i)}{2 * \sum_i (x_i * x_i * w_i)}$$

$$B = \frac{\sum_i (z_i * y_i * w_i)}{3 * \sum_i (y_i * y_i * w_i)}$$

$$C = \frac{\sum_i (z_i * w_i)}{\sum_i (w_i)}$$

Als Exzentrizität wurde $exc = (512/573)/(4/3)$ eingesetzt. Die resultierenden Werte sind zusammen mit den Koordinatenwerten der 20 Punkte in der Tabelle 4.1 angegeben

Tabelle 4.1

X	-1.5	-0.5	0.5	1.5
Y				
-2	0.50	0.70	0.70	0.50
-1	0.61	1.20	1.20	0.61
0	0.67	2.00	2.00	0.67
1	0.61	1.20	1.20	0.61
2	0.50	0.70	0.70	0.50

Die Koeffizienten A, B und C sind wenig anschaulich zur Beschreibung der Ebenen. Sie werden umgerechnet in folgende Parameter.

- den mittleren Grauwert G
das ist der Schnittpunkt der Ebene mit der lokalen z-Achse
- die Neigung N der Ebene aus der Horizontalen
das ist der Winkel der Ebenennormalen mit der z-Achse
- die Richtung R des Normalenvektors
das ist der Winkel seiner Projektion auf die x-y-Ebene mit der x-Achse.

Als Mass für die Güte der Anpassung wird ausserdem noch berechnet

- der mittlere Abstand S der Grauwerte z_i von der Ebene
(andere Masse könnten die Streuung der Koeffizienten oder der lineare Korrelationskoeffizient sein)

Aus der Umformung

$$C = z - A * x - B * y$$

ergibt sich mit

$$\text{NORM} = \sqrt{A^2+B^2+1}$$

die Hessesche Normalform

$$C/\text{NORM} = -x*A/\text{NORM}-y*B/\text{NORM}+z/\text{NORM}$$

und damit

$$G = C$$

$$N = \arccos (1/\text{NORM})$$

$$R = \arccos (-A/\sqrt{\text{NORM}^2-1}),$$

der mittlere Abstand ist

$$S = \frac{\sum_i (|z_i - A*x_i - B*y_i - C| * w_i)}{\sum_i (w_i)}$$

Mit dem beschriebenen Verfahren ist die Grauwertmatrix des Bildes umkodierbar in ein Feld von "Bildvektoren".

4.1.4 Erzeugen eines Vektorfeldes

Der Algorithmus zur lokalen Anpassung von Ebenen an die Bildfunktion über der Bildmatrix ist auf der PDP-10 des Instituts für Informatik implementiert. Die Wortlänge der PDP-10 beträgt 36 Bit, jedoch sind die Eingabe- wie auch die Ausgabedaten in 8 Bit Bytes organisiert, da der Algorithmus auf einen Byte-organisierten Prozessrechner des Typs MINCAL 621 übertragen werden soll.

Die Bildmatrix wird eingelesen und die Grauwert-Bytes des aktuellen Fensters in ein Feld der Länge 20 übertragen. Das Feld wird einem Unterprogramm übergeben, das die Anpassung der Ebenen durchführt und drei Bytes als Resultat liefert:

- die Richtung R der Ebene, kodiert in fünf Bits und die Neigung N der Ebene, kodiert in drei Bits,
- den mittleren Grauwert G
- den mittleren Abstand S multipliziert mit zehn, so dass ein Bereich von 0.1 bis 25.5 darstellbar ist.

Das Unterprogramm berechnet die Summen in den Zählern der Koeffizientengleichungen, die drei Koeffizienten, den mittleren Abstand der Grauwerte und die Parametrisierung in N , R , G . Dabei werden für N acht Werte für die Funktion $1/\cos^2$ - da $1/\cos^2(N) = \text{NORM}^2$ - und für R 32 Werte der Funktion \cos^2 - da $\cos^2(R) = A^2/(\text{NORM}^2-1)$ - in Tabellen gespeichert und zur Berechnung des arccos verwendet. Ebenfalls im Initialisierungsteil werden die Gewichte und alle vom Grauwert unabhängigen Summen berechnet.

Während der Anpassung werden nur Additions- und Multiplikationsoperationen benötigt; der Zeitbedarf liegt unter 3 msec für eine Ebene, für ein ganzes Bild bei 2.5 min einschliesslich der Ein-Ausgabe-Zeiten. Eine weitere Optimierung bezüglich des Zeitbedarfs ist möglich, wird aber erst bei der Implementation im Prozessrechner vorgenommen.

Die Abbildung 4.1 zeigt eine Strassenszene, reproduziert auf einem Ausgabegerät für Bildfunk-Empfang, das über eine in unserem Labor selbst entwickelte Schnittstelle an einen MINCAL 621 Prozessrechner angeschlossen ist [Neumann 78a].

Die Abb. 4.2 zeigt eine Rekonstruktion aus den mittleren Grauwerten der Bildvektoren.

Die Abb. 4.3 zeigt die Grauwertdarstellung der Neigung N multipliziert mit 2^5 . Der Faktor transformiert den Bereich $0 \dots 7$ auf $0 \dots 255$.

Die Abb. 4.4 zeigt die Grauwertdarstellung des mittleren Abstandes S . Hier sieht man, dass in den Bereichen grosser Neigung auch S gross wird; bei zunehmender Steilheit der Grauwertübergänge wächst auch die Differenz zwischen Messwert und approximiertem Wert.

Die Abb. 4.5 zeigt einen Ausschnitt, nämlich ein Fahrzeug, der links oben im Bild eingeblendet ist. Der Ausschnitt umfasst 64×48 Vektoren, wobei jedes Quadrat, das mit dem mittleren Grauwert gefüllt ist, von einem Streifen zerschnitten wird. Dessen Richtung entspricht der des Bildvektors ($+\pi/2 \bmod \pi$) und seine innere Schwärzung ist ein Mass für die Neigung des Vektors. Durch die Drehung der Vektoren um 90° bei der Wiederdarstellung zeigen sie nicht mehr die Richtung des Gradienten an, sondern liegen parallel zum Grauwertübergang. Die Streifen ordnen sich entlang von Grauwertübergängen zu Linienstrukturen, während sie in Gebieten homogenen Grauwertes mehr zufällig ausgerichtet sind. Im unteren Teil der Abbildung

4.5 sieht man am Beispiel des Holmes zwischen den beiden Seitenfenstern, dass linienförmige Grauwertstrukturen dieser Breite noch gut durch parallele Vektorzüge beschrieben werden.

Das Datenvolumen wird von sechs Byte pro Fenster auf drei Byte für die Vektorparameter reduziert.

4.2 Gradientenraum

Die Transformation von der Gradientenmatrix in den Raum, der durch die drei Komponenten R, N und G der Bildvektoren aufgespannt wird, lässt eine leichte Einsicht in die Systematik der Grauwertverteilung im Bild zu. Nicht der einzelne, unsicher gemessene, absolute Wert der Lichtintensität beschreibt ein Bild, sondern - wie wir von Horn [Horn 77] und anderen lernen können - der Wert im Verhältnis zu seinen Nachbarwerten. Trotzdem haben sich bisher nur wenige Autoren damit beschäftigt, eine Kodierung des Bildes zu finden, die das Verhältnis von Grauwerten beschreibt und als Grundlage für weitere Abstraktionen dient.

Chien und Jacobus untersuchen die Richtungsableitung der Bildfunktion und begründen das so [Chien Jacobus 76]:
 "The use of high level information when possible is, of course, to be commended, but when it is used to make up for insufficient low level feature data, systems begin to 'see' what they want to 'see' rather than what is actually in the scene.

To avoid this problem our approach is based on the premise that more work is required at the bottom to extract and order information. To this end we have looked for ways to identify and use more completely the information encoded in picture arrays."

Sie kommen in ihren ersten Untersuchungen zu dem Schluss, dass Betrag und Richtung des Gradienten der Bildfunktion, der Gradient der Richtung und die Änderung des Betrages sinnvolle, bildbeschreibende Größen darstellen.

Yoda et al. verwenden eine der Fernsehkamera nachgeschaltete spezielle Elektronik, um die Bildfunktion in Zeilen- und Spaltenrichtung abzuleiten [Yoda et al. 75]. Das Ergebnis ist ein Mass für die Richtung des Gradienten, kodiert in 6 Bit, und

ein besonderes Signal, das anzeigt, dass der Betrag des Gradienten unter einer Schwelle liegt, unterhalb der vermutet wird, dass Rauschen die Richtungsbestimmung dominiert.

Neumann benutzt ein Gradientenfeld, um gerade Konturen im Bild zu finden [Neumann 78b]. Er sucht es zeilenweise parallel zur x-Achse ab und sammelt benachbarte Gradienten, deren Richtung nahe der positiven x-Achse liegt und deren Betrag einen Schwellenwert überschreitet, zu "Strichen". Das selbe geschieht mit Gradienten, die in etwa die Richtung der negativen x-Achse haben. Überlappende Striche werden zu längeren Strichen verschmolzen. Das Verfahren wird noch sieben mal wiederholt, nachdem jeweils das Koordinatensystem um einen Winkel von $180^\circ/8$ gedreht worden war. Danach liegen Striche in 16 verschiedenen Richtungen vor. Das Verfahren erkennt mehrfach repräsentierte Kanten und beseitigt die überflüssigen Striche.

O'Gorman und Clowes haben eine von Duda und Hart beschriebene [Duda Hart 73] Transformation benutzt, um ebenfalls Geraden zu finden [O'Gorman Clowes 76]. Der Ort und die Richtung eines Gradienten bestimmen eine Gerade, die durch zwei Koeffizienten parametrisiert werden kann. Für das Bild wird ein zweidimensionales Histogramm der Koeffizienten aufgestellt. Seine Häufungsmaxima steuern die Suche nach "Strichen" im Gradientenfeld.

Bowker schlägt vor, beim Suchen nach Konturen Felder von Richtungsvektoren zu verwenden, die mit verschiedener räumlicher Auflösung berechnet worden sind [Bowker 74]. Er macht eine kurze Bemerkung in seinem Artikel, die er aber nicht weiter verfolgt hat, nämlich dass die Unordnung von Richtungsvektoren innerhalb eines von kohärenten Vektorzügen umrandeten Gebietes ein Mass für die Textur der Objektoberfläche sein kann. Ich zeige später Beispiele, wie durch Hervorheben von Bildbereichen, die durch geordnete bzw. ungeordnete Bildvektoren gekennzeichnet sind, eine "duale" Bildbeschreibung erzeugt werden kann.

Die von Bowker vorgeschlagenen Operationen entsprechen etwa denen von Nevatia angewendeten [Nevatia 76b]. Letzterer extrahiert kurze Liniensegmente aus dem Grauwertbild, fasst benachbarte, annähernd parallele Segmente, die etwa in einer vorgegebenen Richtung liegen, zu Paketen zusammen und versucht, in jedem Paket ein Segment zu finden, das sich mit möglichst geringer Richtungsabweichung möglichst lückenlos in eine Gerade einfügt.

Ramer löst das Problem, in einem Feld von Richtungsvektoren zusammenhängende Linien zu finden, nach Transformation in das Problem, einen optimalen Weg in einem Graphen mit bewerteten, ungerichteten Kanten zu finden [Ramer 75]. Die Knoten sind die Richtungsvektoren, die nach einem Schwellenkriterium über die Stärke des lokalen Gradienten noch als Linienelemente in Frage kommen. Die Kanten laufen zwischen zwei im Raster benachbarten (maximal acht Nachbarn) Vektoren, die Kostenfunktion beschreibt die Passung eines Vektors in eine Linie, die von einem Startvektor ausgeht. Ramer sucht in einem zusammenhängenden Gebiet von Linienelement-Kandidaten einen Startvektor und führt eine erschöpfende Suche nach beiden Seiten in den den Startvektor einschliessenden Komponenten des Graphen durch.

Im Gegensatz dazu kann das ähnliche Verfahren von Martelli [Martelli 72,76] den Suchraum heuristisch eingrenzen und setzt keine disjunkte Zerlegung des Graphen voraus. Es arbeitet aber direkt auf Grauwerten und nicht auf Linienelementen. Beide Verfahren haben einen Nachteil gemeinsam. Sie machen die Annahme, dass der Startpunkt der Suche Teil des optimalen Pfades ist. Pohl hat den klassischen A* Algorithmus [Hart et al. 68], der von Martelli verwendet wird (vergl. [Martelli 77]), ausgebaut zu einem bidirektionalen Suchverfahren [Pohl 71]. Sein Einsatz würde den erwähnten Nachteil vermeiden. Der Erfolg des Ramerschen Verfahrens hängt sehr davon ab, ob es immer mit Sicherheit gelingt, den Kandidatenbereich zweier semantisch unterschiedlicher Linien zu trennen, denn er sucht in jedem Kandidatenbereich nur eine Linie.

Das Verfahren von Holdermann und Kazmierczak ist das einzig mir bekannte, das die Richtungsinformation für das Verfolgen eines Konturverlaufs nicht aus der durch Differenzenbildung angenäherten Ableitung der Bildfunktion gewinnt, sondern die Grauwertverteilung lokal durch Ebenen approximiert. Es geht ebenfalls von einem Startvektor aus und versucht sequentiell fortsetzende Vektoren zu finden [Holdermann Kazmierczak 72], [Holdermann 73]. Falls in einem durch Öffnungswinkel und Entfernung vom letzten Vektor definierten Bereich kein Vektor mit ähnlicher Richtung existiert, wird der Öffnungswinkel erweitert und die Hypothese geprüft, ob die Kontur einen Knick macht. Findet sich keine Fortsetzung, wird die Kontur beendet.

5. Gruppierung von Bildvektoren

Das Ziel der Segmentation ist im allgemeinen, die Grauwertmatrix des projizierten Bildes in Bereiche und Konturen zu zerlegen, wobei die Bereiche die Bedeutung von Oberflächen des realen Gegenstandes und Konturen die von Grenzen zwischen aneinander stossenden oder sich verdeckenden Oberflächen haben sollen. Dabei wird fast immer davon ausgegangen, dass sich lokal berechenbare Merkmale der Grauwertverteilung innerhalb des Bildausschnittes, der von einer Oberfläche herrührt, wenig ändern, jedoch beim Überschreiten einer Oberflächengrenze stärkere Schwankungen aufweisen. Ein wichtiger Parameter ist dann der Schwellenwert, mit dessen Hilfe entschieden wird, ob eine Variation noch zulässig innerhalb eines Bereiches ist oder eine Kontur beschreibt. Ein typisches Beispiel sind die Histogrammverfahren von Ohlander und Price [Ohlander 75], [Price 76], ein anderes das von Yakimovsky [Yakimovsky 75]. Letzterer hat ein Likelihood-Mass eingeführt, das die Streuung zweier benachbarter Felder von Grauwerten mit der des aus beiden kombinierten Feldes vergleicht.

$$L = s_{12}^{n+m} / (s_1^n * s_2^m)$$

wobei n und m die Zahl der Messwerte in Feld 1 bzw. Feld 2 ist. Das Mass wird zur Entscheidung benutzt, ob zwischen beiden Feldern ein Teilstück einer Kontur liegen kann. Die Modifikationen des Yakimovsky'schen Verfahrens durch Nagel [Nagel 76a] sowie durch Kraasch und Zach [Kraasch Zach 78] beruhen ebenfalls auf diesem Mass. Diese Verfahren nutzen eine eventuell vorhandene Systematik des Grauwertverlaufes an den Rändern der Oberflächen nicht optimal aus. Die Berechnung von Richtung und Stärke eines Grauwertüberganges durch Approximation des Gradienten der Bildfunktion $g = f(x,y)$ mit Hilfe von Differenzen benachbarter Messwerte gewinnt schon mehr Information aus einer Grauwertmatrix. Aber auch die im Abschnitt 4.2 genannten Autoren gehen von der Vorstellung von homogenen Oberflächen und einem Kontrastmaximum an den Grenzen aus. Dass sie (Holdermann, Bowker, Ramer) jedoch Operationen anwenden müssen, um in einem Strom von parallelen und oftmals auch gleichstarken Vektoren die "zentrale" Linie zu finden, deutet schon darauf hin, dass diese Vorstellung nicht zutrifft.

So kann bei gewölbten Oberflächen (z.B. die Kotflügel des Wagens in Bild 4.5) der Gradient mit dem grössten Betrag in der Nähe des Reflexionsmaximums auftreten (insbesondere bei einer dunklen Färbung) und nicht am Rand der Oberflächen. Als Beispiel für diesen Fall zeigt die Abbildung 5.1 die Zerlegung

eines Bildes durch den Algorithmus von Kraasch und Zach. Kann also nicht garantiert werden, dass an den Grenzen von Oberflächen die grössten Grauwertänderungen auftreten, so ist es ungünstig, eine Segmentation mit den am Anfang des Kapitels genannten Zielen zu erzwingen mit Methoden, die von einer Homogenität von Bereichen ausgehen. Segmentationsalgorithmen, die nicht von der Vorstellung eines im wesentlichen konstanten Grauwertes innerhalb eines Bereiches ausgehen, sondern für den Grauwert eine lineare Funktion über den Koordinaten zugrunde legen, werden von Yakimovsky und Cunningham [Yakimovsky Cunningham 76a] sowie von Somerville und Mundy [Somerville Mundy 76] angegeben. Gerade Verfahren, bei denen durch die Art der Vorverarbeitung Indizien sichtbar werden, dass Oberflächen von Gegenständen nicht durch scharfe Grauwertübergänge im Bild abgegrenzt sind, sollten diese Information nicht durch das Anwenden von Operationen unterdrücken, die auf heuristischem Wege eine scharfe Kontur plazieren.

Besonders beim Vergleich von Bildern aus einer Folge ist es wichtig, Beschreibungsmerkmale zu verwenden, die zuverlässig mit der Bedeutung der Szene zusammenhängen. Ausgehend von der Annahme, dass das im vorigen Kapitel eingeführte Vektorfeld eine in diesem Sinne zuverlässige Abstraktion der Grauwertmatrix darstellt, habe ich nach Relationen gesucht, die wiederum das Vektorfeld beschreiben, ohne auf das klassische Repertoire von Bereichen und Konturen zurückgreifen zu müssen.

5.1 Filtern im Vektorfeld

Anregungen zur formalen Struktur des im folgenden beschriebenen Verfahrens gaben die "Filter"-Verfahren von Waltz [Waltz 72] sowie Tenenbaum und Barrow [Tenenbaum Barrow 76], das "relaxation labelling" [Rosenfeld et al. 76], [Zucker 76b] und die Überlegungen von Mackworth [Mackworth 77] und Freuder [Freuder 76] über Konsistenz in Graphen.

Jeder Vektor eines Feldes ist beschrieben durch sechs Komponenten

$$V = (X, Y, R, N, G, S)$$

den Spalten- und Zeilenkoordinaten, der Richtung und Neigung,

durch Grauwert und Streuung. Die Vektoren bilden die Knoten eines Graphen, dessen Kanten durch die Nachbarschaftsrelation initialisiert sind.

$$\text{NACHBAR}(V_1, V_2) \Leftrightarrow 0 < (x_1 - x_2)^2 + (y_1 - y_2)^2 < 3$$

d.h. jeder Vektor hat acht Nachbarn ausgenommen am Bildrand.

Für das Vektorfeld definiert das Weltmodell Bedingungen, die Werte der Vektorkomponenten einschränken und Relationen zwischen benachbarten Vektoren vorschreiben. In einem dreistufigen Verfahren werden Knoten und Kanten aus dem Graphen entfernt, die mindestens eine der vorgegebenen Bedingungen verletzen. Der mehrstufige Aufbau verringert den Rechenaufwand im Sinne einer hierarchischen Synthese [Barrow et al. 72], [Ambler et al. 75] von Zwangsbedingungen [Freuder 76]. Der Graph zerfällt in voneinander isolierte Subgraphen. Deren Eigenschaften beschreiben Bildteile mit der vom Weltmodell durch den Parametersatz der Zwangsbedingungen formulierten Bedeutung (s. Kap. 3.5). Solche Subgraphen nenne ich auch Bereiche, da sie aus benachbarten Vektoren bestehen. Das Bild wird nicht vollständig in Bereiche zerlegt, nicht jeder Vektor muss in einem Bereich enthalten sein, jedoch ist für einen gerade gewählten Parametersatz ein Vektor in höchstens einem Bereich enthalten. Enthält das Weltmodell mehrere Parametersätze, um unterschiedliche Arten von Bereichen aus dem Bild zu extrahieren, so ist nicht gefordert, dass Bereiche unterschiedlicher Art disjunkt sind. Isolierte Subgraphen, die nur noch aus einem Knoten bestehen, werden aus dem Graphen entfernt.

5.2 Knotenkonsistenz

Der Graph ist als zweidimensionales Feld aus 36 Bit Worten realisiert. Jedes Wort enthält in 24 Bit die vier Komponenten R, N, G und S eines Bildvektors, dessen Koordinaten bestimmen die Position des Wortes im Feld. Somit benötigt ein Bild einen Speicherplatz von $256 * 192 = 49152 = 48\text{K}$ Worte. Die Kanten des Graphen sind in 8 Bit unär kodiert. Von jedem Knoten aus ist sofort festzustellen, mit welchen Nachbarn er verbunden ist, allerdings müssen beim Löschen einer Kante zwei Eintragungen entfernt werden. Da jedoch das Prüfen, ob eine Kante vorhanden ist, häufiger vorkommt als das Löschen, ist die

Darstellung vorteilhaft. Ein weiterer Vorteil besteht darin, dass das Fehlen aller acht Kantenindikatoren als Signal dient, dass der Knoten vollständig isoliert ist, also nicht mehr zum Graphen gehört. Ein einfaches logisches, bitweises UND mit einer entsprechenden Maske gestattet es, diese Aussage schnell zu berechnen.

In der ersten Phase wird das Bild sequentiell, zeilenweise bearbeitet. Aus dem Weltmodell wird ein Parametersatz eingelesen, der die zulässigen Intervalle für die Koordinaten, die Richtung und Neigung, den Grauwert und die Streuung enthält. Liegt der Wert mindestens einer Vektorkomponente ausserhalb eines erlaubten Intervalls, so werden alle Kanten zwischen dem Knoten und seinen Nachbarn gelöscht.

Im Prinzip lässt sich in dieser Phase jede beliebige Funktion der Vektorkomponenten als Diskriminator benutzen. Eine Erweiterung des einfachsten Falles ergibt sich aus der Anwendung von "fuzzy" Logik. Seien sechs F-Mengen definiert durch die charakteristischen Funktionen f_x , f_y , f_r , f_n , f_g und f_s jeweils mit normiertem Wertebereich und einem Bildvektor als Argument, dann legt ein Ausschnitt ("level set") der Durchschnittsmenge die im Graphen verbleibenden Knoten fest.

$$\text{KNOTEN} = \{ V \mid f_x(V) \text{ F-UND } f_y(V) \text{ F-UND } \dots f_s(V) > \theta \}$$

Solche Formulierung hilft bei der Anwendung vager Konzepte wie 'hell' oder 'dunkel' bezogen auf den Grauwert, oder 'horizontal' bzw. 'vertikal' bezogen auf die Richtung oder 'kontrastreich' bezogen auf die Neigung.

Der von mir verwendete Diskriminator würde in obiger Schreibweise lauten

$$\text{KNOTEN} = \{ V \mid \prod_K f_K(V) > 0 \} \text{ mit}$$

$$1 \Leftrightarrow K \in [K_{\min}, K_{\max}]$$

$$f_K(V) = \begin{cases} 1 & \text{falls } K \in [K_{\min}, K_{\max}] \\ 0 & \text{sonst} \end{cases}$$

mit $K \in \{ "X", "Y", "R", "N", "G", "S" \}$

F-UND entspricht hierbei dem arithmetischen Produkt.

Eine weitere Verallgemeinerung ergibt sich, wenn F-Mengen definiert werden, deren charakteristische Funktion von einer Kombination von Vektorkomponenten abhängt oder sogar von

den Komponenten mehrerer Vektoren. So kann z.B. der mittlere Grauwert des ganzen Bildes mit in die Entscheidung einbezogen werden oder, um Konturen zu extrahieren, werden nur solche Vektoren behalten, deren Neigung mindestens ebenso gross ist wie die aller ihrer Nachbarn.

Der Phantasie ist bei der Wahl der Zugehörigkeitsfunktion eines Knoten zum Graphen keine Grenze gesetzt, doch sollten zwei Gesichtspunkte beachtet werden. Zum einen soll die Rechenzeit in vertretbaren (s. Kap. 3.1) Grenzen bleiben, etwa in der Grössenordnung von einigen Sekunden für diese Phase, in der eher eine Vorauswahl im Sinne einer Hierarchie zur Aufwandsreduktion in den nachfolgenden Stufen vorgenommen werden soll.

Zum anderen sollen die Operationen eine einfache Bedeutung haben, so dass schon das "Anschauen" eines Grautonbildes Hinweise auf die Festsetzung des Parametersatzes im Weltmodell geben kann, um Bereiche mit der gewünschten Bedeutung zu extrahieren.

Das erste Bildbeispiel (5.2a) zeigt die Grauwerte von Vektoren in Bereichen mit geringer lokaler Grauwertschwankung. Aus darstellungstechnischen Gründen wurde nur ein Ausschnitt bearbeitet, jenseits seiner Grenzen werden alle Grauwerte wiedergegeben. Die einzigen Einschränkungen waren

$X \in [65,128]$ (Spaltennummer)
 $Y \in [115,162]$ (Zeilennummer)
 $N \in [0,45]$ (Grad)

Die Abbildung 5.2b stellt den zum Ausschnitt gehörenden Graphen dar. Seine Knoten sind als kleine Quadrate, die Kanten als dünne Linien gezeichnet.

Im zweiten Beispiel werden die Bereiche erfasst, in denen ein Grauwertübergang stattfindet (Abb. 5.3a). Entsprechend wurde die Neigung auf

$N \in [63,90]$

eingeschränkt. Den zugehörigen Graphen gibt die Abb. 5.3b wieder.

In beiden Grautonbildern (5.2a und 5.3a) reicht der Kontrastumfang der Reproduktion nicht aus, um die Grauwerte in den Subgraphen im Bereich des hellen Daches sichtbar zu machen.

Als Anmerkung sei darauf hingewiesen, dass Randpunkte und innere Punkte eines Bereiches leicht durch die Zahl ihrer Nachbarn unterschieden werden könnten, wie aus den Bildern 5.2b und 5.3b ersichtlich ist.

5.3 Kantenkonsistenz

Sind alle Vektoren daraufhin überprüft, ob sie die Vorgaben für ihre Komponenten einhalten, resultiert ein knotenkonsistenter Graph. In der nächsten Phase werden Paare von benachbarten Vektoren untersucht, um Kanten-, ja sogar Pfadkonsistenz (als transitive Hülle der Kantenkonsistenz) zu erreichen (s. Kap. 2.2.1, [Mackworth 77]). Das ist gleichbedeutend damit, Beziehungen zwischen Vektoren herzustellen, um so Strukturen im Bild zu erkennen. Auch hier wurde wieder auf einfachste Zusammenhänge zurückgegriffen, die jedoch ausreichen, um Regelmässigkeiten des Vektorfeldes zu beschreiben.

