

Objektdetektion in Straßenverkehrsszenen durch Auswertung von Farbbildfolgen

Von der Fakultät Elektrotechnik der Universität Stuttgart
zur Erlangung des akademischen Grades
eines Doktor-Ingenieurs (Dr.-Ing.) genehmigte Abhandlung

Vorgelegt von Bernd Heisele aus Aalen

Hauptberichter: Prof. Dr. rer. nat. habil. J. H. Werner
Mitberichter: Prof. Dr.-Ing. P. Göhner
Tag der mündlichen Prüfung: 26. Oktober 1998

Institut für Physikalische Elektronik der Universität Stuttgart

Danksagung

Die vorliegende Arbeit entstand während meiner Tätigkeit in der Abteilung für „Bildverstehende Systeme“ (FT3/AB) der Daimler-Benz Forschung in Ulm.

Bei Herrn Prof. Dr. rer. nat. habil. J. H. Werner und Herrn Prof. Dr.-Ing. P. Göhner bedanke ich mich für das Interesse, das sie der Arbeit entgegengebracht haben und für die Übernahme des Erst- bzw. Zweitgutachtens. Für ihre Hilfsbereitschaft während der Erstellung meiner Arbeit bedanke ich mich bei Herrn Dipl.-Ing. P. Schwarzmann und den Mitarbeitern der Bildverarbeitungsgruppe am „Institut für Physikalische Elektronik“ in Stuttgart.

Herzlich danke ich meinen Kollegen der Abteilung FT3/AB, insbesondere Herrn Dr.-Ing. R. Ott, Herrn Dr. rer. nat. W. Ritter und Herrn Dr.-Ing. U. Kreßel, die in zahlreichen Diskussionen zum Gelingen meiner Arbeit beigetragen haben.

Inhaltsverzeichnis

1	Einleitung	1
1.1	Bildverarbeitungssysteme zur Verkehrsszenenanalyse	1
1.2	Objektdetektion in Verkehrsszenen	2
1.3	Lösungsansatz zur Objektdetektion durch Bildfolgenanalyse	3
1.4	Gliederung	6
2	Stand der Technik	8
2.1	Objektdetektion in Verkehrsszenen	8
2.1.1	Binokulare Stereoverfahren	8
2.1.2	Modellierung des Erscheinungsbildes von Objekten in Einzelbildern	9
2.1.3	Bewegungsbasierte Verfahren	10
2.1.4	Bewertung der Verfahren	11
2.2	Bewegungsbestimmung in Bildfolgen	12
2.2.1	Differentielle Verfahren	13
2.2.2	Frequenzbasierte Verfahren	17
2.2.3	Korrespondenzverfahren	18
2.2.4	Bewertung der Verfahren	24
2.3	Bewegungssegmentierung	25
2.3.1	Optischer Fluß als Folge der Bewegung starrer Objekte	25
2.3.2	Modellierung des optischen Flusses für starre, planare Flächen . . .	27
2.3.3	Bewegungssegmentierung anhand affiner Bewegungsmodelle	28
2.3.4	Modellfreie Bewegungssegmentierung	30
2.3.5	Bewertung der Verfahren	30
3	Farbregionenfluß	31
3.1	Farbsegmentierung	31
3.1.1	Farbe	31
3.1.2	Aufgabe der Farbsegmentierung	33

3.1.3	Farbquantisierung durch Clusteranalyse	34
3.1.4	Zusammenhangsanalyse	43
3.2	Verfolgung von Farbregionen in Bildfolgen	43
3.2.1	Festlegung des Zuordnungsfehlers für Regionen	45
3.2.2	Festlegung der Obergrenzen für die Zuordnungsfehler	49
3.2.3	Prädiktion von Bewegungen	50
3.3	Objektdetektion durch Bewegungssegmentierung	58
3.4	Ergebnisse der Objektdetektion anhand des Farbregionenflusses	61
4	Farbclusterfluß	67
4.1	Motivation	67
4.2	Initiale Clusteranalyse	69
4.2.1	Eigenvektor Verfahren	69
4.2.2	Gewichtung zwischen Farbe und Position	69
4.3	Clusteranalyse von Folgebildern	72
4.3.1	Paralleles k-means Verfahren	72
4.3.2	Prädiktion der Referenzvektoren	75
4.4	Bewegungssegmentierung	77
4.5	Ergebnisse der Objektdetektion anhand des Farbclusterflusses	80
5	Zusammenfassung und Ausblick	88
5.1	Zusammenfassung	88
5.2	Ausblick	90
A	Struktogramme	92
A.1	Zuordnung von Regionen im Farbregionenfluß	92
A.2	Bewegungssegmentierung im Farbregionenfluß	93
A.3	Beschleunigung des parallelen k-means Clusterverfahrens	94
B	Kalmanfilterung	95
B.1	Modellierung der Beschleunigung als weißes Rauschen	95
B.2	Modellierung der Ableitung der Beschleunigung als weißes Rauschen	98

Verzeichnis der wichtigsten Abkürzungen und Symbole

MSE	Mittlerer quadratischer Fehler
SSE	Summe der quadratischen Fehler
A	System- bzw. Transitionsmatrix in der Kalmanfilterung
B	Steuermatrix in der Kalmanfilterung
C	Meßmatrix in der Kalmanfilterung
\mathcal{C}	Menge aller Datenvektoren eines Clusters
E_C	Maß für die Abweichung der Farben zweier Regionen
E_F	Maß für die Abweichung des Flächeninhalts zweier Regionen
E_P	Maß für die Abweichung in der Position des Schwerpunkts zweier Regionen
E_S	Maß für die Abweichung des Seitenverhältnisses zweier Regionen
E	Maß für den Unterschied zwischen zwei Regionen
F	System- bzw. Transitionsmatrix bei einem zeitkontinuierlichen System in der Kalmanfilterung
$f(x, y, t)$	Bildintensität in Abhängigkeit der Bildkoordinaten und der Zeit
$g(x, y)$	Intensitätsmuster in Abhängigkeit der Bildkoordinaten
K	Kalman-Gain Matrix in der Kalmanfilterung
k	Zeitindex
λ	Wellenlänge des Lichts
ξ	Positionskoordinate
ξ_x, ξ_y	Schwerpunktkoordinaten
P	Schätzfehlerkovarianzmatrix in der Kalmanfilterung
Φ_X, Φ_Y, Φ_Z	Rotationswinkel in bezug auf das Kamerakoordinatensystem
$p(\cdot)$	Wahrscheinlichkeitsdichte
$p(\cdot \cdot)$	Bedingte Wahrscheinlichkeitsdichte

Q	Kovarianzmatrix des Systemrauschens in der Kalmanfilterung
R, G, B	Farbvektor
R	Kovarianzmatrix des Meßrauschens in der Kalmanfilterung
r	Referenzvektor in der Clusteranalyse
ρ	Korrelation der Trajektorien zweier Regionen
S_X, S_Y, S_Z	Verschiebung in bezug auf das Kamerakoordinatensystem
s	Datenvektor in der Clusteranalyse
σ_x^2	