Folgende Grössen wurden für benachbarte Paare berechnet:

a) Differenz der Grauwerte

$$D_g = |G_1 - G_2|$$

b) Differenz der Richtungen

$$D_r = \min(|R_1 - R_2|, 360 - |R_1 - R_2|)$$

Der Wert von D_r steigt linear zwischen 0° und 180° an und fällt dann ebenso zwischen 180° und 360° ab.

Lässt man einen Zeiger von einem Vektor zu jedem seiner acht Nachbarn zeigen, so nimmt er nacheinander die Winkelpositionen von 0 bis 315 Grad in Schritten von 45 Grad ein. Ich nenne den Winkel Rasterrichtung und bezeichne ihn mit L . Die beiden folgenden Masse bewerten die Beziehung zwischen der Richtung beider Vektoren und der Rasterrichtung, die ihre gegenseitige Lage festlegt.

c) Minimale Differenz zwischen Vektor- und Rasterrichtung

$$LR_i = \min(\min(|R_i - L|, 360 - |R_i - L|), 180 - \min(|R_i - L|, 360 - |R_i - L|))$$

$$i = 1, 2$$

$$DL_{\min} = \min(LR_1, LR_2)$$

d) Summe der Differenzen zwischen Vektor- und Rasterrichtung

$$DL_{\text{sum}} = LR_1 + LR_2$$

LR hat die Gestalt einer Dreiecksfunktion (Fig. 5.1).

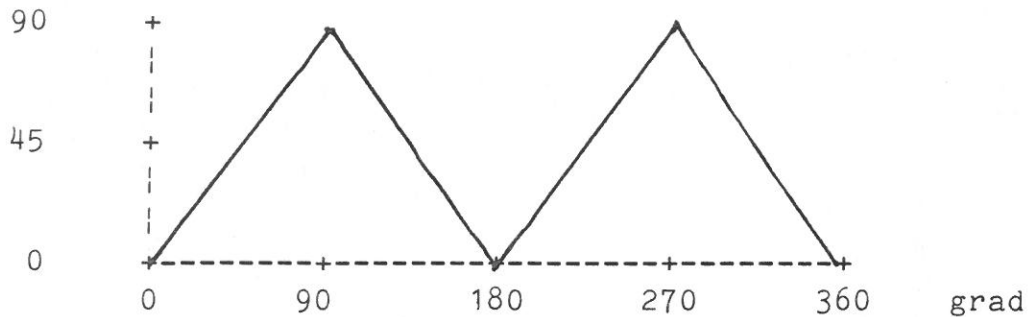


Fig 5.1

Der Diskriminator für das Beibehalten von Kanten ist analog zu dem für die Knoten definiert,

$KANTEN = \{ (V_1, V_2) \mid \prod_D f_D(V_1, V_2) > 0 \}$ mit

$$f_D(V_1, V_2) = \begin{cases} 1 & \Leftrightarrow D \in [D_{min}, D_{max}] \\ 0 & \text{sonst} \end{cases}$$

mit $D \in \{ "D_G", "D_R", "DL_{min}", "DL_{sum}" \}$

Auch hier bleibt wieder ein grosser experimenteller Spielraum für die Auslegung des Diskriminators. Es erscheint mir sinnvoll, erst einmal mit einfachen Entscheidungen zu arbeiten, auch deswegen, um einen Anhaltspunkt für die minimale Laufzeit der Selektion zu bekommen.

Einige Beispiele erläutern den Einfluss von D_R , DL_{min} und DL_{sum} . Aus der Kodierung der Graustufe, der Wert für weiss ist 0, der für schwarz 255, folgt, dass die Bildvektoren zum hellen Gebiet geneigt sind. In der Darstellung in Abbildung 4.5 sind sie um einen Viertelkreis so gedreht, dass links das hellere und rechts das dunklere Gebiet liegen. An der Grenze zweier Bereiche ergibt sich der folgende Verlauf:

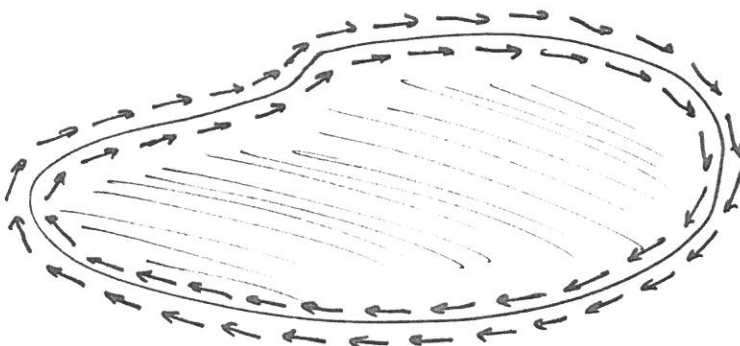


Fig. 5.2

Wie kann nun das Weltmodell eine Kontur zwischen einem hellen und einem dunklen Bereich beschreiben? Wählen wir $D_R=0$, $DL_{min}=0$ und $DL_{sum}=0$, so können nur kollineare Paare auftreten. Bei dieser Wahl bleiben nur Pfade im Graphen erhalten, die Vektorzügen in den acht Hauptrichtungen entsprechen; wie schon erwähnt, ist auf einem rechtwinkligen Raster die relative Lage von Nachbarn auf Vielfache von 45° beschränkt. Die Abbildung 5.4 enthält nochmal eine Darstellung der Vektoren eines Bildausschnittes, 5.5 zeigt den Graphen für die Wahl $D_R=0$, $DL_{min}=0$ und $DL_{sum}=0$.

Um allen 32 von der Kodierung der Richtung her möglichen Geraden zu folgen, müssen wir DL_{min} auf 22.5° und DL_{sum} auf 45° erhöhen. Eine Richtungseinheit ist $360^\circ/32=11.25^\circ$. Verläuft ein Vektorzug unter 22.5° , so ist die Winkeldifferenz zwischen seiner Richtung und der relativen Lage des horizontalen wie auch des diagonalen Nachbarn 22.5° . Die Abbildung 5.6 zeigt das Anwachsen der Kantenmenge. Wenn wir jedoch daran denken, dass wir uns entlang jeder gedachten Geraden auf dem Raster nur wie eine Dame auf dem Schachbrett bewegen können, müssen wir zulassen, dass die Vektoren entlang einer z.B. leicht aus der Horizontalen geneigten Geraden um eine Zeile nach oben bzw. unten springen. Die Konsequenz ist eine Erhöhung von DL_{min} auf 45° . DL_{sum} muss nicht geändert werden, denn wir wissen aus den Arbeiten über die syntaktische Analyse von Geraden auf einem Raster [Brons 74], [Gaafar 77], [Rothstein Weiman 76], dass in einem sechzehntel Quadranten eine "Treppenstufe" mindestens zwei Rastereinheiten breit ist. Die Abbildung 5.7 zeigt das neue Resultat. Hierbei sind Teilgraphen mit weniger als drei Elementen unterdrückt.

Wollen wir auch leicht gekrümmten Konturen folgen, so lockern wir die Bedingung für die Parallelität der Vektoren. Bild 5.8 ist mit dem Parametersatz

$$D_R \leq 45^\circ, DL_{min} \leq 45^\circ, DL_{sum} \leq 56.25^\circ, F \geq 10$$

hergestellt. Die minimale Elementzahl F ist 10.

Um kontrastreiche Konturen zu selektieren, können wir die Neigung der Vektoren einschränken, in Bild 5.9 sind nur noch solche Vektoren enthalten, deren Neigung über 45° liegt.

Das Verfahren erlaubt es aber nicht nur, Konturen zu beschreiben, sondern auch homogene oder fein texturierte Bereiche. Dort haben die Vektoren auf Grund der Textur oder des Amplitudenrauschens eine regellose Ausrichtung. Die Parametrisierung

$$D_R \geq 33.75^\circ, DL_{\min} \geq 33.75^\circ, DL_{\text{sum}} \geq 45^\circ, F \geq 10$$

erzeugt den Graphen in Abbildung 5.10.

Als letztes Beispiel sei die Extraktion von Bereichen vorgestellt, wie sie unter anderem bei der Projektion von gewölbten Oberflächen entstehen. Aus den photometrischen Gesetzen, angewandt von Horn und anderen auf Grautonbilder [Horn 77], [Woodham 77], ergibt sich die Erwartung, dass der Gradient des Grauwertes auf das Reflexionsmaximum ausgerichtet ist. Daher müssen die Bildvektoren eine regelmässige Struktur aufweisen, in der Darstellung von Bild 5.4 ähnlich der von parallelen oder konzentrischen Höhenlinien. Parallele Vektoren erhalten wir dadurch, dass wir den Wert der minimalen Lagendifferenz nach unten beschränken, d.h. Nachbarn bevorzugen, die senkrecht zur Richtung der Vektoren liegen. Dem Bild 5.11 liegt die Parametrisierung

$$D_R \leq 33.75^\circ, DL_{\min} \geq 33.75^\circ, DL_{\text{sum}} \geq 67.5^\circ, N \geq 45^\circ, F \geq 5$$

zu Grunde. Der Schatten unter dem Rad hat einen Grauwertverlauf mit stetigem Gradienten, so dass er durch diese Parameterwahl ebenfalls als Bereich extrahiert wird.

Die drei Beispiele zeigten, dass mit einem einheitlichen Algorithmus nur durch Änderung von wenigen Parameterwerten eine Beschreibung von Bildteilen mit sehr unterschiedlicher Charakteristik erreicht werden kann. Die Parameter hängen auf so einfache Weise mit dem Bildinhalt zusammen, dass es möglich erscheint, ihre Werte für spezielle Anwendungen aus problembezogenen Angaben zu berechnen. Ich denke als Beispiel an die Kontrolle von industriell hergestellten Produkten bei der Fertigung. Hier könnten die Werte bei grober Kenntnis der Beleuchtung und der Kameraposition aus den technischen Zeichnungen abgeleitet werden. Da die Konsistenzfunktion für Knoten und Kanten des Graphen sehr eng im Algorithmus lokalisiert ist (tatsächlich besteht sie im Augenblick aus einer einzigen 'IF'-Abfrage), kann sie ebenfalls leicht spezialisiert werden.

Der Zeitbedarf für die Entfernung inkonsistenter Kanten und Knoten liegt zwischen 0.2msec und 0.4msec pro Vektor im Bildausschnitt je nachdem, wieviele Bedingungen geprüft werden müssen, bevor eventuelle Inkonsistenz festgestellt werden kann. Je schärfer die Selektion insbesondere bei den Knoten ist, denn bei den Kanten muss auf einen zweiten Vektor zugegriffen werden, um so schneller läuft das Verfahren.

Bei der Verarbeitung grösserer Bildserien kann die Reihenfolge der Abfragen so umgestellt werden, dass als erstes das selektivste Mass berechnet wird. Eine dynamische Umordnung lohnt sich für die augenblickliche Entscheidungsfunktion nicht, da der zusätzliche Aufwand den Effekt der Zeiteinsparung zunichte machen würde. Sie sollte jedoch bei komplexen Funktionen erwogen werden.

5.4 Bereichskonsistenz

In der dritten Phase des Verfahrens werden die isolierten Subgraphen besucht, ihre Merkmale berechnet und das Resultat ausgegeben. Dabei können vom Weltmodell Anforderungen gestellt werden, die dazu führen können, dass Subgraphen nicht in die Ergebnisdatei aufgenommen oder noch einmal verändert werden.

5.4.1 Graphsuche

Um die Beschreibung eines Subgraphen zu ermitteln, muss jeder seiner Knoten besucht werden, damit die Komponenten des zugeordneten Bildvektors verwertet werden können. Die Standardalgorithmen ("spanning tree", "travelling salesman" oder ähnliche), die benutzt werden, um jeden Knoten zu besuchen, arbeiten rekursiv. Dabei muss für jeden Aufruf der Prozedur Speicherplatz für die lokalen Variablen und die dynamische Verkettung der Aktivierungsbeschreibung alloziert werden. In meiner Anwendung enthält der Graph einige zehntausend Knoten und eine Komponente kann einige tausend Knoten umfassen, so dass es mir sinnvoll erschien, nach einem bezüglich Speicheraufwand und Zeitbedarf effektiven Algorithmus zu suchen.

In dem einen Knoten beschreibenden Wort sind noch 4 Bit zur freien Verfügung, 24 Bit enthalten die Vektorkomponenten und 8 Bit die unäre Kodierung der Kanten. Die Wortlänge ist 36 Bit. Jede Speicherallokation konnte dadurch vermieden werden, dass die unbesetzten 4 Bit ausgenutzt wurden. Ein Bit wird als Marke verwendet, die an einen Knoten vergeben wird, wenn er einmal besucht worden war. Die restlichen 3 Bit dienen dazu, einen "Ariadne-Faden" zu spinnen, der es dem "Graphbesucher" ermöglicht, in alle Pfade hinein zu laufen und sicher zum Ausgangspunkt zurück zu kehren. Da es höchstens acht verschiedene Richtungen gibt, aus denen man einen Knoten betreten kann, genügen glücklicherweise 3 Bit, um sich den Rückweg zu merken.

Der Algorithmus sucht das Feld sequentiell zeilenweise ab, wobei er alle Knoten überspringt, die nicht zum Graphen gehören oder die schon besucht worden waren. Findet er einen unmarkierten Knoten, muss dies der linke in der obersten Zeile von Knoten eines bisher unbesuchten Subgraphen sein. Ausgehend von diesem Startknoten wird der nächste, über eine Kante zu erreichende, unmarkierte Nachbar gesucht. Wird ein neuer Knoten betreten, werden die Vektorkomponenten zur Bereichsbeschreibung ausgenutzt, die Richtung des Rückweges eingetragen und der Knoten als besucht markiert. Gerät der Graphbesucher in eine Sackgasse, kehrt er auf dem vorgezeichneten Weg zurück, prüft aber in allen dabei passierteten Knoten, ob einer von ihnen noch einen unbesuchten Nachbarn hat. Ist dies der Fall, wird der neue Pfad so weit wie möglich verfolgt.

Im folgenden Algorithmus nimmt die lokale Variable V die Beschreibung eines Knotens aus dem Graphen FELD[Y,X] auf. Die Richtung ist mit 0 für nord, 1 für nordost, 2 für ost, usw. bis 7 für nordwest kodiert; KEINERICHTUNG ist durch den Wert -1 dargestellt. SETZMARKE markiert einen unbesuchten Knoten als besucht, TESTMARKE prüft, ob der Knoten besucht war. SETZWEG trägt die Richtung in den Knoten ein, aus der er betreten worden war, HOLEWEG liest die Rückzugsrichtung aus dem Knoten. KANTE prüft, ob der Nachbar eines Vektors im Feld über eine Kante erreichbar ist, NACHBAR liefert die Koordinaten eines Nachbarknoten ab.

```

V <- FELD[Y,X];

NEUBEREICH:
INITBEREICH;
initialisieren der bereichsbeschreibung;
XSTART <- X; YSTART <- Y;
RUECK <- NORD; VONDORT <- KEINERICHTUNG;
SAMMELN:
KOMONENTENAUSWERTUNG(V)
auswerten der komponenten von v zur bereichsbeschreibung;
SETZMARKE(V); SETZWEG(V,M); FELD[Y,X] <- V;
FOLGEN:
FOR RICHTUNG <- NORD STEP 1 UNTIL NORDWEST DO
alle acht richtungen im uhrzeigersinn;
IF RICHTUNG ≠ RUECK
der nachbar auf dem rü"ckweg war schon markiert;
AND RICHTUNG ≠ VONDORT
bei rü"ckweg von dort gerade zurü"ckgekommen;
AND KANTE(V,RICHTUNG)
kante in richtung RICHTUNG?;
AND NOT TESTMARKE(NACHBAR(V,RICHTUNG))
und nachbar noch unbesucht?;
THEN BEGIN
    RUECK <- RICHTUNG; V <- NACHBAR(V,RICHTUNG);
    VONDORT <- KEINERICHTUNG;
    GOTO SAMMELN
END;
rü"ckzug!!!;
IF X = XSTART AND Y = YSTART THEN GOTO ERSCHOEPFT;
RUECK <- HOLEWEG(V); V <- NACHBAR(V,RUECK);
VONDORT <- (RUECK + SUED) MOD 8;
von dort gerade zurü"ckgekommen;
RUECK <- HOLEWEG(V);
hole rü"ckweg aus aktuellem knoten;
GOTO FOLGEN;
ERSCHOEPFT:
alle knoten des subgraphen ausgewertet;

```


Auf dem Hinweg in einen unbekanntem Teil des Subgraphen braucht bei der Suche nach einer Fortsetzung der Nachbar nicht angeschaut zu werden, der als letzter besucht worden war. Beim Rückweg kann bei der Suche nach einem unbekanntem Zweig der Nachbar ausgelassen werden, den der Algorithmus gerade verlassen hat sowie derjenige, der auf dem vorgemerkten Rückweg für den aktuellen Knoten liegt.

5.4.2 Bereichsbeschreibung

Die Subgraphen werden durch Werte beschrieben, die einen additiven Charakter haben. Der Grund liegt in dem Wunsch, Bereiche zusammenfassen und über die Beschreibung des Resultates leicht verfügen zu können. Folgende Summen werden berechnet

- die Zahl der Knoten in der Graphkomponente,
- die Summe der X-Koordinaten und
- die Summe der Y-Koordinaten.

Aus den drei Werten lassen sich die Schwerpunktskoordinaten des Bereichs bestimmen.

Ein Mass für die Verflechtung der Knoten untereinander ist

- das Verhältnis der Kanten / Knoten im Subgraphen.

Weitere Größen sind

- die Summe der Grauwerte und
- die Summe der Quadrate der Grauwerte;
- die Summe der Neigungswinkel und
- die ihrer Quadrate.

Hieraus können der mittlere Grauwert und seine Streuung sowie die mittlere Neigung und ihre Streuung für den Bereich berechnet werden.

Zwei Häufigkeitsverteilungen werden aufgenommen,

- eine erfasst die Richtung aller Vektoren, das ergibt 32 Werte,
- die andere deren Grauwerte; die 256 möglichen Stufen werden in Gruppen zu 8 zusammengefasst, so dass sich ebenfalls 32 Werte ergeben.

Als Mass für die räumliche Ausdehnung eines Bereiches im Bild wurde folgendes Verfahren gewählt. Die Grauwertmatrix ist in Planquadrate der Grösse 16×16 eingeteilt. Die Planquadrate überlappen sich zur Hälfte, so dass jeder Vektor in genau vier von ihnen enthalten ist, ausgenommen am rechten und unteren Rand. Es sind $256/16 \times 192/16 \times 4 = 768$ Stück. Für jeden Bereich wird

- eine Liste der von ihm berührten Planquadrate

angelegt. Mit Hilfe der Liste können einfache geometrische Beziehungen wie 'über' und 'neben' zwischen Bereichen ausgerechnet werden. Der Mengen-Durchschnitt beider Listen sagt etwas über deren räumliche Nähe aus. Die Mengen-Vereinigung zweier Listen ergibt die Liste des kombinierten Bereiches. Als letztes Wertepaar sind noch

- XSTART und YSTART

in die Bereichsbeschreibung aufgenommen. Mit einem zu dem im vorigen Kapitel beschriebenen analogen Algorithmus kann jeder Vektor des Bereiches aufgesucht werden.

Einige der Werte sind in Halbworten gespeichert, so dass die angegebene Bereichsbeschreibung abgesehen von der Liste der Planquadrate 42 Worte Speicherplatz benötigt.

5.4.3 Bereichsänderung

An das Ergebnis der Bereichsbeschreibung können vom Weltmodell aus wiederum einschränkende Forderungen gestellt werden. Naheliegend ist das Unterdrücken von Bereichen, die eine zu geringe Zahl von Vektoren enthalten, als dass man ihnen eine Bedeutung zumessen könnte.

Erinnert man sich daran, dass die Bereiche unter Berücksichtigung nur von Knoten und Paaren entstanden sind, wobei eben sehr lokale Informationen verwertet wurden, ist es sinnvoll, die globaleren Merkmale der Bereiche gegenüber Zwangsbedingungen zu prüfen. Wird eine geringe Grauwertdifferenz zwischen den Knoten einer Kante gefordert, so kann über einen längeren Pfad hinweg sich doch ein beträchtlicher Grauwertunterschied aufsummieren. Das gleiche gilt für die Richtungsänderung der Vektoren. Um dem Weltmodell die Möglichkeit zu geben, Brücken zwischen zwei Häufungen im Gradientenraum (R, N, G) zu unterbrechen, kann es fordern, dass die beiden Histogramme für Grauwert und Richtung eine bestimmte Form haben. Auch hier wird ein einfaches Mass benutzt, indem das Verhältnis zwischen dem grössten und kleinsten besetzten Wert der Verteilung beschränkt wird. Sind die Merkmale eines Bereiches vollständig berechnet, prüft der Algorithmus, ob Werte für die Richtung oder für den Grauwert zu selten auftauchen. Ist das der Fall, werden sie als verboten markiert. In einem zweiten Durchlauf durch den Subgraphen werden alle Kanten gelöscht, die eine Verbindung zwischen einem Vektor mit erlaubten und einem mit verbotenen Komponenten darstellen. Dadurch kann es passieren, dass der Subgraph in mehrere Komponenten zerfällt. Deswegen werden gleichzeitig mit dem Lösen der unerwünschten Verbindungen bei allen Knoten des Subgraphen die Markierungen gelöscht. Da alle Knoten der neuen Subgraphen rechts oder unterhalb des Startknotens liegen, werden im Laufe des weiteren sequentiellen Absuchens des Vektorfeldes die Subgraphen mit Sicherheit wieder besucht.

Die Abbildung 5.12 zeigt den Graphen für unseren Beispiel-Ausschnitt mit der Einschränkung $N \geq 63^\circ$. Die Forderung, dass der niedrigste Wert der Richtungsverteilung mindestens 10% vom höchsten sein soll, erzeugt Bild 5.13. Die Abbildungen 5.14 bis 5.17 sind die Graphen für die Werte von 20% bis 50%.

Ein Beispiel für die Bewältigung eines oft schwer zu lösenden Problems, nämlich die Trennung von Schatten und Gegenstand, zeigt der Graph 5.19. Er beschreibt einen Ausschnitt von Bild 5.18, in dem der Kastenwagen enthalten ist. Die beiden unteren grossen Bereiche in Bild 5.19 stellen die in Fahrtrichtung linke Seitenfläche des Fahrzeuges und den Schatten auf der Strasse dar. Beide Subgraphen sind voneinander getrennt! Dabei liegt die Trennlinie ziemlich genau am unteren Rand der Karosserie. Die Räder sind in beiden Bereichen nicht enthalten!

Der Effekt kommt dadurch zustande, dass die Vektoren in

den beiden Bildteilen entgegengesetzte Richtungen haben. Das Licht fällt von oben ein, dadurch wird die leicht gewölbte Seitenfläche nach unten dunkler, die Bildvektoren zeigen nach oben. Direkt unter dem Fahrzeug ist der dunkelste Punkt erreicht und der Grauwert des Schattens muss mit zunehmender Entfernung vom Wagen abnehmen, daher sind im Schatten die Vektoren nach unten gerichtet. Die Kollinearitätsbedingung verhindert wirkungsvoll eine Brücke zwischen beiden Bereichen. Der Vorteil durch das Ausnutzen von Grauwert-Beziehungen zeigt sich ganz deutlich daran, dass die Räder vom Bereich des Schattens nicht eingemeindet wurden, obwohl sich die Grauwerte kaum unterscheiden. Die Trennung wäre mit einem Segmentationsverfahren, das sich nur auf Mittelwerte oder Streuung von Grauwerten stützt, sicherlich schwieriger gewesen.

In Fig. 5.3 sind die Häufigkeitsverteilungen der Vektorrichtung im Bereich der Seitenwand und des Schattens gegen den Winkel aufgetragen. An den Richtungshistogrammen sieht man, dass die Richtung der Vektoren in der Seitenwand um 270° liegen, das ist Norden, während der Häufungspunkt für den Bereich des Schattens bei $90^\circ = \text{Süden}$ ist.

(Da der Nullpunkt des Koordinatensystems in der linken, oberen Ecke des Bildes liegt (s. Kap. 4.1.1) und die positive y-Achse zum unteren Bildrand zeigt, ergibt sich eine Winkelzählung im Uhrzeigersinn.)

6. Bildfolgen mit bewegten Objekten

6.1 Vergleich von Bereichen

Der Vergleich von Bereichen aus zwei verschiedenen Bildern hat das Ziel, zwei symbolische Beschreibungen einander zuzuordnen, die mit demselben Gegenstand in der realen Szene, aber zu verschiedenen Zeiten zusammenhängen. Die grösste Schwierigkeit dabei ist es, die Unzuverlässigkeit der symbolischen Beschreibung zu kompensieren, d.h. zu entscheiden, ob sich in einem Teil der Szene die Bedeutung geändert hat oder ob die symbolischen Beschreibungen auf Grund der stochastischen oder systematischen Störungen der Aufnahmen unterschiedlich ausgefallen sind. Obwohl Price ein aufwendiges Segmentationsverfahren beützt [Price 76], muss er doch zugeben [Price Reddy 77, p.623]:

"If the recognition procedure were given the segments which were extracted, several errors might be expected because of the segmentations. Several ships are broken into two regions and these may be recognized as two individual ships or as no ships. Also, some of the piers are broken into several pieces and some of these pieces may resemble a ship more than a pier."

Das Zitat bezieht sich auf einen Vergleich zweier Bilder einer Hafenszene.

Muss man mit solchen Fehlern rechnen, sollte sich der Vergleich von Bereichen nicht so sehr auf dagegen empfindliche Merkmale stützen, wie etwa Form oder Fläche, die Price benutzt. Die von mir durchgeführte symbolische Beschreibung von Bildern durch Bereiche, die einen Ausschnitt eines Vektorfeldes darstellen, bietet die Möglichkeit, die innere Struktur von Bereichen zum Vergleich heranzuziehen. Nehmen wir den Fall an, dass ein Bereich in einem Bild die Oberfläche eines Gegenstandes repräsentiert. Dieselbe Oberfläche habe in einem anderen Bild zwei Bereiche als Repräsentanten. Dann sollten sich Merkmale aus beiden Bereichen des einen Bildes in dem einen Bereich des anderen Bildes wiederfinden lassen. Die im vorigen Kapitel vorgestellte Bereichsbeschreibung ist daraufhin angelegt, eine Zuordnung von Teilbereichen oder Kombinationen von Bereichen zu ermöglichen.

6.1.1 Vergleich von Merkmalen

Ein Mass für die Ähnlichkeit zweier Bereiche wird aus dem Vergleich der einzelnen Merkmale zusammengesetzt. Im einzelnen sind das Fläche, Verflechtung, Schwerpunkt, Grauwert, Neigung, Grauwertverteilung, Richtungsverteilung und die Liste der berührten Planquadrate. Die Ergebnisse des Einzelvergleiches sind normiert für alle Merkmale.

Fläche, F ist die Zahl der Vektoren

$$D_F = 1 - |F_1 - F_2| / (F_1 + F_2)$$

Verflechtung, P ist die Zahl der Paare

$$D_P = 1 - |P_1/F_1 - P_2/F_2| / (P_1/F_1 + P_2/F_2)$$

mittlerer Grauwert, G ist die Summe der Grauwerte

$$D_M = 1 - |G_1/F_1 - G_2/F_2| / (G_1/F_1 + G_2/F_2)$$

Neigung, N ist die Summe der Neigungswinkel

$$D_N = 1 - |N_1/F_1 - N_2/F_2| / (N_1/F_1 + N_2/F_2)$$

Für den Vergleich der Verteilungen kann man eine Kreuzkorrelationsfunktion einsetzen, die unabhängig von der Fläche der Bereiche ist, also nur die Form der Verteilung berücksichtigt. Das normierte Produkt hat diese Eigenschaft, ich gebe es als Beispiel für die Grauwertverteilung an.

$$H_i \text{ mit } i = 1, \dots, 32$$

seien die Werte der Verteilung.

Grauwertverteilung

$$D_{HG} = \sqrt{(\sum_i (H_{1i} * H_{2i}))^2 / (\sum_i (H_{1i}^2) * \sum_i (H_{2i}^2))}$$

Eine analoge Formel erzeugt die Werte D_{HR} für die Richtungsverteilung.

Sei Q die Liste der Planquadrate, die ein Bereich berührt, aufgefasst als Menge. Das Verhältnis des Durchschnittes zur Vereinigung der Listen von zwei Bereichen ergibt ein Mass für den Unterschied der von den Bereichen besetzten Bildausschnitte.

$$D_Q = |Q_1 \cap Q_2| / |Q_1 \cup Q_2|$$

Die Betragsstriche meinen die Mächtigkeit der beiden Mengen.

Die Werte aller Operatoren liegen zwischen 0 und 1, wobei 1 die beste und 0 keine Übereinstimmung bedeuten. Sie sind leicht berechenbar, unabhängig voneinander und enthalten kein spezielles Vorwissen über die Art der Szenen oder die Wahl der Parameter, die die Erzeugung der Bereiche beeinflusst haben.

6.1.2 Vergleich zweier Bereiche

Wir haben die Grauwertmatrix als Relation über dem durch die Kameraoptik projizierten Bild aufgefasst. Die Bildvektoren bilden eine Relation über den Grauwerten. Die Bereiche wiederum sind eine Relation über den Bildvektoren. Wir gehen nun wieder eine Ebene höher in der Hierarchie der Relationalstruktur und definieren die Identität von Bereichen aus verschiedenen Bildern einer zeitlich geordneten Folge. Sie soll die Vermutung ausdrücken, dass Bereiche dieselben Teile der Oberfläche eines realen Gegenstandes beschreiben. Vielleicht möchte ich hier betonen, dass nicht die Merkmale der Bereiche gleich sein sollen, sondern sozusagen der Verursacher einer Folge von Bereichen derselbe sein soll. Gerade bewegte Objekte erzeugen ja in der Sequenz der Bilder nicht in allen Merkmalen übereinstimmende Bereiche, zumindest deren Schwerpunkt muss sich geändert haben; oder falls sich das Objekt entlang der optischen Achse des Kameraobjektives bewegt, ändert sich die Fläche seiner Projektion, oder falls es aus einem Schatten heraustritt, ändert sich sein Grauwert, oder ...

Die Identität ist eine F-Relation, deren F-Wert aus dem Vergleich der Einzelmerkmale ermittelt wird. Sie werden durch eine F-UND Operation verknüpft, was die Erwartung ausdrückt, dass wegen der Beschränkung auf geringe Änderungen von Bild zu Bild identische Bereiche in allen Merkmalen ähnlich sind. Das F-UND wird als Minimum der Einzelvergleiche gebildet.

$$F\text{-UND} = \min_X D_X \text{ mit} \\ X \in \{F, P, M, N, HR, HG, Q\}$$

Wenn Bereiche als Repräsentanten von Grauwertübergängen erzeugt werden, hat sich besonders das Richtungshistogramm als brauchbarer Diskriminator erwiesen. Die Verwendung von Verteilungen hat den Vorteil, dass sie, wenn erst einmal die Identitätsrelation etabliert ist, durch Aufsuchen des Maximums der Kreuzkorrelation Unterschiede bei zwei Bildern oder Trends bei einer längeren Folge feststellbar machen. Gerade das Richtungshistogramm ist einer Beschreibung der Aussenform überlegen. Auch wenn die Aussenform des Bereiches gestört wird, kann die Charakteristik der Richtungsverteilung erhalten bleiben. Gleichzeitig gestattet sie trotzdem, eine Aussage über die Orientierungsänderung der korrespondierenden Bereiche zu machen. Die beiden Abbildungen 6.1 und 6.3 zeigen die Vektordarstellung eines Ausschnittes aus den Bildern 6.2 und 6.4. Das darin enthaltene Fahrzeug fährt eine Kurve. Fig. 6.1 zeigt die Histogramme der Richtungsverteilung für die beiden Ausschnitte (zur Erinnerung, die Vektoren sind in der Darstellung um 90° gedreht, s.a. Kap. 4.1.4). Die untere Verteilung ist dem späteren Zeitpunkt zugeordnet. Sie zeigt eine Verschiebung zu niedrigeren Winkeln, d.h. der Wagen hat sich entgegen dem Uhrzeigersinne gedreht.

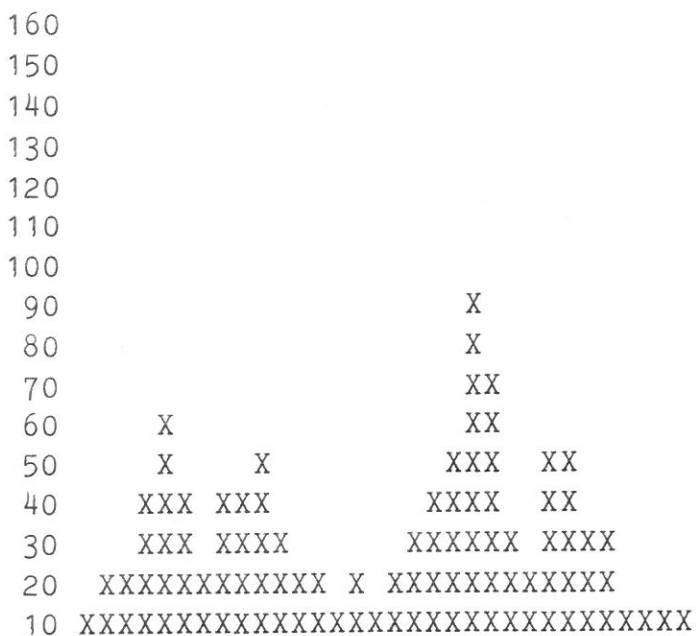
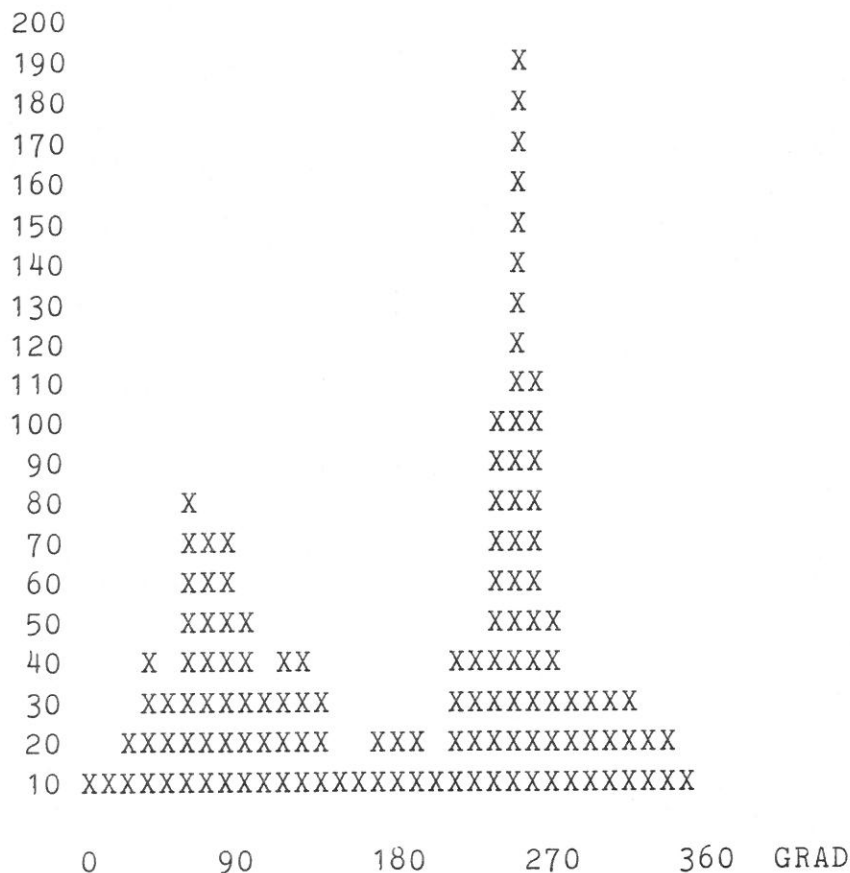


FIG. 6.1

Doch zurück zum Vergleichsalgorithmus. Es werden alle Bereiche aus einem Bild mit allen aus dem nächsten Bild verglichen. Als erstes Merkmal werden die Listen der berührten Planquadrate verglichen. Damit werden schnell alle Paarungen verworfen, die der Kontinuitätsanforderung für die Bewegung von Gegenständen widersprechen würden. Ist der Mengendurchschnitt beider Listen leer, wird nicht einmal mehr die Vereinigung berechnet, D_Q ist dann auf jeden Fall Null.

Die Kontinuitätsanforderung wird durch einen Parameter im Weltmodell beschrieben, der einen minimalen Wert für D_Q festlegt. Je näher sein Wert an eins kommt, umso grösser muss die Zahl der von beiden Bereichen gemeinsam berührten Planquadrate sein. Umso geringer wird damit die erlaubte Verschiebung eines Gegenstandes von Bild zu Bild. (Ich erinnere daran, dass die Planquadrate eine Fläche von 16×16 Vektoren haben und sich überlappen.)

Die Reihenfolge der weiteren Bereichsmerkmale ist so gewählt, dass diejenigen zuerst verglichen werden, die erfahrungsgemäss für nicht identische Bereiche den schlechtesten Wert liefern.

Für jeden Bereich wird gespeichert, welche drei Bereiche im nächsten Bild und welche drei Bereiche im vergangenen Bild seine besten Partner waren. Die Bereiche sind durch Nummern identifiziert, es werden je drei Paare von Bereichsnummer und F-Wert der Identitätsrelation zu jeder Bereichsbeschreibung hinzugefügt. Jedes Paar benötigt ein Speicherwort. Die Zahl drei ist willkürlich. Lowerre hat in seinem HARPY-Spracherkennungs-System [Lowerre 76] untersucht, welche Fehler dadurch entstehen, dass nicht alle Hypothesen für die Interpretation eines gesprochenen Satzes verfolgt werden und wieviel Rechenzeit andererseits eingespart werden kann, wenn bis auf wenige gute alle Hypothesen verworfen werden. Natürlich lassen sich seine Ergebnisse, etwa eine Aussage der Art, dass in 99% aller Fälle die richtige Interpretation immer unter den zehn besten Hypothesen ist, nicht auf diesen Fall übertragen, jedoch aber das Prinzip.

Sinkt beim Vergleich der Merkmale zweier Bereiche die Konfidenz für ein Merkmal unter den F-Wert der drittbesten Zuordnungshypothese beider Bereiche (mit 0 vorbesetzt, falls noch nicht berechnet), so wird der Vergleich sofort abgebrochen, denn die Paarung kann nicht mehr unter den drei besten Hypothesen in einem oder beiden Bereiche auftauchen. Die untere Grenze für die Rechenzeit liegt bei etwa 25msec pro Bereich.

6.1.3 Bereichsaddition

Bevor das Verfahren fortschreitet, um aus Paaren von korrespondierenden Bereichen eine ganze Kette solcher Bereiche aufzureihen, ist es schon jetzt möglich, das Ergebnis der oben geschilderten Berechnungen auf seine Konsistenz mit den einschränkenden Annahmen aus Kap. 3.1 zu überprüfen. Ich habe im folgenden die Bedingung ausgenutzt, dass sich die Ansicht eines Gegenstandes nicht abrupt ändern soll. Deswegen sollten sich die Merkmale der von seinen Oberflächen erzeugten Bereiche von Bild zu Bild nicht sprunghaft ändern. Taucht jedoch solch ein Fall auf, so ist es wert der Vermutung nachzugehen, dass eine Unstetigkeit in der Bereichszerlegung aufgetreten ist.

In meinem Verfahren ist folgendes Vorgehen realisiert: Nachdem die Identitätsrelation für Bereiche aus allen Bildpaaren der Bildsequenz berechnet worden ist, wird untersucht, ob für einen Bereich A aus einem Bild S unter den drei besten Zuordnungshypothesen zu den Bereichen B_1 , B_2 und B_3 in einem Nachbarbild T der Fall auftritt, dass A die beste Zuordnung für zwei der Bereiche, z.B. B_1 und B_3 ist. Dann wird die Beschreibung eines neuen Bereiches B_4 berechnet, indem die Werte der Bereichsmerkmale von B_1 und B_3 addiert werden (Mengenvereinigung für die Listen der Planquadrate). Die Zuordnungskonfidenz (F-Wert der Identitätsrelation) zwischen A und B_4 wird berechnet. Liegt ihr Wert höher als der von A und B_1 , so ist B_4 der beste Kandidat für A und B_4 wird in die Liste der Bereiche von Bild T aufgenommen. Wie es vorher für die vom Segmentierungsprozess erzeugten Bereiche geschehen ist, werden auch die für Bild T erzeugten Summenbereiche mit den Bereichen aus Bild U verglichen, wobei Bild U auf S und T folgt.

Das Verfahren wird sowohl in Richtung aufsteigender als auch in Richtung abnehmender Bildnummern angewendet. Damit werden die beiden Fälle erfasst, dass sich ein Bereich in dem zeitlich folgenden Bild in zwei Bereiche zerlegt und dass zwei Bereiche im späteren Bild zu einem zusammenfallen. In meinen bisherigen Experimenten konnte ich beobachten, dass etwa 1% bis 5% der in einer ganzen Bildfolge vorhandenen Bereiche solche Summenbereiche sind. Fig. 6.2 zeigt ein Beispiel für die Bereichsmerkmale zweier Bereiche aus einem Bild, des Summenbereiches und des zugeordneten Bereiches im nächsten Bild. Die Histogramme stellen die Verteilung der Vektorrichtung dar.

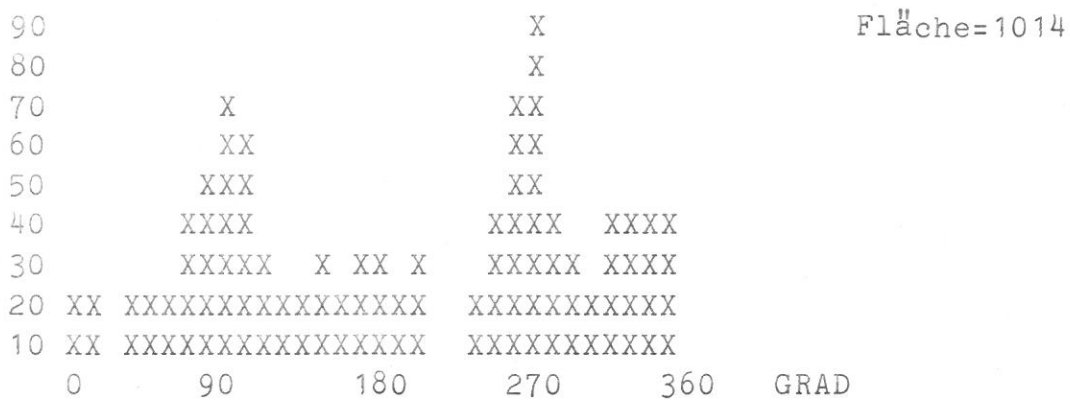
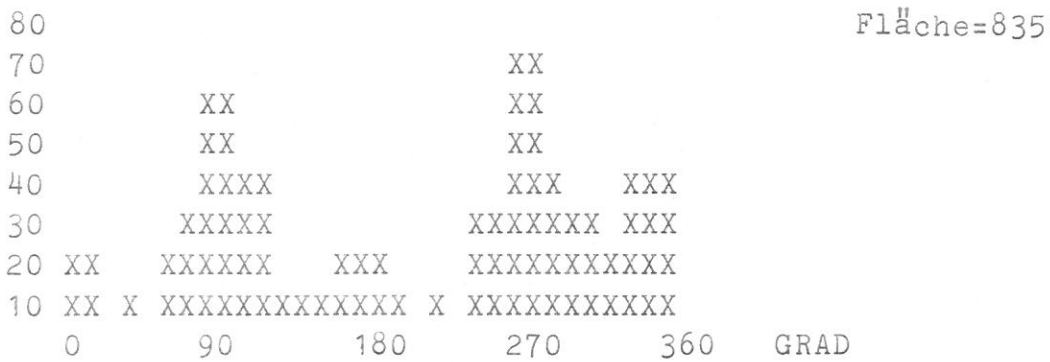
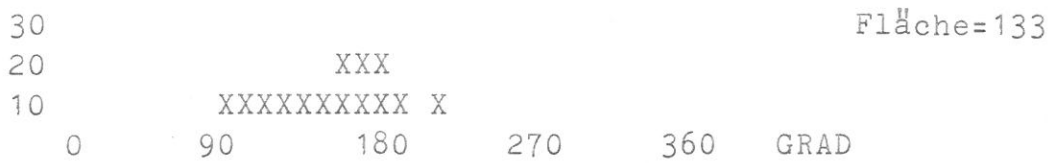
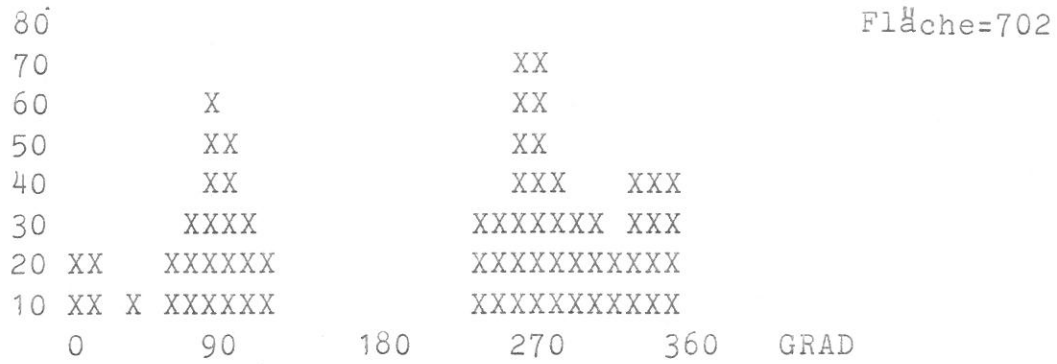


Fig. 6.2

Von oben nach unten: Fläche und Richtungsverteilung der Vektoren vom ersten Bereich, zweiten Bereich, der Summe aus beiden und dem im nächsten Bild zugeordneten Bereich.

6.2 Bereichsfolgen

Bis jetzt wurde die Berechnung von Konfidenzwerten für Bereichspaare betrachtet. In diesem Abschnitt geht es darum, die Identitätsrelation transitiv auf Folgen von Bereichen zu erweitern, die aus zeitlich geordneten Bildern stammen.

6.2.1 Transitive Identitätsrelation

Grundsätzlich wird jede Folge aus solchen Bereichen gebildet, deren gegenseitige Zuordnung (für alle zeitlich benachbarten Paare) die grösste Konfidenz unter den jeweils drei Hypothesen besitzt. D.h. eine Folge wird dann fortgesetzt, wenn ihr letzter Bereich A, der aus Bild S stammen möge, einen Zuordnungspartner B aus dem folgenden Bild T besitzt, so dass sowohl B die beste der drei Zuordnungshypothesen von A als auch A die beste Zuordnung von B ist. Es lassen sich Kriterien verwenden, die ein Abbrechen einer Folge bewirken, obwohl die obige Bedingung erfüllt ist, z.B. das Unterschreiten eines Schwellenwertes für die Zuordnungskonfidenz. Ebenso sind Kriterien denkbar, die eine Fortsetzung der Folge erlauben, obwohl obige Bedingung nicht erfüllt ist. Ich gebe später noch jeweils ein Beispiel an.

Die Beschreibung einer Folge enthält eine geordnete Liste der in ihr vertretenen Bereiche. Weiterhin ist ihre Länge vermerkt. Zwecks Berechnung der mittleren Geschwindigkeit wird die Verschiebung der Bereichsschwerpunkte in x-Richtung, d.h. die horizontale Bewegung gemessen als Anzahl der Spalten, und in y-Richtung als Anzahl der Zeilen ermittelt. Andere Merkmale der Folge können bei Bedarf aus den Daten der Bereiche errechnet werden.

Das Verfahren zum Aufbau der Folgen beginnt bei einem Bereich des ersten Bildes. Der Speicherplatz für die Datenstruktur zur Folgenbeschreibung wird alloziert und der Bereich an der ersten Stelle in die Bereichsliste aufgenommen. Ausserdem wird der Bereich in eine Menge eingetragen, in der für jedes Bild vermerkt ist, welche Bereiche aus dem Bild schon in einer Folge enthalten sind. Wird ein Bereich im nächsten Bild gefunden, der die Folge fortsetzt, so wird er ans Ende der Bereichsliste gehängt und in die Menge der schon aus dem

betreffenden Bild verwendeten Bereiche eingetragen. Ist die Folge beendet, sucht der Algorithmus im ersten Bild nach einem anderen Bereich, um wiederum eine Folge zu beginnen. Sind alle Bereiche des ersten Bildes daraufhin überprüft, ob sie als Anfang einer Folge dienen können, sucht der Algorithmus im nächsten Bild nach Bereichen, die noch nicht in einer Folge enthalten sind, und versucht mit denen als Ausgangspunkt neue Folgen aufzubauen. Alle Folgen, die mindestens drei Bereiche enthalten, werden in eine Liste aufgenommen, die anderen ignoriert. Das Verfahren endet, nachdem die Suche im drittletzten Bild der Sequenz abgeschlossen ist.

Das Bild 6.5 ist das erste einer Sequenz von 66 Bildern, die eine Strassenszene mit einem sich nach links bewegendem Fahrzeug beschreibt. Der Bildabstand beträgt 40 msec, so dass insgesamt ein Zeitabschnitt von 2.64 sec erfasst wird. Zur Bereichsextraktion wurden Parameter gewählt, die einen Bildausschnitt mit schwach geneigten Vektoren regelloser Ausrichtung und von hellem Grauton beschreiben. Zur Markierung der Identitätsrelation wurden die Schwerpunkte der Bereiche aus einer Folge durch eine schwarze Linie verbunden. Bild 6.6 zeigt das Resultat der Zuordnung für das Dach des Fahrzeugs in der ersten Hälfte der Bildsequenz, das Fahrzeug hat hier schon einen Bogen begonnen, der zum Ende der Sequenz hin enger wird. Die Abbildung 6.7 zeigt den Weg des Schwerpunktes für den Bereich des Daches über die gesamte Sequenz.

Auf Grund der perspektivischen Verzerrung hat sich die Fläche des Bereiches, der das Dach beschreibt, von etwa 1100 auf 350 Vektoren verringert. Der Grauwert nahm um 15 % zu, d.h. das Dach wurde dunkler. Der Weg des Schwerpunktes betrug 97 Zeilen und 179 Spalten, das sind etwa 1,5 Zeilen und 3 Spalten im Mittel pro Bild. Die Abbildung 6.8 ist aus dem ersten und letzten Bild der Sequenz zusammengesetzt, so dass das Fahrzeug in beiden Positionen gleichzeitig zu sehen ist.

6.2.2 Alternative Zerlegung

Für die folgenden Beispiele wurde neben der oben erwähnten eine zweite Art von Bereichen aus den Vektorfeldern extrahiert. Es wurden solche Bereiche gebildet, die Grauwertübergänge beschreiben, also aus in etwa gleich ausgerichteten Vektoren starker Neigung bestehen. Die Fläche sollte mindestens 100 Vektoren betragen. Die Abbildung 6.9 zeigt das Resultat der Folgenbildung in der Bildsequenz mit beiden Bereichstypen. Zu erkennen ist die Spur der Seitenwand des fahrenden Wagens sowie darunter die seines Schattens. Der Schatten wird zeitweise durch die Baumkrone im Vordergrund verdeckt. Die Unruhe gegen Ende der Sequenz rührt daher, dass die Seitenwand des Fahrzeugs und der Seitenschatten auf Grund der Drehung verschwinden; dafür wird die Rückwand in der Projektion grösser und die rechte Seitenwand taucht auf.

Interessant ist noch das parallel zur Anfangsposition des Kastenwagens parkende Fahrzeug. Sein Dach bildet einen sehr stabilen Bereich, dessen Schwerpunkt kaum seine Lage verändert. Der Fahrzeugschatten jedoch wird in seinem vorderen Teil zeitweise vom Dach des passierenden Kastenwagens verdeckt, so dass eine horizontale Verschiebung des Schattenschwerpunktes erfolgt.

Als ein Beispiel für die in Abschnitt 6.2.1 erwähnte Möglichkeit, die Fortsetzung einer Folge durch Kriterien zu verhindern, die aus den Bereichsmerkmalen berechenbar sind, zeige ich die Verwendung der Schwerpunktsverschiebung. In den Abbildungen 6.10 bis 6.12 sind die Identitätsrelationen für den Fall dargestellt, dass sich die Schwerpunktskoordinaten zweier in einer Folge benachbarten Bereiche um weniger als 10, 6 bzw. 3 unterscheiden. Als Metrik wurde die Summe aus den Absolutwerten der Differenz in x- und y-Richtung gewählt. Wie zu erwarten, werden die Spuren zunehmend begradigt; dort wo vorher Ausbuchtungen waren, entstehen kleine Lücken.

(Bemerkung: Jetzt ist die Identitätsrelation keine transitive Erweiterung der Bereichsidentität mehr.)

Gleichzeitig entstehen aber auch am Anfang der Sequenz Unterbrechungen. Im Bild 6.11 rühren die Lücken daher, dass die systematische Schwerpunktsverschiebung, hervorgerufen durch die Bewegung des Fahrzeuges, zusammen mit einer zufälligen Verschiebung die gesetzte Schwelle gerade übersteigt. In der Abbildung 6.12 fehlt der erste Teil der Spur des Daches, da hier die Fahrzeuggeschwindigkeit (in der Projektion) zu gross ist. Bemerkenswert sind in diesem Bild zwei Phänomene. Rechts vom Baum ist ein kurzes Stück der Schattenspur erhalten

geblieben. Die Ursache dafür ist, dass die Vorderfront des Schattens zunehmend vom Baum verdeckt wird und damit die Schwerpunktgeschwindigkeit verringert wird. Einen ähnlichen Grund hat die Existenz einer kurzen Spur in der Seitenwand des Wagens am Beginn der Sequenz. Hier verlängert sich der Bereich der Seitenwand solange, bis das Fahrzeug vollständig im Bild sichtbar ist, dadurch wird die Schwerpunktsverschiebung eine kurze Weile verzögert.

6.3 Kombination von Folgen

Die zuletzt gezeigten Abbildungen legen nahe, nach Wegen zu suchen, um die Lücken in den Spuren zu überbrücken. Das bedeutet, diejenigen Bereiche in einer einzigen Folge zusammenzufassen, die von demselben Oberflächenbereich eines Gegenstandes verursacht worden sind. Damit wird der im vorigen Abschnitt durch den Gebrauch von Begriffen wie "Lücke" oder "Unterbrechung" vorweggenommene semantische Zusammenhang zwischen Folgen im Programmsystem berechenbar gemacht.

6.3.1 Zeitliche Konkatenation

Das Auftreten von Unterbrechungen in den Spuren kann drei Ursachen haben

- Es wurden während eines Abschnittes der Beobachtungszeit keine Bereiche von dem betreffenden Teil der Oberfläche eines Gegenstandes erzeugt. Ein Grund dafür kann sein, dass dieser Teil der Oberfläche von einem anderen Objekt verdeckt wurde. Ein Beispiel ist die Unterbrechung der Schattenspur in Abb. 6.10.
- Störungen können verhindern, dass eine eindeutige Zuordnung zweier Bereiche gefunden wird. So kann sich zum Beispiel durch eine plötzlich auftretende Spiegelreflexion die Charakteristik eines Bereiches, der von dem reflektierenden Teil der Gegenstandsoberfläche herrührt, so stark ändern, dass schon auf der Ebene des Bereichsvergleiches die Identitätsrelation nicht mehr verifiziert werden kann.
- Besteht auf der Bereichsebene die Identitätsrelation, so kann ihre Erweiterung auf Bereichsfolgen durch zusätzliche

Bedingungen eingeschränkt werden, wie im vorigen Abschnitt an einem Beispiel gezeigt wurde.

Um einzelne Folgen miteinander in Beziehung zu setzen, müssen die Eigenschaften ausgewählt werden, auf deren Basis ein Vergleich geschehen soll. Auch hier habe ich wieder versucht, die Brauchbarkeit des Ansatzes an einer Minimallösung zu demonstrieren.

Ausgehend von der Einschränkung (Kap. 3.1), dass sich die Bedeutung einer Szene nur allmählich ändern soll, habe ich als Repräsentanten einer Folge ihren ersten und ihren letzten Bereich gewählt. Für alle Folgen werden die Merkmale des letzten Bereiches mit denen des ersten Bereiches aller derjenigen Folgen verglichen, die zu einem späteren Zeitpunkt beginnen. Der Vergleich beider Bereiche geschieht auf genau dieselbe Art, wie in Kapitel 6.1 beschrieben. Die einzige Abweichung ist, dass auf das Kriterium D_Q zur Bewertung der Bereichsverschiebung verzichtet wurde. Ist die Ähnlichkeit für alle möglichen Paarungen berechnet, werden in der Reihenfolge abnehmender Konfidenz Paare von Folgen vereinigt, indem die spätere Folge an die frühere angehängt wird. Dabei werden die Merkmale der Folgen aufsummiert bzw. die Listen der Bereiche konkateniert. Die resultierenden Folgen sind so ebenfalls eine Erweiterung der paarweisen Bereichsidentität, wobei die Bedingung aufgehoben ist, dass Bereichspaare immer aus zeitlich benachbarten Bildern stammen müssen. In diesem Sinne verstehe ich die Zusammenfassung von Folgen als das in Kapitel 6.2.1 angekündigte Beispiel für die Fortsetzung von Folgen.

Die Abbildung 6.13 zeigt das Ergebnis einer Konkatenation von Folgen. Dabei sind die Strecken zwischen den Bereichsschwerpunkten jeder Folge schwarz dargestellt, wenn sich die Bildnummern der Bereiche nur um eins unterscheiden. Stammen die Bereiche nicht aus benachbarten Bildern, ist der Weg zwischen ihren Schwerpunkten weiss gekennzeichnet. Eigenschaften von Folgen, die man in komplexeren Situationen hinzuziehen könnte, wären die mittlere Schwerpunktsgeschwindigkeit oder die zeitliche bzw. räumliche Nähe der Folgenenden. Als weitere Möglichkeit könnten geometrische Beziehungen zwischen Folgen ausgenutzt werden.

6.3.2 Räumliche Selektion

Habe ich im vorigen Abschnitt die Konkatenation von Folgen als Beziehung zwischen zeitlich nicht unmittelbar benachbarten Beschreibungen von Objekten vorgestellt, so will ich als Abschluss in der Hierarchie der Relationalstrukturen die Auswertung einer geometrischen Beziehung zwischen Folgen zeigen. Das Beispiel soll nur den ersten Schritt in Richtung einer weiterführenden Interpretation der Szene andeuten.

Als Prototyp des Kastenwagens wurde folgende Beschreibung gewählt: Das Fahrzeug besteht aus einem Dach, das durch Bereiche des Typs 1 (homogener Grauwert, regellose Ausrichtung der Vektoren) beschrieben wird. Eine Seitenwand befindet sich unterhalb des Daches und wird durch Bereiche des Typs 2 (Grauwertgradient, gleichgerichtete Vektoren) beschrieben. Die Relation 'unter' wird durch die Differenzbildung der y-Koordinaten zweier Bereiche berechnet. In der Abbildung 6.14 sind die beiden Folgen dargestellt, die diesen Bedingungen genügen.

Bei der Selektion der Bereiche kann man nun zwei Strategien einschlagen. Einerseits können alle Bereichspaare in jedem Bild überprüft werden, ob sie die Bedingungen des Prototyps erfüllen und aus den so selektierten Bereichen werden Folgen gebildet. Andererseits können erst alle Bereichsfolgen berechnet werden und die Übereinstimmung mit dem Prototyp an einem Paar von Repräsentanten der Folgen, z.B. dem ersten Bereich jeder Folge, geprüft werden. Ein Entscheidungskriterium zur Auswahl zwischen beiden Strategien könnte der Rechenaufwand sein. Im vorliegenden Falle bietet sich das erstere Vorgehen an, da die vorgegebenen Bedingungen nur die Überprüfung des Bereichstyps und die Berechnung einer Differenz erfordern. Durch die Vorselektion wird die Zahl der Bereiche vor der Berechnung der Identitätsrelation vermindert, dort also kombinatorischer Aufwand eingespart.

Es mag aber auch Prototypen geben, die eine aufwendigere Berechnung erfordern, so dass die zweite Strategie vorzuziehen ist. Ich wollte mit dem vorgestellten Beispiel darauf hinweisen, dass es auf diesem Weg möglich ist, die Berechnung zeitlicher Relationen mit der Berechnung geometrischer, topologischer oder - auf höherem Abstraktionsniveau - semantischer Relationen zu mischen, um ein gutes Interpretationsergebnis schnell zu erreichen.

7. Ausklang

Ich bin der Meinung, dass die hier vorgestellten Verfahren eine solide Basis für die Interpretation natürlicher Szenen mit sich bewegenden Objekten bilden. Besonders die Extraktion von Bereichen aus einem Vektorfeld mit graphtheoretischen Methoden hat sich als effektives und flexibles Verfahren erwiesen. Die vorgeschlagene hierarchische Relationalstruktur

- Grauwert
- Bildvektoren
- Vektorpaare
- Bereiche
- Bereichspaare
- Bereichsfolgen
- Kombinationen von Folgen

ist nach oben ausbaufähig und kann in ein weiterführendes Interpretationssystem, wie z.B. eine Modifikation des von Bertelsmeier und mir konzipierten [Bertelsmeier Radig 77a], eingegliedert werden.

Ich möchte noch einen kurzen Überblick über Verbesserungen des bestehenden Systems geben. Die Digitisierungsapparatur hat sich bewährt, hinzugekommen ist inzwischen ein Lichtzeiger, mit dessen Hilfe die Eingabe von Bildkoordinaten in den Rechner möglich ist. Als Anwendung kommt zum Beispiel die interaktive Unterstützung bei der Formulierung von Prototypen in Frage. So könnten, um hier bei dem Beispiel aus dem vorigen Kapitel zu bleiben, das Dach des Wagens und die Seitenwand umfahren werden und aus der Statistik über die Komponenten der in diesen Ausschnitten enthaltenen Bildvektoren die Parameter für die Bereichsextraktion genommen werden.

In dem gesamten Verarbeitungsprozess einer Bildserie benötigt die Umformung der Grauwertmatrix in ein Feld von Bildvektoren die meiste Rechenzeit. Hier muss untersucht werden, ob die Bestimmung der Ebenenkoeffizienten nach der Methode der kleinsten Fehlerquadrate durch ein weniger aufwendiges Verfahren ersetzt werden kann; z.B. könnte jede Ebenengleichung mit Hilfe dreier Messwerte aus der Grauwertmatrix bestimmt werden.

Die Berechnung der Bildvektoren sowie die Bereichsextraktion werden auf Prozessrechner verlagert. Hier, wie auch bei der gleichzeitigen Extraktion von Bereichen verschiedener Art, wird eine Parallelverarbeitung den Zeitaufwand verkürzen. Die apparativen Voraussetzungen sind

gegeben und die programmtechnischen werden gerade erarbeitet.

Eine weitere Möglichkeit der Aufwandsreduktion besteht darin, die Bearbeitung auf diejenigen Bildteile zu beschränken, in denen sich die interessierende Handlung abspielt. Als Hilfsmittel wäre die in Kapitel 3.3 skizzierte Kosten/Nutzen-Analyse einzusetzen. Um solche Bildteile festzustellen, könnte das von Jain und Nagel entwickelte Verfahren verwendet werden [Jain Nagel 77].

Weiterhin wird es notwendig sein, bei der Kombination von Bereichsfolgen Wissen über die dreidimensionale Natur der Szene heranzuziehen. Sowohl die Verdeckung von Objekten als auch das Verschwinden oder Erscheinen von Teilen der Gegenstandsoberfläche ist ein Phänomen, das der Dreidimensionalität natürlicher Szenen entspringt. Ein geeignetes, schon weit entwickeltes Werkzeug scheinen mir Produktionensysteme zu sein. In ihren Regeln könnte man oben erwähntes Wissen formulieren und ausserdem die in Kapitel 3.1 angegebenen Einschränkungen explizit als Interpretationsunterstützung zur Verfügung stellen.

Literaturverzeichnis

IJCAI = International Joint Conference on Artificial
Intelligence
IJCPR = International Joint Conference on Pattern Recognition

[Aggarwal Duda 75]

J. K. Aggarwal, R. O. Duda
Computer Analysis of Moving Polygonal Images
IEEE Trans. Comp. C-24 (1975) 966-976

[Akin Reddy 77]

O. Akin, R. Reddy
Knowledge Acquisition for Image Understanding Research
Computer Graphics and Image Processing 6 (1977) 307-334

[Albrecht et al. 75]

M. Albrecht, B. Fischer, D. Platz
Anwendungen spezieller Bildverarbeitungsverfahren zur
Erkennung von Organgrenzen und Tumoren in
lokalisationsdiagnostischen Leberszintigrammen in einem
interpretativen Dialogsystem
Fachbereich Informatik, Mitteilung IfI-HH-M-17/75, Univ.
Hamburg 1975

[Ambler et al. 75]

A. P. Ambler, H. G. Barrow, C. M. Brown, R. M. Burstall,
R. J. Popplestone
A Versatile System for Computer-Controlled Assembly
Artificial Intelligence 6 (1975) 129-156

[Arbib Riseman 76]

M. A. Arbib, E. M. Riseman
Computational Techniques in Visual Systems. Part I The
Overall Design
COINS Technical Report 76-10, Dept. Comp. and Inf. Sc.,
Univ. of Massachusetts at Amherst, Amherst/Mass., Juli
1976
(Proc. IEEE, im Druck)

[Arking 67]

A. Arking

Processing of TIROS Cloud Cover Pictures on a Digital
Computerin F. Gruenberger (ed.) Computer Graphics, Thomson Book,
Washington D.C., Academic Press, London 1967

[Badler 75]

N. I. Badler

Temporal Scene Analysis: Conceptual Descriptions of
Object Movements

Dissertation, Report 80, Univ. Toronto, Februar 1975

[Baird 76]

M. L. Baird

An Application of Computer Vision to Automated IC Chip
Manufacture

IJCPR-76, 3-7

[Baird Kelly 74a]

M. L. Baird, M. D. Kelly

Recognizing Objects by Rules of Inference on Sequentially
Thresholded Gray-Level Pictures

Computer Graphics and Image Processing 3 (1974) 1-22

[Baird Kelly 74b]

M. L. Baird, M. D. Kelly

A Paradigm for Semantic Picture Recognition

Pattern Recognition 6 (1974) 61-74

[Bajcsy Lieberman 74]

R. Bajcsy, L. I. Lieberman

Computer Description of Real Outdoor Scenes

IJCPR-74, 174-179

[Bajcsy Lieberman 76]

R. Bajcsy, L. I. Lieberman

Texture Gradient as a Depth Cue

Computer Graphics and Image Processing 5 (1976) 52-67

[Bajcsy Soroka 77]

R. K. Bajcsy, B. I. Soroka

Steps Towards the Representation of Complex
Three-Dimensional Objects

IJCAI-77, 596

[Baker 76]

H. H. Baker
Building Models of Three-Dimensional Objects
Master Thesis, Univ. of Edinburgh, 1976

[Baker 77]

H. H. Baker
Three-Dimensional Modelling
IJCAI-77, 649-655

[Barrow et al. 72]

H. G. Barrow, A. P. Ambler, R. M. Burstall
Some Techniques for Recognizing Structures in Pictures
in S. Watanabe (ed.), Frontiers of Pattern Recognition,
Academic Press, New York 1972, pp. 1-29

[Barrow Popplestone 71]

H. G. Barrow, R. J. Popplestone
Relational Descriptions in Picture Processing
Machine Intelligence 6 (B. Meltzer, D. Michie, eds.)
University Press Edinburgh, 1971, pp. 377-396

[Baumgart 74]

B. G. Baumgart
Geometric Modeling for Computer Vision
Report STAN-CS-74-463, Comp. Sc. Dept., Stanford Univ.,
Okt. 1974

[Bertelsmeier Radig 77a]

R. Bertelsmeier, B. Radig
Kontextunterstützte Analyse von Szenen mit bewegten
Objekten
GI/NTG Fachtagung Digitale Bildverarbeitung (H.-H.
Nagel, ed.) München, 28.-30. März 1977, Informatik
Fachberichte Bd. 8 pp. 101-128, Springer,
Berlin-Heidelberg-New York 1977

[Bertelsmeier Radig 77b]

R. Bertelsmeier, B. Radig
Context Guided Analysis of Scenes With Moving Objects
Bericht IfI-HH-B-41/77, Fachbereich Informatik, Univ.
Hamburg, Februar 1977

[Bertelsmeier]

R. Bertelsmeier
Struktur eines Systems zur Erkennung von Objekten in
Bildern und Bildfolgen
Fachbereich Informatik, Univ. Hamburg, in Vorbereitung

[Bolles 76]

R. C. Bolles
Verification Vision Within a Programmable Assembly System
Report STAN-CS-77-591, Dissertation, Dept. of Comp. Sc.,
Stanford Univ. 1976

[Bonde 78]

T. Bonde
Untersuchungen zur dreidimensionalen Modellierung von
bewegten Objekten durch Analyse von Formveränderungen der
Objektbilder in TV-Aufnahmefolgen
Fachbereich Informatik, in Vorbereitung

[Bowker 74]

J. K. Bowker
Edge Vector Image Analysis
IJCPR-74, 520-524

[Braid 75]

I. C. Braid
The Synthesis of Solids Bounded by Many Faces
C. ACM 18 (1975) 209-216

[Brayer Fu 76]

J. M. Brayer, K. S. Fu
Application of a WEB Grammar Model to an Earth Resources
Satellite Picture
IJCPR-76, 405-410

[Brice Fennema 70]

C. R. Brice, C. L. Fennema
Scene Analysis Using Regions
Artificial Intelligence 1 (1970) 205-226

[Bron Kerbosch 73]

C. Bron, J. Kerbosch
Finding All Cliques of an Undirected Graph
C. ACM 16 (1973) 525-577

- [Brons 74]
R. Brons
Linguistic Methods for the Description of a Straight Line
on a Grid
Computer Graphics and Image Processing 3 (1974) 48-62
- [Burr Chien 77]
D. J. Burr, R. T. Chien
A System for Stereo Computer Vision with Geometric Models
IJCAI-77, 583
- [Budy Dreschler 76]
B. Budy, L. Dreschler
Untersuchungen zur Identifikation eines bewegten Objektes
(PKW) in der Video-Bildsequenz einer Strassenszene
Mitteilung IfI-HH-M-33/76, Fachbereich Informatik, Univ.
Hamburg, 1976
- [Chow Aggarwal 77]
W. K. Chow, J. K. Aggarwal
Computer Analysis of Planar Curvilinear Moving Images
IEEE Trans. Comp. C-26 (1977) 179-185
- [Chien Jacobus 76]
R. T. Chien, C. Jacobus
Directional Derivatives in Computer Image Processing
IJCPR-76, 684-688
- [Chien Jones 75]
R. T. Chien, V. C. Jones
Acquisition of Moving Objects and Hand-Eye Coordination
IJCAI-75, 737-741
- [Chang Wang 75]
S.-K. Chang, Y. R. Wang
Three-Dimensional Object Reconstruction from Orthogonal
Projections
Pattern Recognition 7 (1975) 167-176
- [Clowes 71]
M. B. Clowes
On Seeing Things
Artificial Intelligence 2 (1971) 79-116

[Davis et al. 77]

R. Davis, B. Buchanan, E. Shortliffe
Production Rules as a Representation for a
Knowledge-Based Consultation Program
Artificial Intelligence 8 (1977) 15-45

[Davis Rosenfeld 76]

L. S. Davis, A. Rosenfeld
Applications of Relaxation Labeling, 2: Spring-Loaded
Template Matching
Techn. Report 440, Comp. Sc. Center, Univ. of Maryland,
1976

[Davis 73]

L. S. Davis
A Survey of Edge Detection Techniques
Report TR-273, Univ. of Maryland, 1973
Computer Graphics and Image Processing 4 (1975) 248-270

[Demuth 76]

H. B. Demuth
Feature Extraction through Least Squares Fit to a Simple
Model
IJCPR-76, 37-41

[Dierckx 77]

P. Dierckx
An Algorithm for Least-Squares Fitting of Cubic Spline
Surfaces to Functions on a Rectilinear Mesh over a
Rectangle (Algorithm 008)
J. of Comp. and Appl. Math. 3 (1977) 113-129

[Dreschler 77]

L. Dreschler
Isolierung sich bewegender Objekte in TV-Bildfolgen unter
Ausnutzung der Formkonstanz bei variabler Objektbildgrösse
Diplomarbeit, Fachbereich Informatik, Univ. Hamburg, Aug.
1977

[Duda Hart 73]

R. O. Duda, P. E. Hart
Pattern Classification and Scene Analysis
John Wiley & Sons, New York, 1973

[Duda et al. 76]

R. O. Duda, P. E. Hart, N. J. Nilson
 Subjective Bayesian Methods for Rule-Based Inference
 Systems
 Nat. Comp. Conf. July 7-10, 1976, New York, AFIPS Conf.
 Proceedings 45 (1976) 1075-1082

[Duda Nitzan 76]

R. O. Duda, D. Nitzan
 Low Level Processing of Registered Intensity and Range
 Data
 IJCPR-76, 598-601

[Ehrich Foith 76]

R. W. Ehrich, J. P. Foith
 Representation of Random Waveforms by Relational Trees
 IEEE Trans. Comp. C-25 (1976) 725-736

[Ellis McLain 77]

T. M. R. Ellis, D. H. McLain
 A New Method of Cubic Curve Fitting Using Local Data [E2
] (Algorithm 514)
 ACM Trans. Math. Softw. 3 (1977) 175-179

[Ernst et al. 76]

D. Ernst, B. Bargel, F. Holdermann
 Processing of Remote Sensing Data by a Region Growing
 Algorithm
 IJCPR-76, 679-683

[Falk 70]

G. Falk
 Computer Interpretation of Imperfect Line Data as a
 Three-Dimensional Scene
 Report AIM 132, Stanford Univ. 1970
 Scene Analysis Based on Imperfect Edge Data
 IJCAI-71, 8-16
 Interpretation of Imperfect Line Data as a
 Three-dimensional Scene
 Artificial Intelligence 3 (1972)

[Foith 77]

E. Foith
 Symbolische Repräsentation von Grauwertbildern für
 Szenenanalysen
 Institut für Informationsverarbeitung in Technik und
 Biologie, Karlsruhe, März 1977

[Freuder 76]

E. C. Freuder
Synthesizing Constraint Expressions
Memo AI-370, MIT, Juli 1976

[Fu 73]

K. S. Fu
Stochastic Languages for Picture Analysis
Computer Graphics and Image Processing 2 (1973) 433-453

[Fu 74]

K. S. Fu
Syntactic Methods in Pattern Recognition
Academic Press, New York 1974

[Gaafar 77]

M. Gaafar
Convexity Verification, Block-Chords, and Digital
Straight Lines
IJCPR-76, 514-518
Computer Graphics and Image Processing 6 (1977) 361-370

[Garvey 76]

T. O. Garvey
An Experiment with a System for Locating Objects in
Multisensory Images
IJCPR-76, 567-575

[Gimelfarb et al. 75]

G. Gimelfarb, E. Kushner, V. Rybak
A "Hand-Eye" Robot Simulation System
IJCAI-75, 766-770

[Guzman 68]

A. Guzman
Computer Recognition of Three Dimensional Objects in a
Visual Scene
(Dissertation) Report MAC TR 59, MIT, Cambridge/Mass.
1968

[Guzman 71]

A. Guzman
Analysis of Curved Line Drawings Using Context and Global
Information
Machine Intelligence 6 (B. Meltzer, D. Michie, eds.),
University Press Edinburgh 1971, pp. 325-375

[Hannah 74]

M. J. Hannah
 Computer Matching of Areas in Stereo Images
 Dissertation, Memo AIM 239, Stanford Univ., Juli 1974

[Hall et al. 72]

D. J. Hall, R. M. Endlich, D. E. Wolf, A. E. Brain
 Objective Methods for Registering Landmarks and
 Determining Cloud Motions from Satellite Data
 IEEE Trans. Comp. C-21 (1972) 768-776

[Hart et al. 68]

P. E. Hart, N. Nilson, B. Raphael
 A Formal Basis for the Heuristic Determination of Minimum
 Cost Paths
 IEEE Trans. Sys. Sci. and Cyber. SSC-4 (1968) 100-107
 Correction to "A Formal Basis for the Heuristic
 Determination of Minimum Cost Paths"
 SIGART Newsletter No. 37, Dec. 1972, 28-29
 siehe auch
 D. Gelperin
 On the Optimality of A*
 Artificial Intelligence 8 (1977) 69-76

[Hanson Riseman 76]

A. R. Hanson, E. M. Riseman
 A Progress Report on VISIONS: Representation and Control
 in the Construction of Visual Models
 COINS Techn. Report 76-9, Comp. and Inf. Sc., Univ. of
 Massachusetts at Amherst, Amherst/Mass., Juli 1976

[Hewitt 69]

C. Hewitt
 PLANNER: A Language for Manipulating Models and Proving
 Theorems in a Robot
 IJCAI-69

[Hewitt 71]

C. Hewitt
 Procedural Embedding of Knowledge in PLANNER
 IJCAI-71, 167-182

[Hewitt 72]

C. Hewitt
 Description and Theoretical Analysis (Using Schemata)
 of PLANNER. A Language for Proving Theorems and
 Manipulating Models in a Robot

Dissertation, Dept. of Math. AI-TR-258, MIT,
Cambridge/Mass., 1972

[Hogg 76]

D. C. Hogg

Visual Perception in a Changing Environment: A Paradigm
for Real-Time Analysis

Univ. of Western Ontario, Dept. of Comp. Sc., Report 29,
Okt. 1976

[Holdermann Kazmierczak 72]

F. Holdermann, H. Kazmierczak

Preprocessing of Gray-Scale Pictures

Computer Graphics and Image Processing 1 (1972) 66-80

[Holdermann 73]

F. Holdermann

Processing of Gray Scale Pictures

NTG/GI Fachtagung Cognitive Verfahren und Systeme,

Hamburg 11.-13. April 1973, (T. Einsele, W. Giloi, H.-H.

Nagel, eds.) Springer Berlin-Heidelberg-New York 1973

[Horowitz Pavlidis 74]

S. L. Horowitz, T. Pavlidis

Picture Segmentation by a Directed Split-and-Merge

Procedure

IJCPR-74, 424-433

[Horn 74]

B. K. P. Horn

Determining Lightness from an Image

Computer Graphics and Image Processing 3 (1974) 277-299

[Horn 75]

B. K. P. Horn

Obtaining Shape from Shading Information

in P. H. Winston (ed.), The Psychology of Computer

Vision, McGraw-Hill, New York, 1975, pp. 115-155

[Horn 77]

B. K. P. Horn

Understanding Image Intensities

Artificial Intelligence 8 (1977) 201-231

[Huang Fu 72]

T. Huang, K. S. Fu
Stochastic Syntactic Analysis for Programmed Grammars and
Syntactic Pattern Recognition
Computer Graphics and Image Processing 1 (1972) 257-283

[Hueckel 71]

M. Hückel
An Operator which Locates Edges in Digitized Pictures
J. ACM 18 (1971) 113-125
und
A Local Visual Operator which Recognizes Edges and Lines
J.ACM 20 (1973) 634-647

[Huffman 71]

D. A. Huffman
Impossible Objects as Nonsense Sentences
Machine Intelligence 6 (B. Meltzer, D. Michie, eds.),
University Press Edinburgh, 1971, pp. 295-324

[Ishi Nagata 76]

M. Ishi, T. Nagata
Feature Extraction of Three Dimensional Objects and
Visual Processing in a Hand-Eye System Using Laser
Tracker
Pattern Recognition 8 (1976) 229-237

[Jain et al. 77]

R. Jain, D. Militzer, H.-H. Nagel
Separating Non-Stationary from Stationary Scene
Components in a Sequence of Real World TV-Images
IJCAI-77, 612-618

[Jain Nagel 77]

R. Jain, H.-H. Nagel
Analysing a Real World Scene Sequence Using Fuzziness
IEEE Conference on Decision and Control, New Orleans,
Dec. 7-9, 1977

[Kelly 70]

M. D. Kelly
Visual Identification of Peoples by Computer
Memo AI-130, Comp. Sc. Dept. Stanford Univ. 1970

[Kelly 71]

M. D. Kelly
Edge Detection in Pictures by Computer Using Planning
Machine Intelligence 6 (B. Meltzer, D. Michie, eds.),
Edinburgh University Press, 1971, pp. 379-409

[Kiessling 76]

A. Kiessling
A Fast Scanning Method for Three Dimensional Scenes
IJCPR-76, 586-589

[Klinger Dyer 76]

A. Klinger, C. R. Dyer
Experiments on Picture Representation Using Regular
Decomposition
Computer Graphics and Image Processing 5 (1976) 68-105

[Kraasch Zach 78]

R. Kraasch, W. Zach
SERF - Ein sequentieller Segmentierungsalgorithmus
Bericht Fachbereich Informatik IfI-HH-M-59/78, Univ.
Hamburg 1978

[LeFaivre 74]

R. A. LeFaivre
The Representation of Fuzzy Knowledge
J. of Cybernetics 4 (1974) 57-66

[LeFaivre 76]

R. A. LeFaivre
Procedural Representation in a Fuzzy Problem-Solving
System
Nat. Comp. Conf. 7.-10. Juni 1976, New York, AFIPS Conf.
Proceedings 45 (1976) 1069-1074

[Lesser Erman 77]

V. R. Lesser, L. D. Erman
A Retrospective View of the HEARSAY-II Architecture
IJCAI-77, 790-800

[Levine et al. 73]

M. D. Levine, D. A. O'Handley, G. M. Yagi
Computer Determination of Depth Maps
Computer Graphics and Image Processing 2 (1973) 131-150

[Leese et al. 70]

J. A. Leese, C. S. Novak, V. R. Taylor
The Determination of Cloud Pattern Motions from
Geosynchronous Satellite Image Data
Pattern Recognition 2 (1970) 272-292

[Limb Murphy 75]

J. O. Limb, J. A. Murphy
Estimating the Velocity of Moving Images in Television
Signals
Computer Graphics and Image Processing 4 (1975) 311-327

[Lowerre 76]

B. T. Lowerre
The HARPY Speech Recognition System
Dissertation, Dept. Comp. Sc., Carnegie-Mellon Univ.,
April 1976

[Mackworth 73]

A. K. Mackworth
Interpreting Pictures of Polyhedral Scenes
Artificial Intelligence 4 (1973) 121-137

[Mackworth 77]

A. K. Mackworth
Consistency in Networks of Relations
Artificial Intelligence 8 (1977) 99-118

[Marr 75]

D. Marr
Early Processing of Visual Information
Memo AI-340, MIT Cambridge/Mass., Dez. 1975
Phil. Trans. Royal Society, Series B, 275 (1976) 483-524

[Marr 76]

D. Marr
Analysis of Occluding Contour
Memo AI-372, MIT Cambridge/Mass., Okt. 1976

[Marr Nishihara 76]

D. Marr, H. K. Nishihara
Representation and Recognition of the Spatial
Organization of Three Dimensional Shapes
Memo AI-377 MIT Cambridge/Mass., Aug. 1976

[Martelli 72]

A. Martelli

Edge Detection Using Heuristic Search Methods

Comp. Graphics and Image Processing 1 (1972) 169-182

[Martelli 76]

A. Martelli

An Application of Heuristic Search Methods to Edge and
Contour Detection

C. ACM 19 (1976) 73-83

[Martelli 77]

A. Martelli

On the Complexity of Admissible Search Algorithms

Artificial Intelligence 8 (1977) 1-13

[Martin Aggarwal 77]

W. N. Martin, J. K. Aggarwal

Dynamic Scene Analysis: The Study of Moving Images

Techn. Report 184, Inf. Syst. Research Lab., The Univ. of
Texas at Austin, Austin, Texas, Jan. 1977

[McKee Aggarwal 76]

J. W. McKee, J. K. Aggarwal

Computer Recognition of Partial Views of Three
Dimensional Curved Objects

IJCPR-76, 499-503

[Michie 68]

D. Michie

"Memo" Functions and Machine Learning

Nature 218 (1968) 19-22

[Minsky 65]

M. Minsky

Matter, Mind and Models

Proceedings of IFIP Congress 65 (W. A. Kalenich ed.)

Spartan Books, Washington D.C., Macmillan, London, 1965

[Milgram Rosenfeld 71]

D. L. Milgram, A. Rosenfeld

Array Automata and Array Grammars

Proceedings of IFIP Congress 1971, (C. V. Freiman, ed.),

pp. 69-74, North-Holland, Amsterdam-London 1972

[Mohr Haton 76]

R. Mohr, J.-P. Haton

A Parsing Algorithm for Imperfect Patterns and its Applications to Speech and Image Recognition

Vortrag IJCPR-76

Bericht 76-R-018, Univ. de Nancy I 1976

[Mori et al. 73]

K.-I. Mori, M. Kidode, H. Asada

An Iterative Prediction and Correction Method for Automatic Stereocomparison

Computer Graphics and Image Processing 2 (1973) 393-401

[Mulligan 77]

G. D. Mulligan

Algorithms for Finding Cliques of a Graph

Report TR 41, Dept. of Comp. Sc., Univ. Toronto, 1977

[Nagel 76a]

H.-H. Nagel

Experiences with Yakimovsky's Algorithm for Boundary and Object Detection in Real World Images

Bericht IfI-HH-B-23/76, Fachbereich Informatik, Univ. Hamburg 1976

IJCPR-76, 753-758

[Nagel 76b]

H.-H. Nagel

Formation of an Object Concept by Analysis of Systematic Time Variations in the Optically Perceptible Environment

Bericht IfI-HH-B-27/76, Fachbereich Informatik, Univ. Hamburg, Juli 1976

Computer Graphics and Image Processing 7 (1978) 149-194

[Nagel 77]

H.-H. Nagel

Analyzing Sequences of TV-Frames: System Design Considerations

Bericht IfI-HH-B-33/77, Fachbereich Informatik, Univ. Hamburg, März 1977, und IJCAI-77, 626

[Nagin et al. 77]

P. A. Nagin, A. R. Hanson, E. M. Riseman

Region Extraction and Description through Planning

COINS Techn. Report 77-8, Univ. of Massachusetts at Amherst, Amherst/Mass., May 1977

[Neumann 78a]

B. Neumann

Preiswertes Digital-Hardcopy-Gerät

Nachrichtentechnische Zeitschrift ntz 31 (1978) 110-112

[Neumann 78b]

B. Neumann

Identifikation von gestörten Objektansichten unter

Verwendung geradliniger Konturapproximationen

Bericht IfI-HH-B-42/78, Fachbereich Informatik, Univ.

Hamburg 1978

[Nevatia 74]

R. Nevatia

Structured Description of Complex Curved Objects for
Recognition and Visual Memory

Memo AIM 250, STAN-CS-74-464, Stanford Univ.,

Dissertation Comp. Sc. Dept., Okt. 1974

[Nevatia 76a]

R. Nevatia

Depth Measurement by Motion Stereo

Computer Graphics and Image Processing 5 (1976) 203-214

[Nevatia 76b]

R. Nevatia

Locating Object Boundaries in Textured Environments

IEEE Trans. Comp. C-25 (1976) 1170-1175

[Newell Ernst 65]

A. Newell, G. Ernst

The Search for Generality

Proceedings of IFIP Congress 65 (W. A. Kalenich ed.)

Spartan Books, Washington D.C., Macmillan, London, 1965

[Newsweek 77]

T. Stoppard

zitiert nach Newsweek, Vol. XC, Nr. 7, 15. August 1977,

p. 36

[Ohlander 75]

R. B. Ohlander

Analysis of Natural Scenes

Dissertation, Comp. Sc. Dept., Carnegie-Mellon Univ.,

Pittsburgh/Pa., April 1975

[O'Gorman Clowes 76]

F. O'Gorman, M. B. Clowes
Finding Picture Edges Through Collinearity of Feature
Points
IEEE Trans. Comp. C-25 (1976) 449-456

[Onoe et al. 73]

M. Onoe, N. Hamano, K. Ohba
Computer Analysis of Traffic Flow Observed by Subtractive
Television
Computer Graphics and Image Processing 2 (1973) 377-392

[Onoe Ohba 76]

M. Onoe, K. Ohba
Digital Image Analysis of Traffic Flow
IJCPR-76, 803-808

[Parrish Goksel 77]

E. A. Parrish, A. K. Goksel
A Camera Model for Natural Scene Processing
Pattern Recognition 9 (1977) 131-136

[Pavlidis 76]

T. Pavlidis
A Review of Algorithms for Shape Analysis
Techn. Report 218, Dept. of EE & CS, Princeton Univ.,
Princeton/N.J. 1976

[Perkins 76]

W. A. Perkins
Multilevel Vision Recognition System
IJCPR-76, 739-744
Model-Based Vision System for Scenes Containing Multiple
Parts
IJCAI-77, 678-684

[Pfaltz 72]

J. L. Pfaltz
WEB Grammars and Picture Description
Computer Graphics and Image Processing 1 (1972) 193-220

[Pfaltz Rosenfeld 69]

J. L. Pfaltz, A. Rosenfeld
Web Grammars
IJCAI-69, 609-619

[Pingle Thomas 74]

K. K. Pingle, A. J. Thomas
A Fast, Feature Driven Stereo Depth Program
Memo AIM 248, Stanford Univ., May 1974

[Pohl 71]

I. Pohl
Bi-Directional Search
Machine Intelligence 6 (B. Meltzer, D. Michie, eds.),
Edinburgh University Press 1971, pp. 127-140

[Poplestone et al. 75]

R. J. Poplestone, C. M. Brown, A. P. Ambler, G. F.
Crawford
Forming Models of Plane-and-Cylinder Faceted Bodies from
Light Stripes
IJCAI-75, 664-668

[Potter 75a]

J. Potter
Motion as a Cue to Segmentation
IEEE Trans. Sys. Man Cyb. SMC-5 (1975) 390-394

[Potter 75b]

J. Potter
Scene Segmentation by Velocity Measurements Obtained with
a Cross-Shaped Template
IJCAI-75, 803-810

[Price 76]

K. E. Price
Change Detection and Analysis in Multi-Spectral Images
Dissertation, Carnegie Mellon Univ., Pittsburgh/Pa., Dez.
1976

[Price Reddy 77]

K. E. Price, R. Reddy
Change Detection and Analysis in Multispectral Images
IJCAI-77, 619-625

[Prager et al. 77]

J. M. Prager, A. R. Hanson, E. M. Riseman
Extracting and Labelling Boundary Segments in Natural
Scenes
COINS Techn. Report 77-7, Univ. of Massachusetts at
Amherst, Amherst/Mass., 1977

[Ramer 75]

U. Ramer
Extraction of Line Structures from Photographs of Curved
Objects
Computer Graphics and Image Processing 4 (1975) 81-103

[Riseman Arbib 77]

E. M. Riseman, M. A. Arbib
Computational Techniques in the Visual Segmentation of
Static Scenes
Computer Graphics and Image Processing 6 (1977) 221-276

[Roberts 65]

L. G. Roberts
Machine Perception of Three-Dimensional Solids
in Optical and Electro-Optical Information Processing
(Tippet et al., eds.), MIT Press, Cambridge/Mass., 1965,
pp. 159-197

[Rosenfeld 71]

A. Rosenfeld
Isotonic Grammars, Parallel Grammars, and Picture
Grammars
Machine Intelligence 6 (B. Meltzer, D. Michie, eds.)
Edinburgh University Press 1971, pp. 281-294

[Rosenfeld et al. 76]

A. Rosenfeld, R. A. Hummel, S. W. Zucker
Scene Labelling by Relaxation Operations
IEEE Trans. Systems, Man, and Cybernetics SMC-6 (1976)
420-433

[Rosenfeld Kak 76]

A. Rosenfeld, A. C. Kak
Digital Picture Processing
Academic Press, New York, 1976

[Rothstein Weiman 76]

J. Rothstein, C. Weiman
Parallel and Sequential Specification of a Context
Sensitive Language for Straight Lines on Grids
Computer Graphics and Image Processing 5 (1976) 106-124

[Rubin Reddy 77]

S. M. Rubin, R. Reddy
The Locus Model of Search and it's Use in Image
Interpretation
IJCAI-77, 590-595

[Rüenaufner 76]

P. Rüenaufner
Untersuchungen zur Ermittlung von Möglichkeiten und
Grenzen der Einbeziehung des elektronischen Fernsehens in
die Analyse des Strassenverkehrsablaufs
Dissertation, Techn. Hochschule Aachen, Dez. 1976

[Sakai et al. 76]

T. Sakai, T. Kanade, Y. Ohta
Model-Based Interpretation of Outdoor Scenes
IJCPR-76, 581-585

[Shapira 74]

R. Shapira
A Technique for the Reconstruction of a Straight-Edge,
Wire-Frame Object from Two or More Central Projections
Computer Graphics and Image Processing 3 (1974) 318-326

[Shapira Freeman 77]

R. Shapira, H. Freeman
Reconstruction of Curved-Surface Bodies from a Set of
Imperfect Projections
IJCAI-77, 628-634

[Shirai 75]

Y. Shirai
Edge Finding, Segmentation of Edges and Recognition of
Complex Objects
IJCAI-75, 674-681

[Shortliffe 76]

E. H. Shortliffe
Computer-Based Medical Consultation: MYCIN
American Elsevier, New York, 1976

[Smith Phillips 72]

E. A. Smith, D. R. Phillips
Automated Cloud-Tracking Using Precisely Aligned Digital
ATS Pictures
IEEE Trans. Comp. C-21 (1972) 715-729

[Sobel 74]

I. Sobel

On Calibrating Computer Controlled Cameras for Perceiving
3-D Scenes

Artificial Intelligence 5 (1974) 185-198

[Somerville Mundy 76]

C. Somerville, J. L. Mundy

One Pass Contouring of Images through Planar
Approximation

IJCPR-76, 745-748

[Sties 73]

M. Sties

Zur Beschreibung und Klassifizierung von Linienstrukturen
Dissertation, Univ. Karlsruhe, Mai 1973

[Tanimoto Pavlidis 75]

S. L. Tanimoto, T. Pavlidis

A Hierarchical Data Structure for Picture Processing

Computer Graphics and Image Processing 4 (1975) 104-119

[Tanimoto Pavlidis 76]

S. L. Tanimoto, T. Pavlidis

Graph Labelling Algorithms for Picture Analysis

IJCPR-76, 749-752

[Tenenbaum 73]

J. M. Tenenbaum

On Locating Objects by Their Distinguishing Features in
Multisensory Images

Computer Graphics and Image Processing 2 (1973) 308-320

[Tenenbaum Barrow 76]

J. M. Tenenbaum, H. G. Barrow

IGS: A Paradigm for Integrating Image Segmentation and
Interpretation

in (C. H. Chen, ed.) Pattern Recognition and Artificial
Intelligence, pp. 472-507, Academic Press, New York, 1976
und

Experiments in Interpretation Guided Segmentation

Artificial Intelligence 8 (1977) 241-274

[Tsotsos 76]

J. K. Tsotsos
A Prototype Motion Understanding System
Master Thesis, Dept. Comp. Sc., Techn. Report TR 93,
Univ. Toronto, Juni 1976

[Turner 74]

K. J. Turner
Computer Perception of Curved Objects Using a Television
Camera
Dissertation, Dept. of Machine Intelligence, Univ.
Edinburgh, 1974

[Uhr 73a]

L. Uhr
The Description of Scenes over Time and Space
Nat. Comp. Conf. 1973, pp 509-517, AFIPS Conf.
Proceedings Vol. 42

[Uhr 73b]

L. Uhr
Describing, Using "Recognition Cones"
Techn. Report 176, Comp. Sc. Dept., Univ. of Wisconsin,
Madison/Wisc. 1973

[Uhr 75a]

L. Uhr
Toward Integrated Cognitive Systems, which Must make
Fuzzy Decisions about Fuzzy Problems
in L. A. Zadeh, K. S. Fu, K. Tanaka, M. Shimura (eds.)
Fuzzy Sets and Their Applications to Cognitive and
Decision Problems, pp. 353-393, Academic Press, New York,
1975

[Uhr 75b]

L. Uhr
A Wholistic Integrated Cognitive System (SEER-T1) that
Interacts with it's Environment over Time
Techn. Report 236, Comp. Sc. Dept., Univ. of Wisconsin,
Madison/Wisc., 1975

[Uhr 75c]

L. Uhr
Recognition Cones that Perceive and Describe Scenes that
Move and Change over Time
Techn. Report 235, Comp. Sc. Dept., Univ. of Wisconsin,
Madison/Wisc., 1975

IJCPR-76, 287-293

[Uhr Douglass 77]

L. Uhr, R. Douglass
A Parallel-Serial "Recognition Cone" System for
Perception: Some Test Results
Techn. Report 292, Comp. Sc. Dept., Univ. of Wisconsin,
Madison/Wisc., 1977 und IJCAI-77, 597

[Ullman 77]

S. Ullman
The Interpretation of Visual Motion
Dissertation, MIT, Mai 1977

[Underwood Coates 75]

S. A. Underwood, C. L. Coates
Visual Learning from Multiple Views
IEEE Trans. Comp. C-24 (1975) 651-661

[Uno et al. 76]

T. Uno, M. Ejiri, T. Tokunaga
A Method of Real-Time Recognition of Moving Objects and
its Application
Pattern Recognition 8 (1976) 201-208

[Wahlster 77]

W. Wahlster
Die Repräsentation von vagem Wissen in
natürlichsprachlichen Systemen der Künstlichen
Intelligenz
Diplomarbeit, Fachbereich Informatik, Univ. Hamburg Juli
1977

[Waltz 72]

D. Waltz
Generating Semantic Descriptions from Drawings of Scenes
with Shadows
Report TR 271, MIT, Cambridge/Mass., Nov. 1972 und in
The Psychology of Computer Vision (P. H. Winston, ed.)
McGraw-Hill, New York, 1975

[Waterman 77]

U. A. Waterman
An Introduction to Production Systems
AISB European Newsletter 25 (1977) 7-10

[Williams Lowrance 77]

T. P. Williams, J. D. Lowrance
Model-Building in the VISIONS High Level System
COINS Techn. Report 77-1, Univ. of Massachusetts at
Amherst, Januar 1977

[Winston 75]

P. H. Winston
Learning Structural Descriptions from Examples
in P. H. Winston (ed.) The Psychologie of Computer
Vision, McGraw Hill, New York, 1975

[Woodham 77]

R. J. Woodham
A Cooperative Algorithm for Determining Surface
Orientation from a Single View
IJCAI-77, 635-641

[Wolferts 73]

K. Wolferts
Ein interaktives Verfahren zur teilautomatischen
Auswertung von Luftbildern für Verkehrsanalysen
NTG/GI Fachtagung Cognitive Verfahren und Systeme,
Hamburg 11.-13. April 1973, (T. Einsele, W. Giloi, H.-H.
Nagel, eds.) Springer Berlin-Heidelberg-New York 1973,
pp. 307-331

[Wolferts 74]

K. Wolferts
Special Problems in Interactive Image-Processing for
Traffic Analysis
IJCPR-74, 1-2

[Yakimovsky 73]

Y. Yakimovsky
Scene Analysis Using a Semantic Base for Region Growing
Report AIM-209, Stanford Univ. 1973

[Yakimovsky 75]

Y. Yakimovsky
Boundary and Object Detection in Real World Images
IJCAI-75, 695-702
J. ACM 23 (1976) 599-618

[Yakimovsky Cunningham 76a]

Y. Yakimovsky, R. Cunningham

On the Problem of Embedding Picture Elements in Regions
Techn. Memo 33-774, Jet Propulsion Lab., Cal. Inst. of
Technologie, Pasadena/Cal., 1976

[Yakimovsky Cunningham 76b]

Y. Yakimovsky, R. Cunningham

DABI - A Data Base for Image Analysis with
Nondeterministic Inference Capability
in (C. H. Chen, ed.) Pattern Recognition and Artificial
Intelligence, pp. 554-592, Academic Press, New York, 1976

[Yegi et al. 75]

M. G. Yegi, N. V. Zavalishin, R. L. Sheinin

Analysis of Three-Dimensional Scene Images on the
Knowledge of their Texture Features
IJCAI-75, 705-709

[Yoda et al. 75]

H. Yoda, J. Motoike, M. Ejiri

Direction Coding Method and its Application to Scene
Analysis
IJCAI-75, 620-627

[Zadeh 65]

L. A. Zadeh

Fuzzy Sets

Information and Control 8 (1965) 338-353

[Zadeh 68]

L. A. Zadeh

Fuzzy Algorithms

Information and Control 12 (1968) 94-102

[Zadeh 72]

L. A. Zadeh

Fuzzy Languages and their Relation to Human and Machine
Intelligence

Man and Computer, Proc. Int. Conf. Bordeaux 1970, pp
130-165, Karger, Basel 1972

[Zucker 76a]

S. W. Zucker

Region Growing: Childhood and Adolescence

Computer Graphics and Image Processing 5 (1976) 382-399

[Zucker 76b]

S. W. Zucker
Relaxation Labelling and the Reduction of Local
Ambiguities
Techn. Report 451, Comp. Sc. Center, Univ. of Maryland,
1976
und IJCPR-76, 852-861 und
Relaxation Labelling, Local Ambiguity, and Low Level
Vision
in (C. H. Chen, ed.) Pattern Recognition and Artificial
Intelligence, pp. 593, Academic Press, New York, 1976

[Zucker et al. 75]

S. W. Zucker, A. Rosenfeld, L. S. Davis
General Purpose Models: Expectations About the Unexpected
IJCAI-75, 716-721

Und zum Schluss noch ein Appell . . .

you send me your poems
I'll send you mine.

Things tend to awaken
even through random communication.

Let us suddenly
proclaim spring. And jeer

at the others,
all the others.

I will send a picture too
if you will send me one of you.

Robert Creeley
"The Conspiracy"

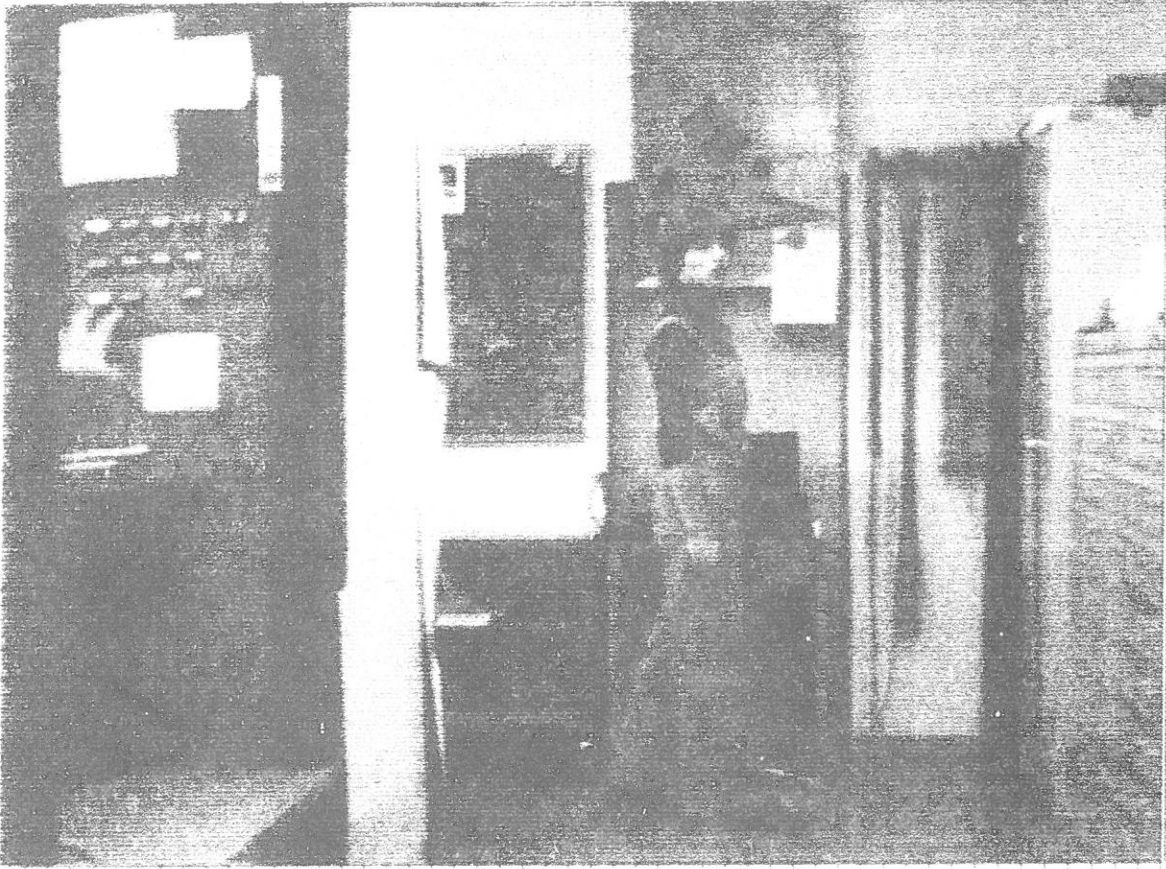


FIG. 1. LABORATORY

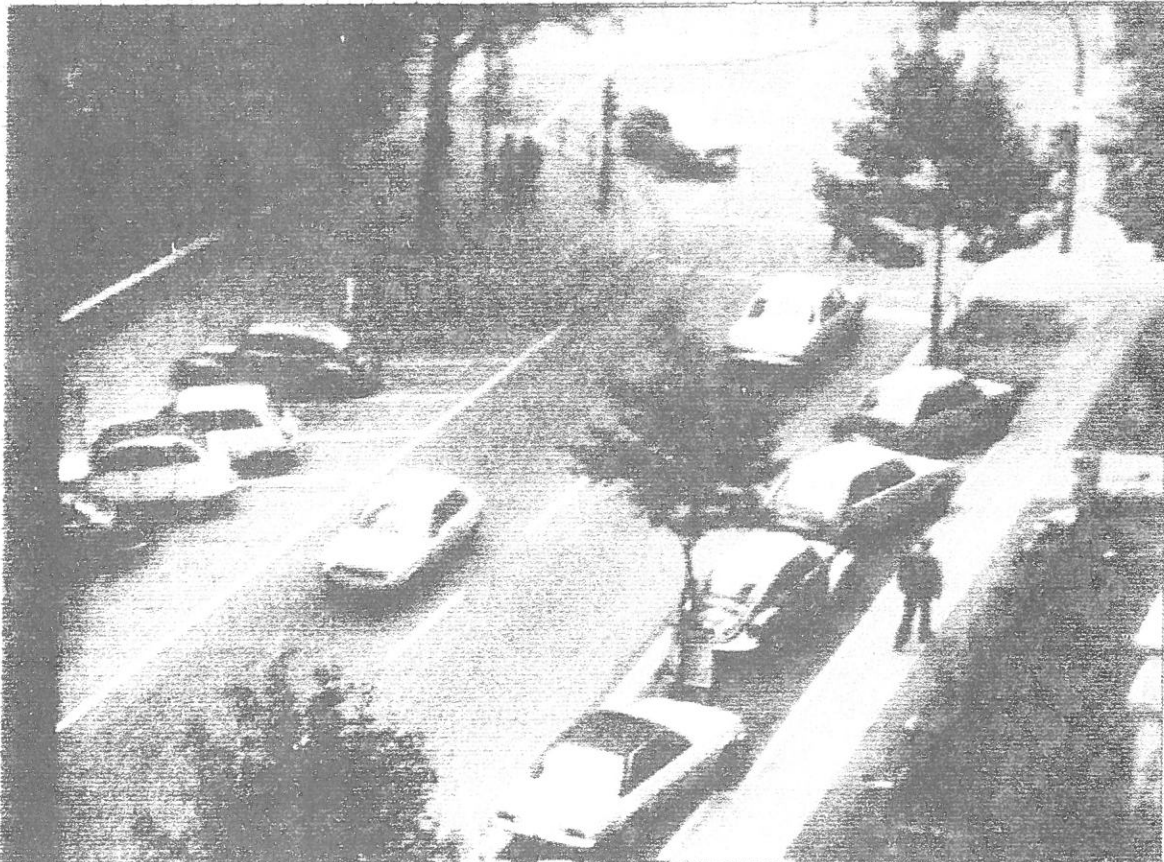


FIG. 2. STREET SCENE



REF. 42 STRASSE-BEIM ORIGINALFLUSS



REF. 42 BEI DISTRIBUTION AUF WITTLEREN BRUNNEN
DAM. DER BRUNNENPUNKT WURDE REDUKTIERT



Abb. 4.3: Bei Topfneigung (Dunkel+Stark)

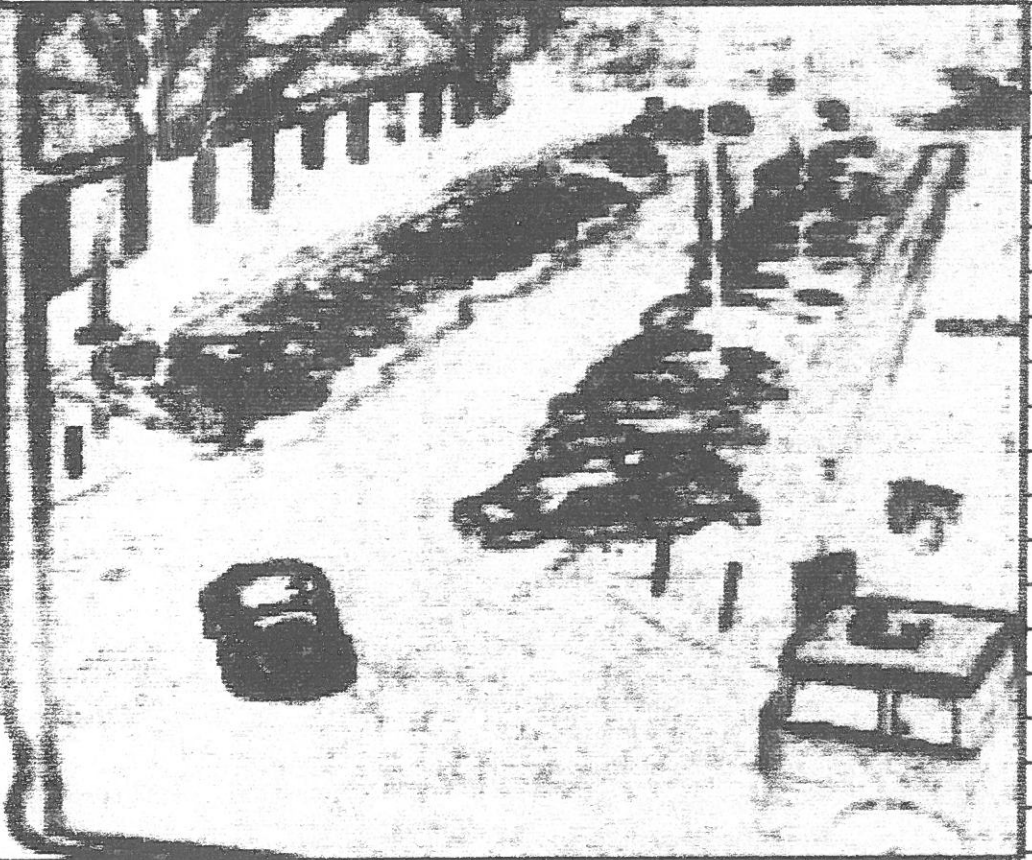


Abb. 4.4: Differenzielle Grauwert-Anpassung (Schwarz-25,5)

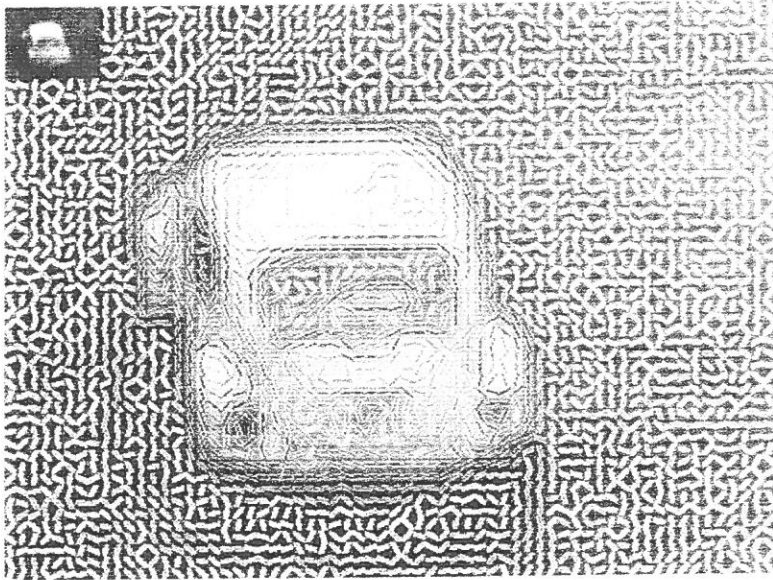
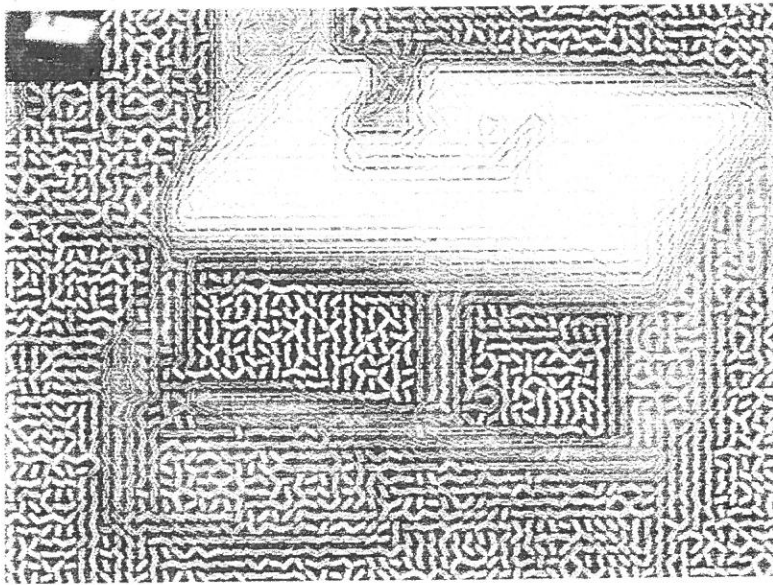


ABB. 4.6 Vektoren: RICHUNG GRÄDWERT NEIGUNG



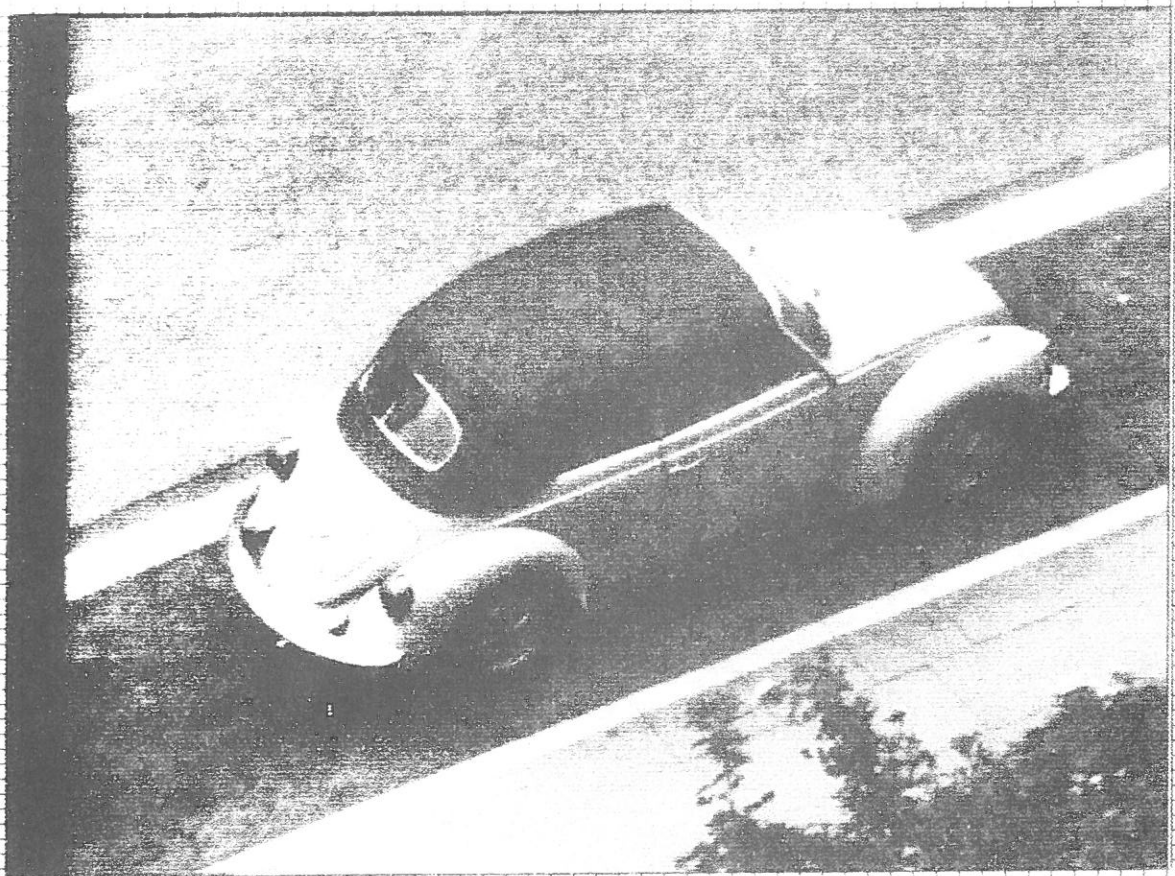
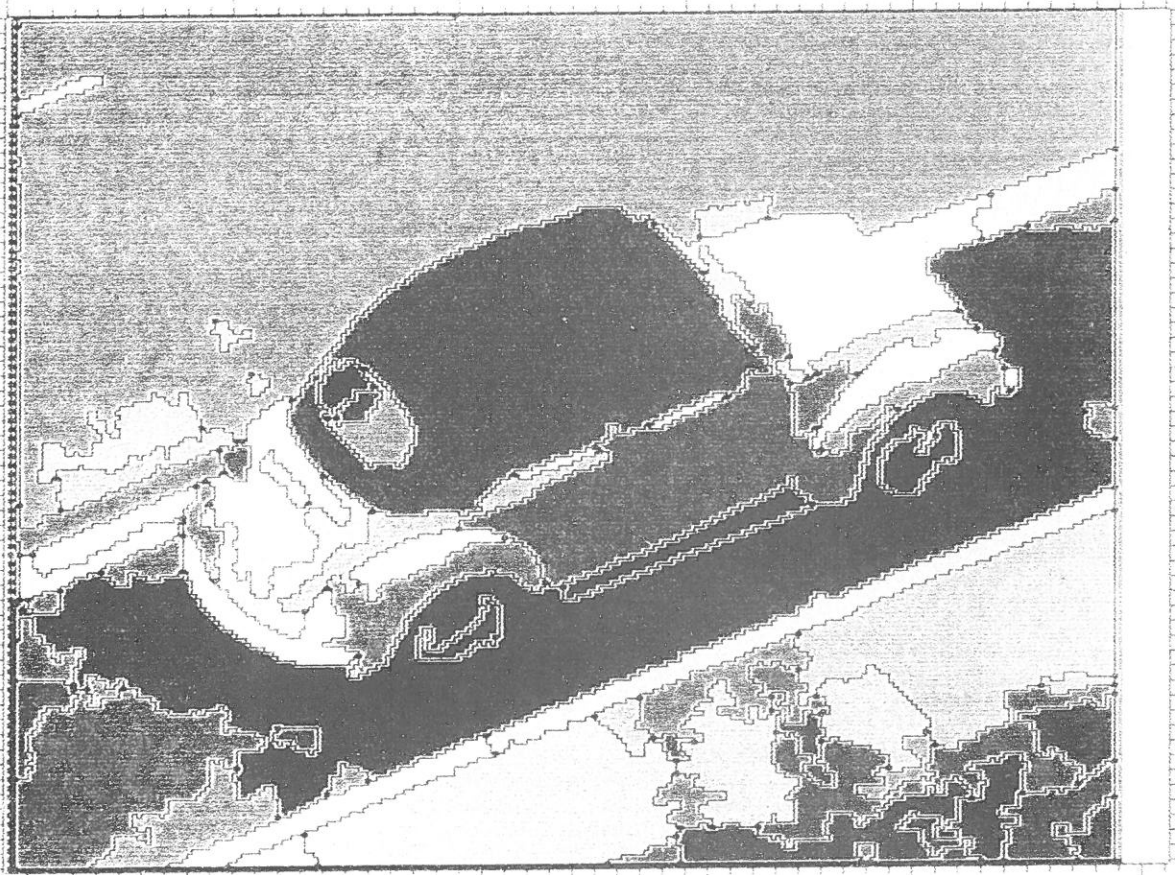
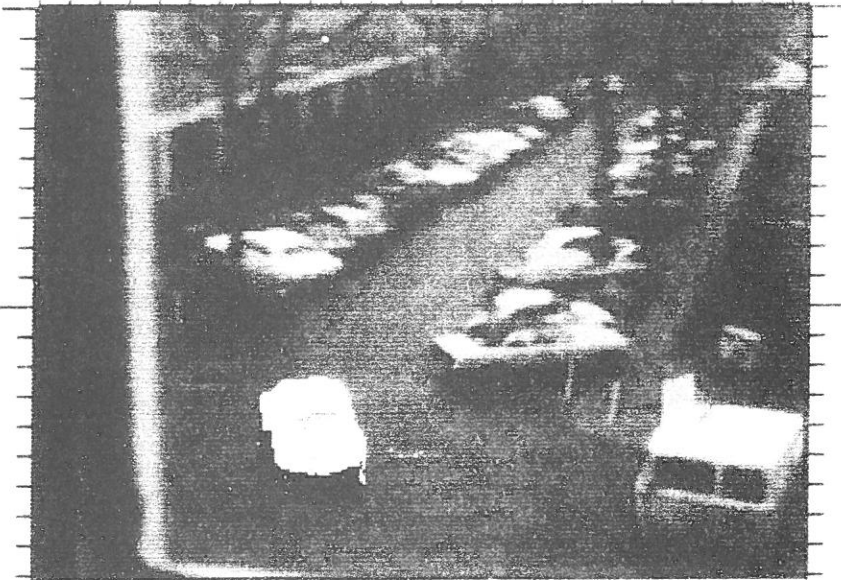


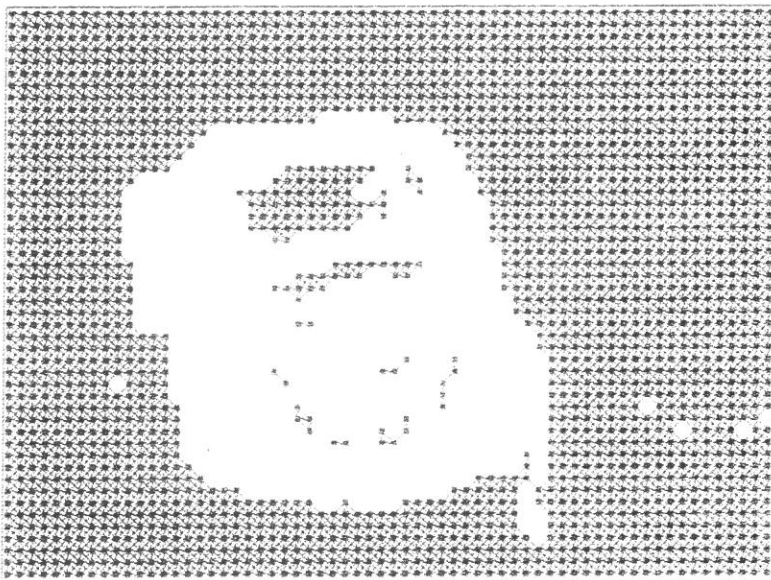
Abb. 5.1 Segmentierung durch "ESDF" (EPDA 77)





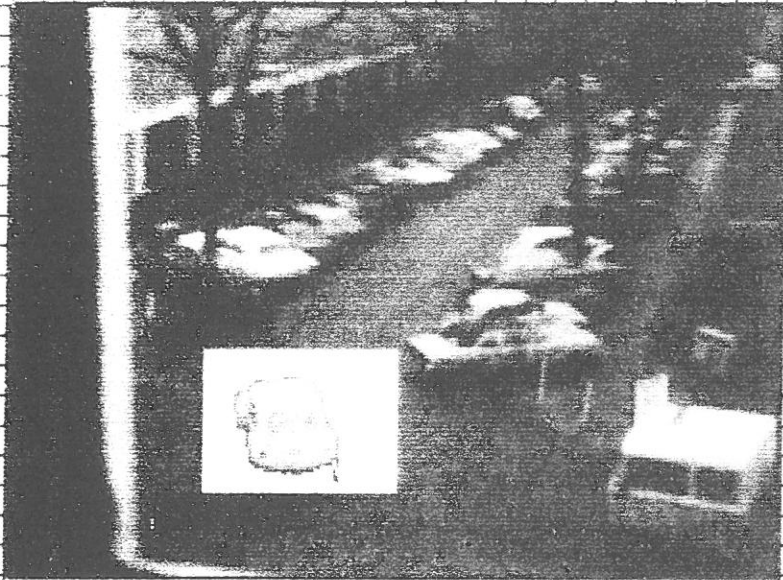
A19.EBN 15-JAN-1978 0:47:14
 Z 115-162 S 65-128 P52.PAR

ABB. 5.2A IM AUSSCHNITT GRAUNERT VON VEKTOREN GERINGER NEIGUNG.
 VEKTOREN STÄRKER NEIGUNG WEISS



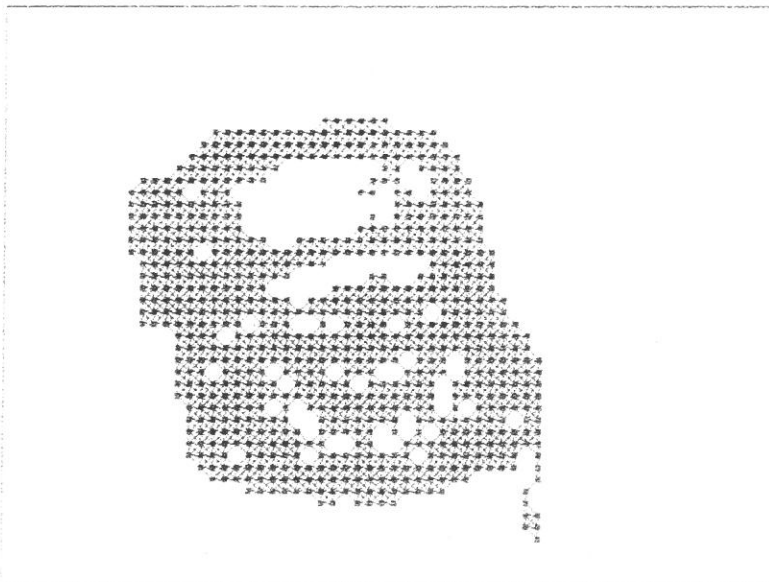
A19.EBN 15-JAN-1978 0:47:14 Z 115-162 S 65-128 P52.PAR

ABB. 5.2B GRAPH DER VEKTOREN GERINGER NEIGUNG IM AUSSCHNITT



A19.EBN 15-JAN-1978 1:1:21
 Z 115-162 S 65-128 P53.PAR

Abb. 5.34 WERTOPFEN GROSSEER NEIGUNG GRAU SONST WEISS



A19.EBN 15-JAN-1978 1:1:27 Z 115-162 S 65-128 P53.PAR

Abb. 5.35 GRAPH DER WERTOPFEN GROSSEER NEIGUNG

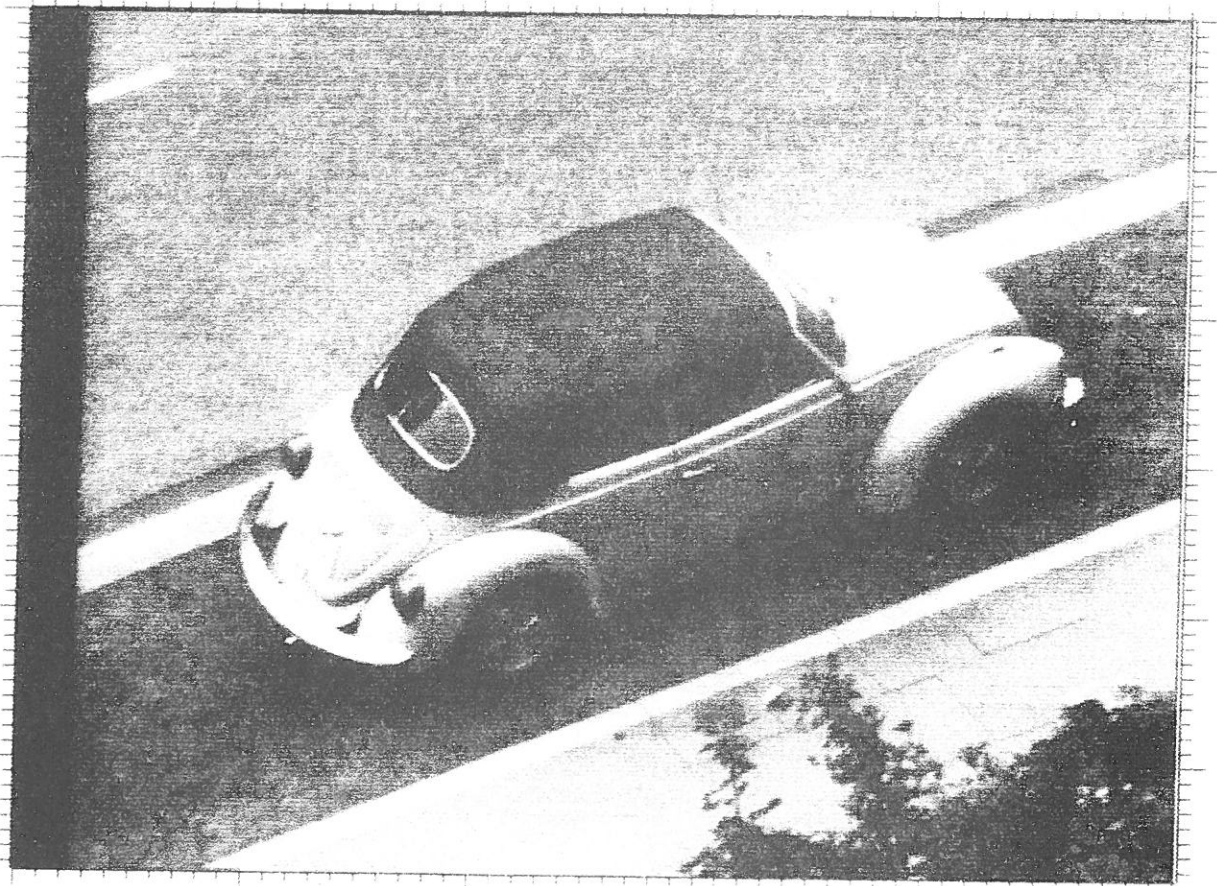
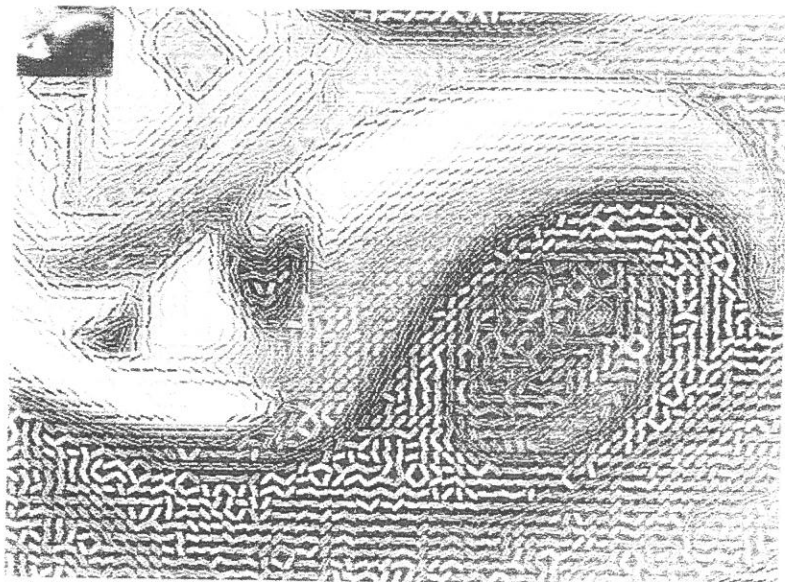
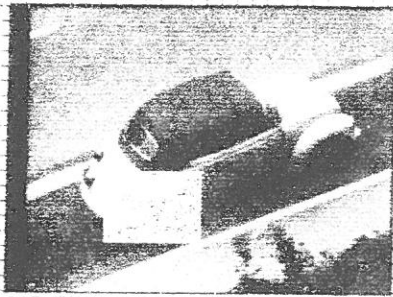


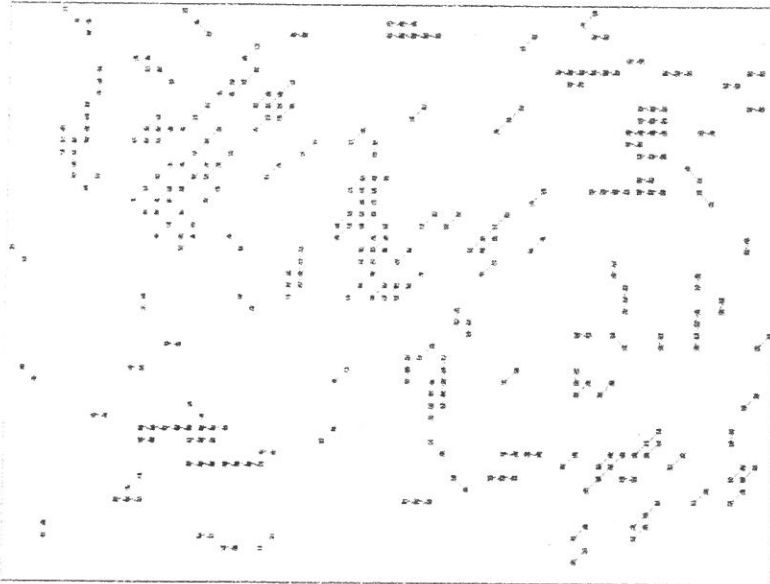
ABB. 5.4 AUTOSZENE UND VEKTORDARSTELLUNG AUSSECHNITT



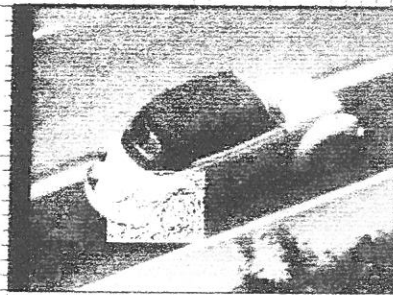


CAR. EBN 15-JAN-1978 1:33:26
 Z 110-157 S 65-128 PSS.PAR

ABB. 5.6 TELEGRAPHEN BESTEHEN AUS STRENG KOLLINEAREN VEKTOREN



CAR. EBN 15-JAN-1978 1:33:33 Z 110-157 S 65-128 PSS.PAR

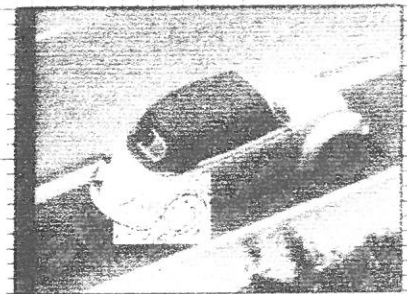


CAR. EBN 15-JAN-1978 1:43:58
 Z 110-157 S 65-128 PSS.PAR

ABB. 5.6 VEKTOREN PARALLEL, JEDOCH AUSLENKUNG VON 22.5 GRAD
 SPLÄUBT ZWISCHEN VENTORRICHTUNG UND RASTERRICHTUNG



CAR. EBN 15-JAN-1978 1:44:19 Z 110-157 S 65-128 PSS.PAR

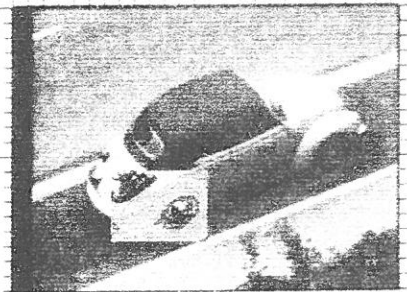


OPR.EBN 15-JAN-1978 1:54:33
Z 110-157 S 65-128 P57.PAR

ABB. 5.7 VEKTOREN PARALLEL; AUSLENKUNG MAX. 45 GRAD;
IM AUSSCHNITT TEILGRAPHEN MIT EINHEITLICHEN GRAUWERT



OPR.EBN 15-JAN-1978 1:54:46 Z 110-157 S 65-128 P57.PAR



OPR.EBN 15-JAN-1978 2:15:20
Z 110-157 S 65-128 P59.PAR

ABB. 5.8 DIFFERENZ DER VEKTORRICHTUNGEN MAX. 45 BRAD

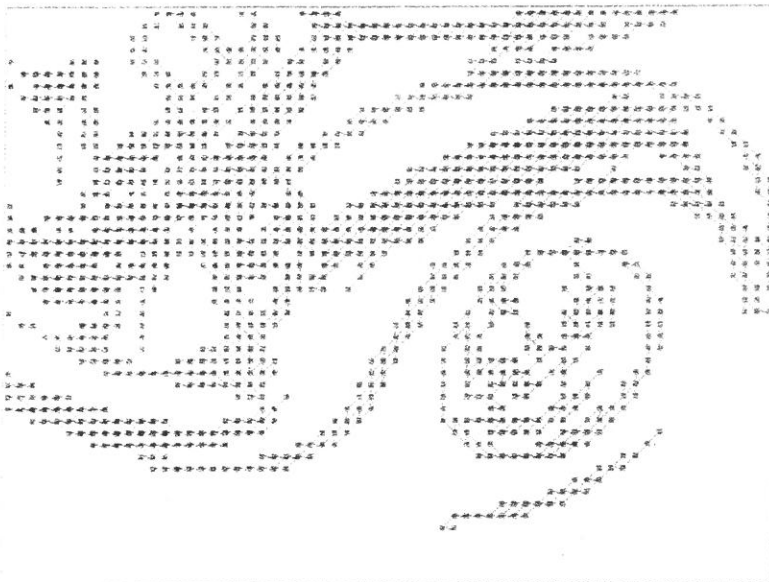


OPR.EBN 15-JAN-1978 2:16:28 Z 110-157 S 65-128 P59.PAR



ABP. 874 15-JAN-1978 2:16:12
 JL 110-167 S 65-128 P89.PAP

ABB. 8.9. NÜBERSTELLUNG DER Vektoren GEFOHRT



ABP. 874 15-JAN-1978 2:18:00 C 110-167 S 65-128 P89.PAP

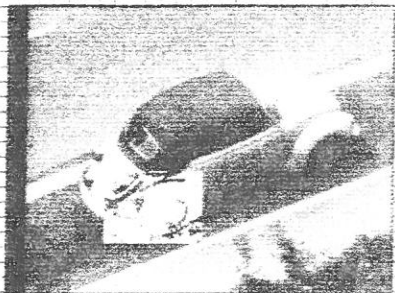


ABP. 874 15-JAN-1978 2:17:21
 JL 110-167 S 65-128 P910.PAP

ABB. 8.10. RICHTUNGSDIFFERENZEN Vektor-Vektor, Vektoren-PASTER
 (WINKELN 30 GRAD (NATÜRLICHE GRADE))



ABP. 874 15-JAN-1978 2:17:27 C 110-167 S 65-128 P910.PAP

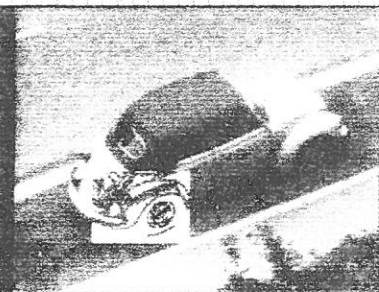


CAR.EBN 15-JAN-1978 2:30:16
IZ 110-157 S 65-123 PS11.PAR

ABB. 5.11 PARALLELE Vektoren SENKRECHT ZUR FASTERICHTUNG
(KUNSTLICHE GRAUERTE)

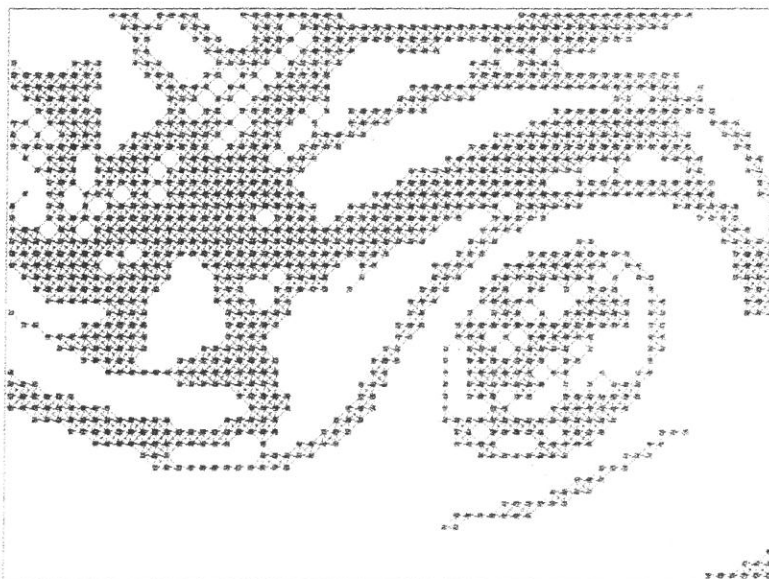


CAR.EBN 15-JAN-1978 3:38:14 IZ 110-157 S 65-123 PS11.PAR

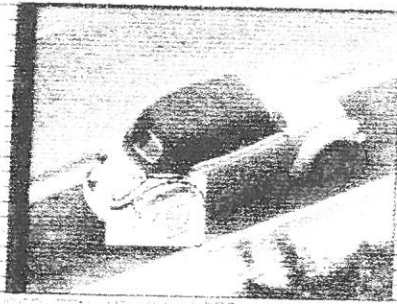


CAR.EBN 15-JAN-1978 3:48:56
IZ 110-157 S 65-123 PS12.PAR

ABB. 5.12 Vektoren GROSSENE NEIGUNG (KUNSTL. GRAUTON)

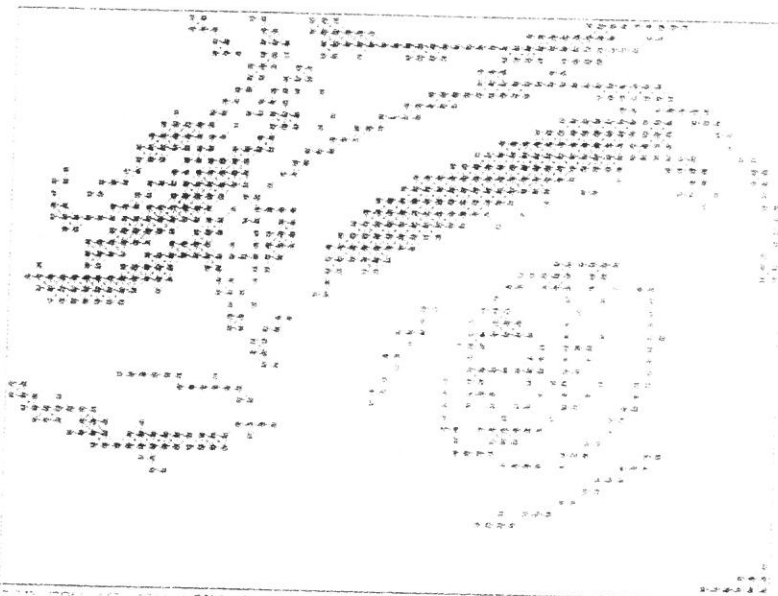


CAR.EBN 15-JAN-1978 3:45:03 IZ 110-157 S 65-123 PS12.PAR

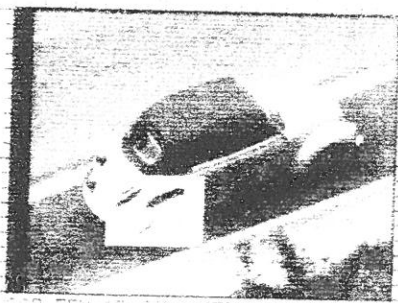


APP. ESN 15-JAN-1975 0411 75
SI 110-187 S 45-126 0511 75

455. S.18 SPANBERT-- FICHTLBERG-STRANDBERG 0411 75

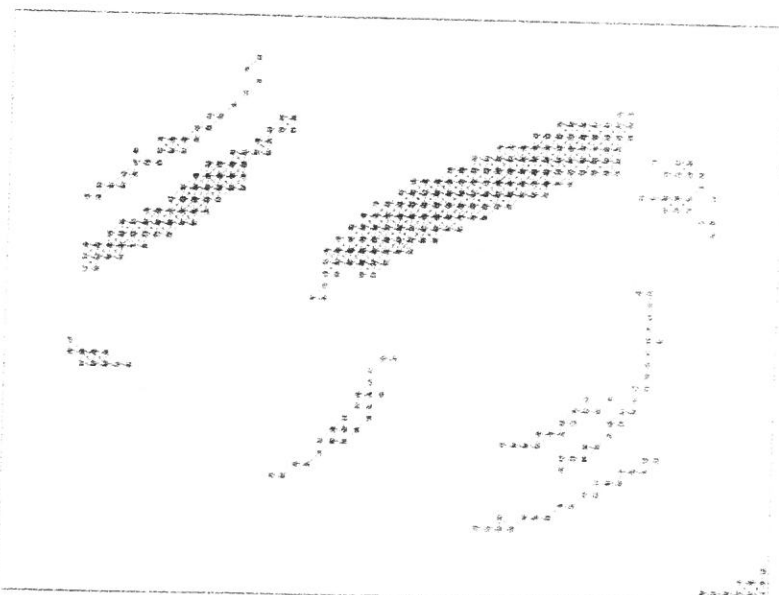


APP. ESN 15-JAN-1975 0411 75 SI 110-187 S 45-126 0511 75

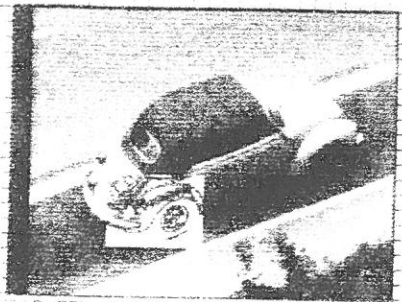


APP. ESN 15-JAN-1975 0411 75
SI 110-187 S 45-126 0511 75

456. S.18 SPANBERT-- FICHTLBERG-STRANDBERG 0411 75

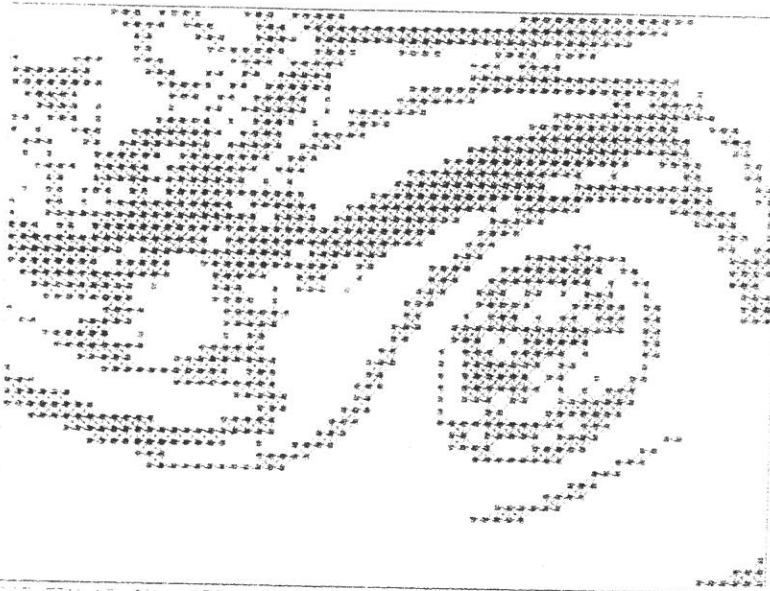


APP. ESN 15-JAN-1975 0411 75 SI 110-187 S 45-126 0511 75

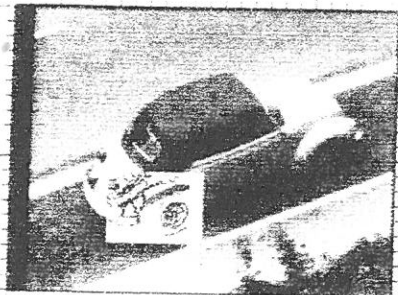


04P.85H 15-JUN-1978 2159149 Z 110-157 S 65-126 P513.PAF

ABB. 8.12 (PAWERT-) RICHTUNGSHISTOGRAMM MIT TA=10

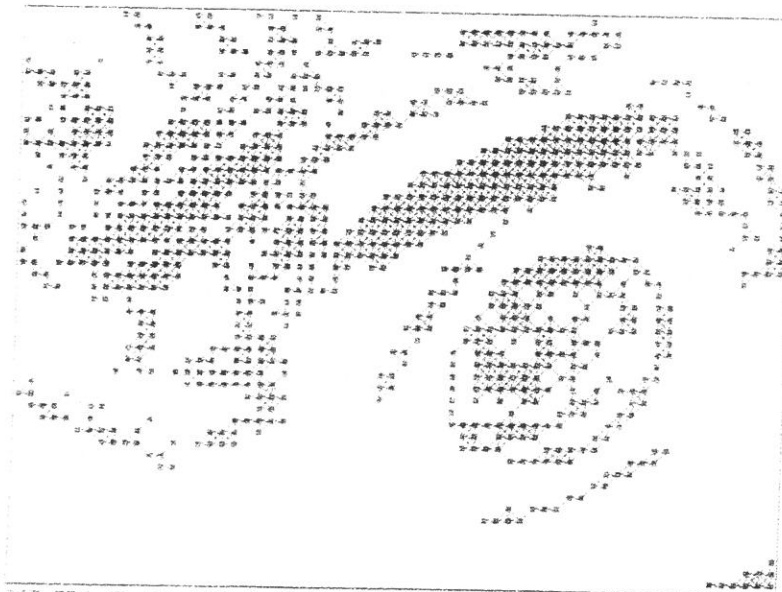


04P.85H 15-JUN-1978 2159149 Z 110-157 S 65-126 P513.PAF

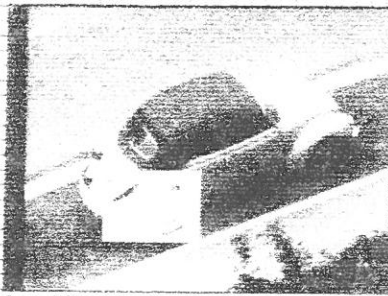


04P.85H 15-JUN-1978 2159149 Z 110-157 S 65-126 P514.PAF

ABB. 8.14 (PAWERT-) RICHTUNGSHISTOGRAMM MIT TA=10

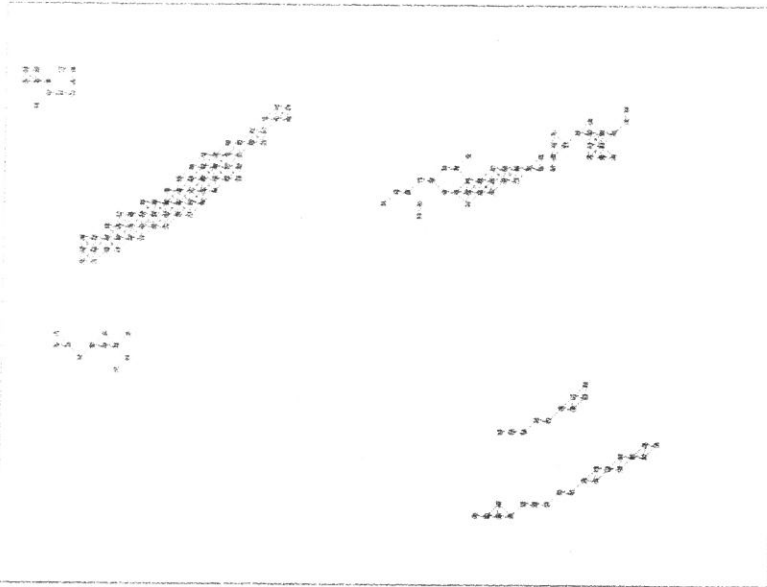


04P.85H 15-JUN-1978 2159149 Z 110-157 S 65-126 P514.PAF



428. EBN 15-JAN-1978 3:42:36
 110-157 S 85-128 PS17.PAF

429. 5.17 BRUNNERT-, RICHTUNGSHISTOGRAMM (NIN-YAN-50)



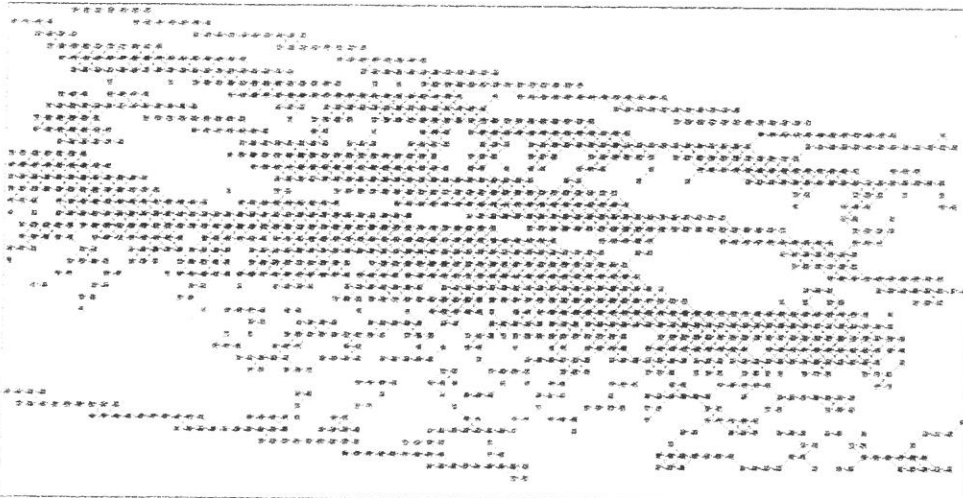
430. EBN 15-JAN-1978 3:42:44 Z 110-157 S 85-128 PS17.PAF

428. 5.18 (ASTENWAGEN MIT SEITENWAND UND SCHATTEN)



431. EBN 15-JAN-1978 3:54:51
 110-147 S 172-251 NWS19.PAF

429. 5.16 SCHATTEN UND SEITENWAND GETRENNTE BEREICHE. OBEN: HELL SEITENWAND. DUNKEL SCHATTEN. UNTEN: GRAU



432. EBN 15-JAN-1978 3:54:57 Z 110-147 S 172-251 NWS19.PAF

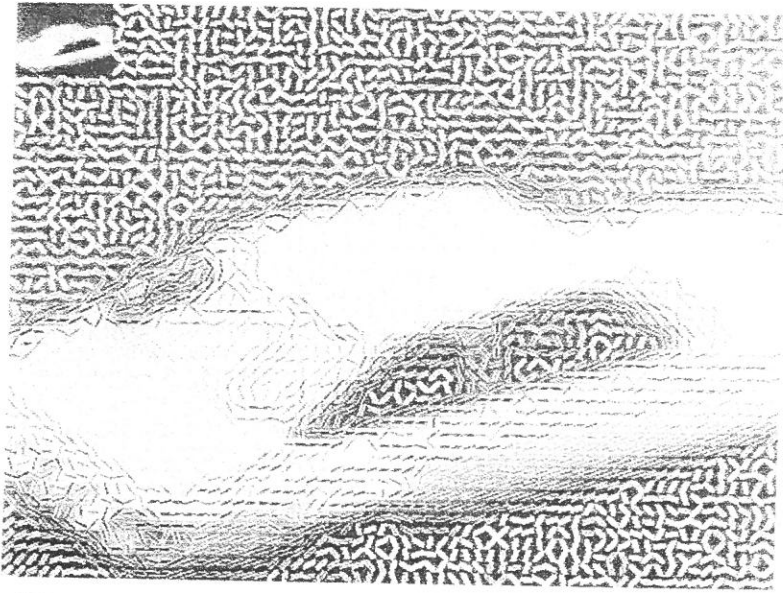


ABB. 6.1 VERTORDARSTELLUNG DES FAHRZEUGES IN POS. 1

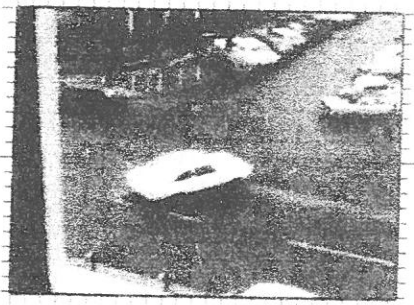


ABB. 6.2 FAHRZEUG IN POSITION 1

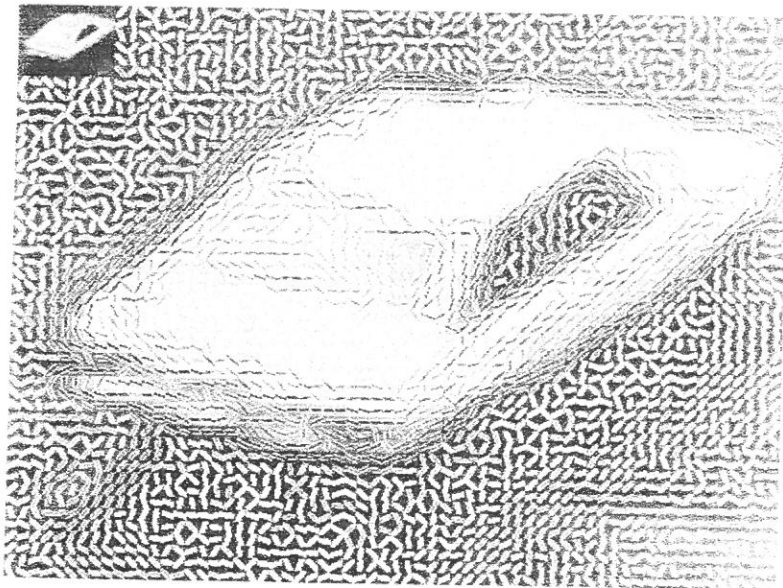


ABB. 6.3 VERTORDARSTELLUNG DES FAHRZEUGES IN POS. 2



ABB. 6.4 FAHRZEUG IN POSITION 2



ABB. 5.6 ERSTES BILD DER FOLGE MIT BEHEBTEM UNTERSCHNITT



ABB. 5.6 ERSTE ERGÄNZUNG DES DACHSCHIEBEPUNKTES SCHWARZ



ABB. 5.7 2.6 BEWANDEN BILDSEQUENZ



ABB. 6.9 FAHRZEUG IN ANFANGSPOSITION EINGEBLENDET

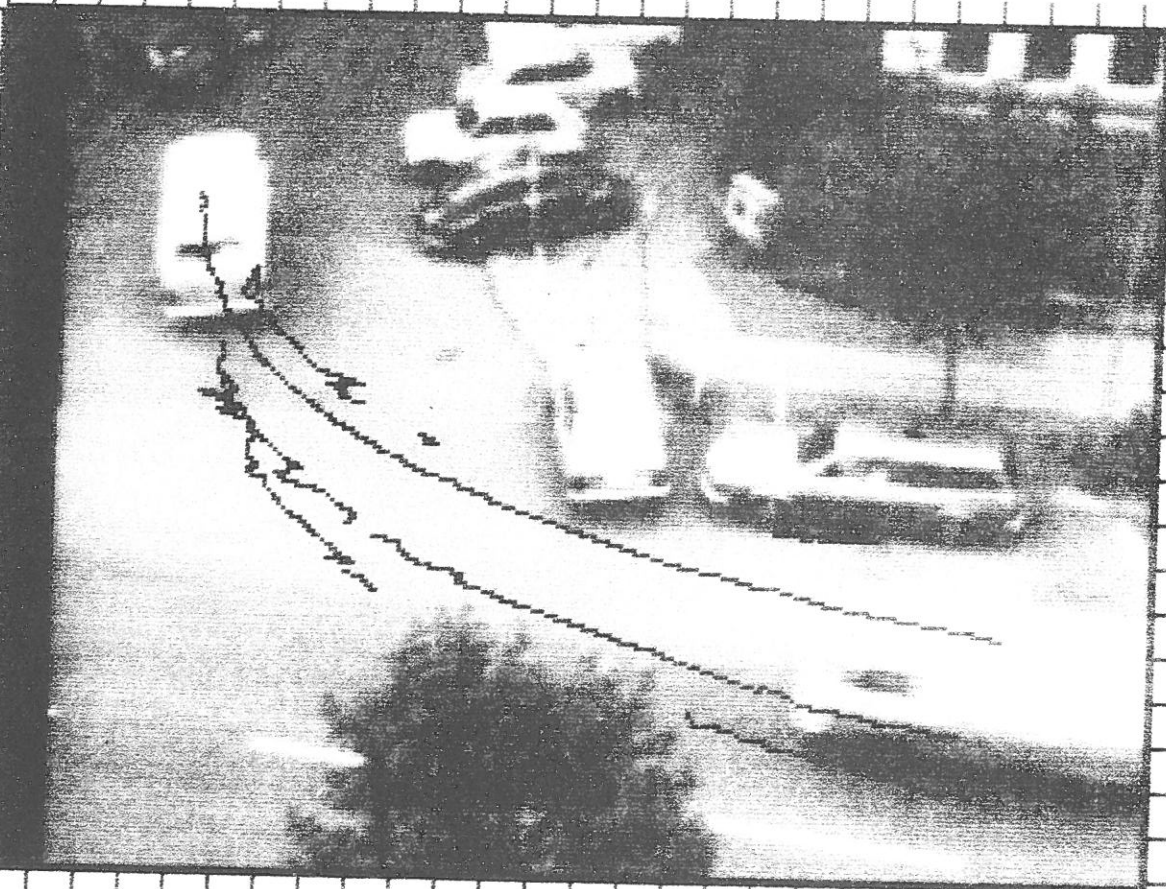


ABB. 6.9 ZWEI VERSCHIEDENE BEREICHETYPEN

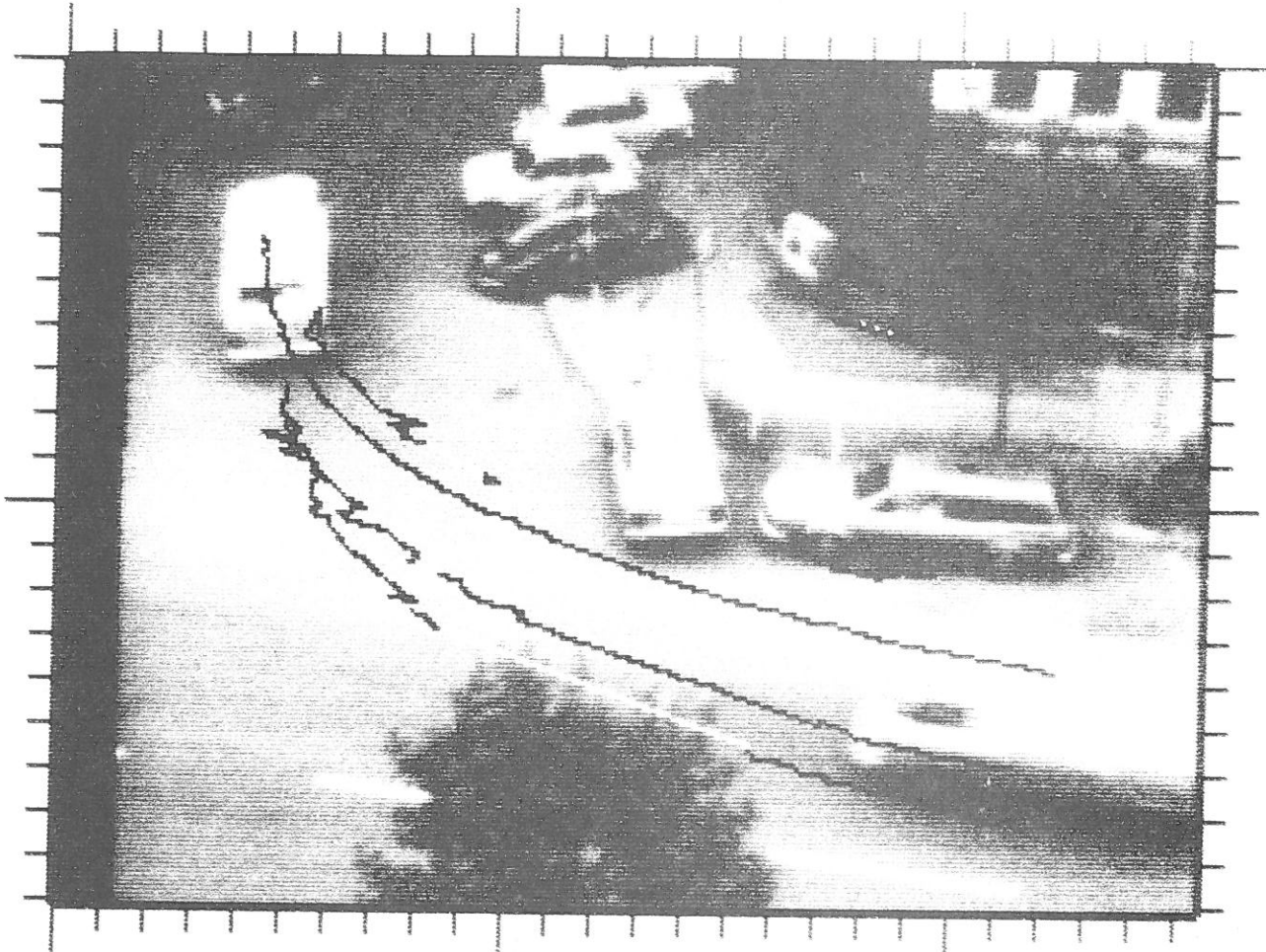


ABB. 5.13 SPUR: SCHWARZ-SCHWERPUNKT, WEISS-ÜBERSPRECHUNG

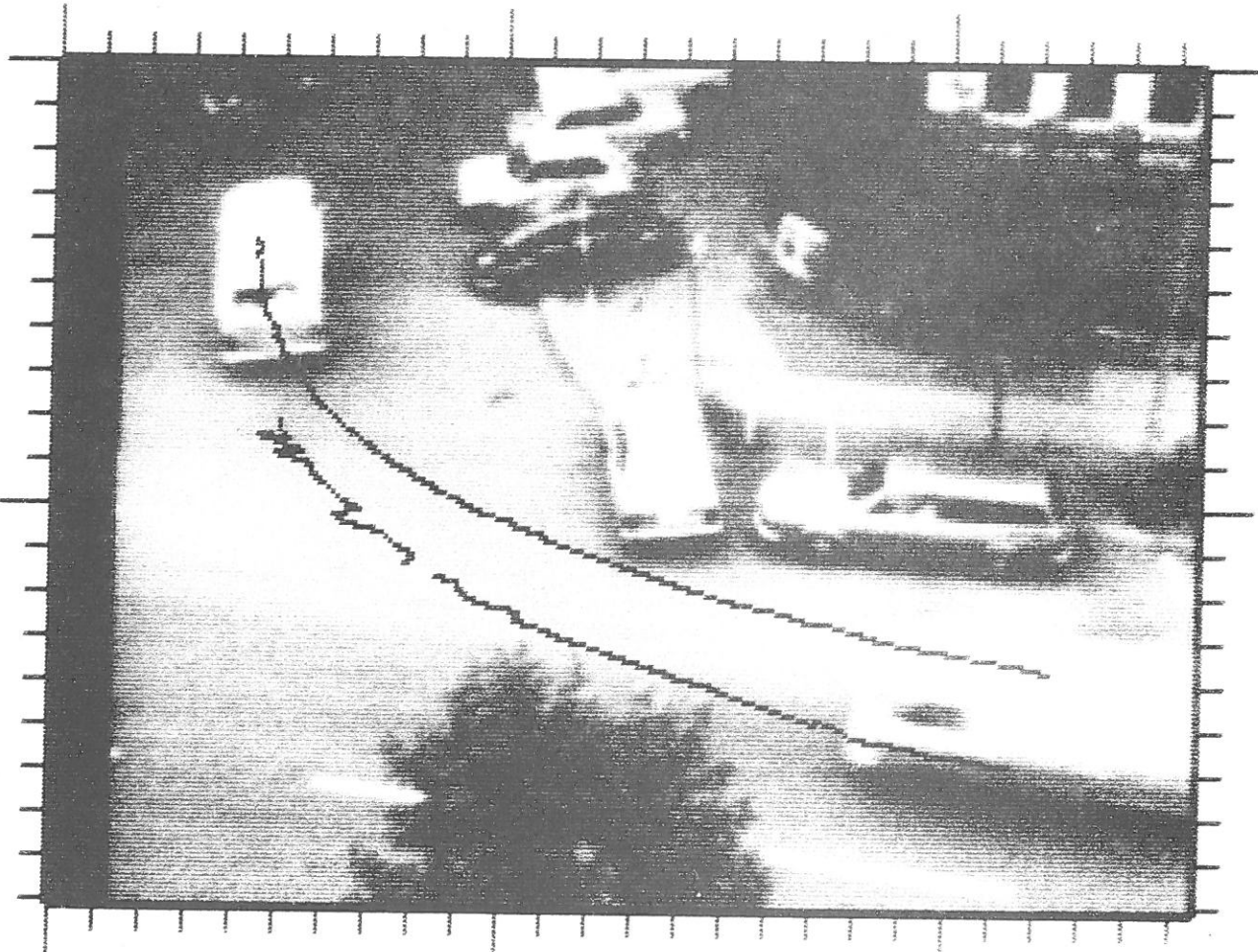
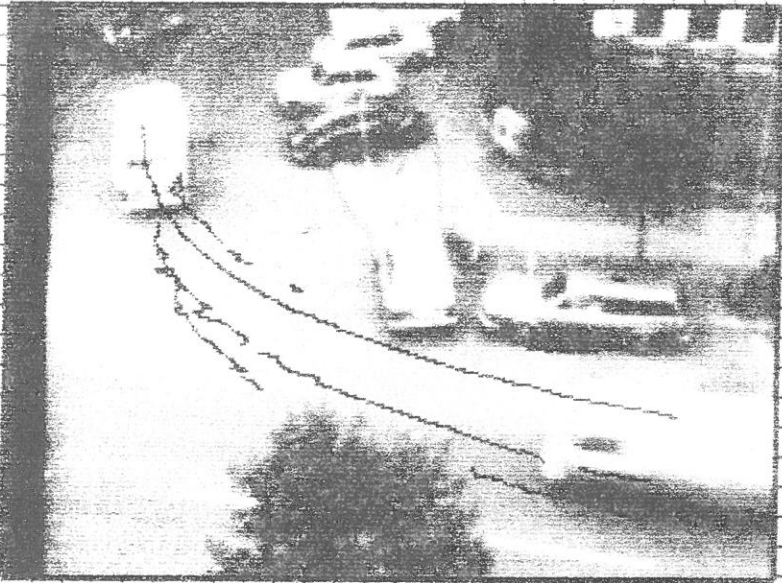
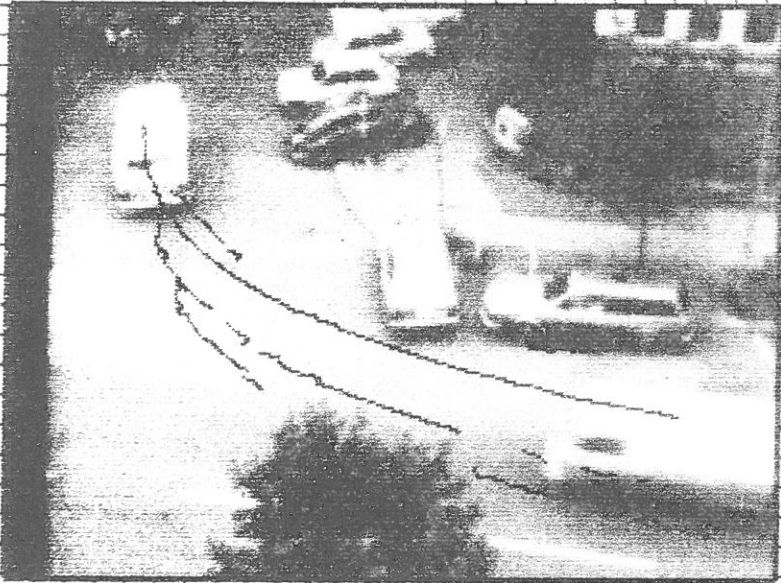


ABB. 5.14 FÄHRZEUG ALS DACH ÜBER SEITENWÄND



REF. 5110 194 - ERSCHÜBUNG 10 SPALTEN-ZEILEN



REF. 5110 194 - ERSCHÜBUNG 8 SPALTEN-ZEILEN



REF. 5110 194 - ERSCHÜBUNG 6 SPALTEN-ZEILEN

Verzeichnis der Autoren

Aggarwal	34-35, 40, 104, 108, 117
Akin	49, 104
Albrecht	62, 104
Ambler	16, 25, 39-40, 73, 104, 106, 121
Arbib	16, 21, 39, 53, 104, 122
Arking	62, 105
Asada	26, 118
Badler	37, 40, 48, 105
Baird	16, 21, 24, 105
Bajcsy	16, 25, 27, 105
Baker	26, 106
Bargel	18, 110
Barrow	16, 20, 35, 39-40, 72-73, 104, 106, 124
Baumgart	24, 27, 106
Bertelsmeier	35, 38, 40-41, 106-107
Bolles	23, 53, 107
Bonde	24, 107
Bowker	69, 107
Braid	27, 107
Brain	30, 112
Brayer	22, 107
Brice	18, 107
Bron	40, 107
Brons	43, 78, 108
Brown	25, 40, 73, 104, 121
Buchanan	54, 109
Budy	30, 62, 108
Burr	26, 108
Burstall	16, 39-40, 73, 104, 106
Chang	25, 108
Chien	26, 29, 68, 108
Chow	35, 40, 108
Clowes	19, 69, 108, 120
Coates	26, 126
Crawford	25, 121
Cunningham	18, 54, 72, 128
Davis	16, 18, 20, 54, 109, 129
Demuth	24, 109
Dierckx	62, 109
Douglass	51, 126
Dreschler	30-32, 62, 108-109
Duda	18, 24-25, 34, 69, 104, 109-110

Dyer	18, 115
Ehrich	18, 110
Ejiri	29, 68, 126, 128
Ellis	62, 110
Endlich	30, 112
Erman	58, 115
Ernst	15, 18, 110, 119
Falk	19, 110
Fennema	18, 107
Fischer	62, 104
Foith	18-19, 110
Freeman	27, 123
Freuder	20, 72-73, 111
Fu	22-23, 54, 107, 111, 114
Gaafar	43, 78, 111
Garvey	25, 111
Gelperin	112
Gimelfarb	25, 111
Goksel	24, 120
Guzman	2, 16, 19, 111
Hall	30, 112
Hamano	29, 120
Hannah	24-25, 112
Hanson	18, 21, 39, 112, 118, 121
Hart	18, 24, 69-70, 109-110, 112
Haton	23, 118
Hewitt	53, 112
Hogg	37, 113
Holdermann	18, 70, 110, 113
Horn	17, 47, 68, 79, 113
Horowitz	18, 113
Huang	23, 54, 114
Huffman	19, 114
Hummel	20, 72, 122
Hückel	44
Hückel	114
Ishi	25, 114
Jacobus	68, 108
Jain	32-33, 54, 114
Jones	29, 108
Kak	16, 18, 122

Kanade	16, 123
Kazmierczak	70, 113
Kelly	16, 21, 34, 105, 114-115
Kerbosch	40, 107
Kidode	26, 118
Kiessling	25
Kiessling	115
Klinger	18, 115
Kraasch	18, 39, 43, 71-72, 115
Kushner	25, 111
Leese	29, 116
Lefaiivre	53, 56, 115
Lesser	58, 115
Levine	26, 115
Lieberman	16, 25, 105
Limb	29, 116
Lowerre	36, 93, 116
Lowrance	21, 127
Mackworth	17, 19, 72, 76, 116
Marr	2, 18, 27, 45, 116
Martelli	18, 70, 117
Martin	35, 117
Mckee	35, 117
Mclain	62, 110
Michie	55, 117
Milgram	22, 117
Militzer	32, 114
Minsky	13, 117
Mohr	23, 118
Mori	26, 118
Motoika	68, 128
Mulligan	40, 118
Mundy	72, 124
Murphy	29, 116
Nagata	25, 114
Nagel	18, 31-34, 47, 54, 71, 114, 118
Nagin	18, 118
Neumann	67, 69, 119
Nevatia	26-27, 69, 119
Newell	15, 119
Nilson	24, 70, 110, 112
Nishihara	27, 116
Nitzan	25, 110
Novak	29, 116
O'gorman	69, 120

O'handley	26, 115
Ohba	29, 120
Ohlander	16, 18, 27, 71, 119
Ohta	16, 123
Onoe	29, 120
Parrish	24, 120
Pavlidis	18-19, 113, 120, 124
Perkins	19, 120
Pfaltz	22, 120
Phillips	30, 123
Pingle	26, 121
Platz	62, 104
Pohl	70, 121
Popplestone	25, 35, 40, 73, 104, 106, 121
Potter	30, 121
Prager	21, 121
Price	18, 27-28, 71, 88, 121
Radig	35, 38, 106
Ramer	70, 122
Raphael	70, 112
Reddy	16, 28, 49, 88, 104, 121, 123
Riseman	16, 18, 21, 39, 53, 104, 112, 118, 121-122
Roberts	19, 122
Rosenfeld	16, 18, 20, 22, 72, 109, 117, 120, 122, 129
Rothstein	43, 122
Rothstein	78
Rubin	16, 123
Rybak	25, 111
Rüenaufener	29
Rüenaufener	61
Rüenaufener	123
Sakai	16, 123
Shapira	27, 123
Sheinin	25, 128
Shirai	19, 35, 123
Shortliffe	54, 109, 123
Smith	30, 123
Sobel	24, 124
Somerville	72, 124
Soroka	27, 105
Sties	22, 124
Stoppard	2, 119

Tanimoto	19, 124
Taylor	29, 116
Tenenbaum	16, 20, 25, 72, 124
Thomas	26, 121
Tokunaga	29, 126
Tsotsos	37, 40, 125
Turner	16, 19, 35, 40, 43, 47, 125
Uhr	51-52, 125-126
Ullman	14, 27, 50, 126
Underwood	26, 126
Uno	29, 126
Wahlster	56, 126
Waltz	16, 19, 47, 72, 126
Wang	25, 108
Waterman	21, 126
Weiman	43, 78, 122
Williams	21, 127
Winston	37, 127
Wolf	30, 112
Wolferts	29, 127
Woodham	17, 79, 127
Yagi	26, 115
Yakimovsky	18, 31-32, 54, 71-72, 127-128
Yegi	25, 128
Yoda	68, 128
Zach	18, 39, 43, 71-72, 115
Zadeh	23, 34, 52, 128
Zavalishin	25, 128
Zucker	18, 20, 72, 122, 128-129

Bisher erschienene Berichte:

- 1 Monien, B.
Beziehungen zwischen Zeitkomplexitätsklassen und Kellerautomaten
mit Zählern
X/71, 36 p
- 2 Kupka, I., Wilsing, N.
A Formal Framework for Dialog Languages
XII/72, 64 p
- 3 Kupka, I., Wilsing, N.
Syntax und Semantik des Dialogsprachenkonzepts HDL
I/73, 54 p
- 4 Kupka, I., Wilsing, N.
An Approach to Formal Semantics of Dialog Languages
/73, 46 p
- 5 Brunnstein, K., Schmidt, J. W.
Teilnehmersysteme
II/73, 167 p
- 6 Monien, B., Weicker, R.
Ueber die Effizienz von Tabulator-Turinemaschinen mit beschränkter
Markenzahl im Vergleich zu anderen Maschinenmodellen
X/73, 34 p
- 7 Woehl, K.
PREVAN - ein Vorcompiler zur Vektoranalysis
X/73, 67 p
- 8 Monien, B.
Komplexitätsklassen von Automatenmodellen und beschränkte Rekursion
II/74, 95 p
- 9 Nagel, H.-H.
On Processes, High Level Languages and Structured Programs for
Computer Networks
V/74, 30 p
- 10 Fleischhauer, G.
Beziehungen zwischen den Zeitkomplexitätsklassen verschiedener
Automatenmodelle
IX/74, 42 p
- 11 Grosse-Lindemann, C.-D., Nagel, H.-H.
Postlude to a Pascal-Compiler Bootstrap on a DECSystem-10
X/74, 28 p
- 12 Hoepner, M.
Ueber den Zusammenhang von Szilardsprachen und Matrixgrammatiken
XII/74, 119 p
- 13 Friesland, G., Sensler, H.-E.
Zur Uebertragung von Compilern durch Selbstcompilation am
Beispiel des Pascal-Compilers
XII/74, 190 p
- 14 Monien, B.
Rekursive und grammatikalische Charakterisierung der
Exponentialzeitsprachen und ein neuer Beweis des
Kleene'schen Normalformtheorems
III/75, 24 p

- 15 Heibey, H.-W.
Ein Modell zur Behandlung mehrdimensionaler Strukturen unter Berücksichtigung der in ihnen definierten Laserrelationen
V/75, 140 p
- 16 Heibey, H.-W., Lutterbeck, B., Schueler, U., Senzler, H.-E.
Nichttechnische Auswirkungen bei der EDV-Anwendung
V/75, 64 p
- 17 Hoepner, M.
Durch Mischoperationen definierte Sprachfamilien
VII/75, 26 p
- 18 Kupka, I., Oberquelle, H., Wilsing, N.
An Experimental Language for Conversational Use
IX/75, 105 p
- 19 Kudlek, M.
Reducing Operators for General Formal Systems
XI/75, 21 p
- 20 Opp, M.
Charakterisierungen erkennbarer Termnennungen in absolut freien universellen Algebren
I/76, 109 p
- 21 Schaeffler, G.
Stochastische OL-Systeme
II/76, 66 p
- 22 Juers, J.
Indexgrammatiken mit eingeschränkter Reselanwendung
II/76, 39 p
- 23 Nagel, H.-H.
Experiences with Yakimovsky's Algorithm for Boundary and Object Detection in Real World Images
III/76, 23 p
- 24 Wittis, T.
Semantische Analyse von Sätzen zur Erfassung eines Sachverhaltes
V/76, 137 p
- 25 Brunnstein, K., Lewandowski, U., Schmidt, J. W.
Rechner fuer den Schuleinsatz - Ein Schulrechnerkatalog
X/75, 260 p
- 26 Albrecht, M.
Automatische Erkennung von Organengrenzen in Leber-Szintigrammen und Erfassung von Parametern zur Beschreibung des Leber-Abbildes im Szintigramm
III/76, 82 p
- 27 Nagel, H.-H.
Formation of an Object Concept by Analysis of Systematic Time Variations in the Optically Perceptible Environment
VII/76, 36 p
- 28 Oberquelle, H.
Grundbegriffe zur Beschreibung von Dialogen und dialogfähigen Systemen
IX/76, 60 p

- 29 Kries, B.
Petrietze und Zustandsgraphen
I/77, 100 p
- 30 Bruesse, B., Gisch, B., Kahl, T., Linde, H.,
Mittelstein, M., Westphal, H.
Concurrent PASCAL - Compiler fuer Kleinrechner
XII/76, 190 p
- 31 Schmidt, J. W.
Some High-Level Language Constructs for Data
of Type Relation : An Investigation Based on PASCAL Extensions
I/77, 18 p
- 32 Jain, R., Militzer, D., Nagel, H.-H.
Separating Non-stationary from Stationary Scene Components
in a Sequence of Real World TV-Images
III/77, 18 p
- 33 Nagel, H.-H.
Analysing Sequences of TV-Frames : System Design Considerations
III/77, 44 p
- 34 Valk, R.
Self-Modifying Nets
VII/77, 36 p
- 35 Hoepner, M., Opp, M.
Renaming and Erasing in Szilard Languages
V/77, 37 p
- 36 Ullrich, G.
Der Entwurf von Steuerstrukturen fuer parallele
Ablaeufe mit Hilfe von Petri-Netzen
XII/76, 268 p
- 37 Schmidt, J. W.
Type Concepts for Database Definition :
An Investigation Based on Extensions to PASCAL
V/77, 22 p
- 38 Wahlster, W.
Zur Repraesentation von unserem Wissen in natuerlich-
sprachlichen Systemen der Kuenstlichen Intelligenz
VII/77, 130 p
- 39 Damm, F.-E.
Entwurf einer Dialogsprache zum computergestuetzten
Ableiten und Beweisen
V/77, 142 p
- 40 Bonk, M.
Rudimentaere Klassen
III/77, 63 p
- 41 Bertelsmeier, R., Radig, B.
Context-Guided Analysis of Scenes with Moving Objects
IV/77, 18 p
- 42 Neumann, B.
Identifikation von gestoerten Objektansichten unter
Verwendung geradliniger Konturapproximation
XII/77, 34 p

- 43 Materna, W.
Ein allgemeines Verfahren zur Konstruktion von repräsentativen Modellen
XII/77, 187 p
- 44 Dreschler, L., Nagel, H.-H.
Using 'Affinity' for Extracting Images of Moving Objects from TV-Frame Sequences
II/78, 28 p
- 45 Froelich, R.
Zwei neue Verfahren zur Aufwandsminderung diskreter Simulation von Rechensystemen
II/78, 157 p
- 46 Fischer, H., Jarke, M., Meyer, D., Schmidt, J. W., Ullmer, W.
A Structured Framework for the Implementations of Relations
II/78, 22 p
- 47 Seehusen, J.
Ein nichtlineares Optimierungsverfahren mit Restriktionen zur Dimensionierung von Rechnerkonfigurationen
IV/78, 164 p
- 48 Jain, R., Nagel, H.-H.
On a Motion Analysis Process for Image Sequences from Real World Scenes
IV/78, 37 p
- 49 Friede, D.
Über deterministische kontextfreie Sprachen und rekursiven Absties
VI/78, 102 p
- 50 Dehnig, W., Essig, H., Maass, S.
Zur Anpassung virtueller Mensch-Rechner-Schnittstellen an Benutzererfordernisse im Dialog - dargestellt am Beispiel von Datenbanksystemen
VII/78, 272 p
- 51 Brodie, M. L.
The Application of Data Types to Databases
XII/78, 20 p
- 52 Meyer, D.
Entwurf und Implementation von PASCAL-Erweiterungen für die Bearbeitung relationaler Datenbanken
VI/78, 104 p
- 53 Hoffmann, H.-J., Kupka, I.
Arbeiten mit interaktiven Systemen, erläutert am Vorgang des Programmierens - Theoretische Konzepte und Gesichtspunkte der Realisierung -
X/78
- 54 Kupka, I., Maluszynski, J.
A Dialog Language with an ALGOL 68 - Like Kernel.
Draft Description
XII/78, 43 p
- 55 Kowalk, W.
Reduktion paralleler Programme
XII/78, 152 p
- 56 Scheffe, P.
Foundations of Reasoning with Uncertain Facts and Vague Concepts
IX/78, 51 p
- 57 Boley, H.
Five Views of FIT Programming
XI/78, 68 p

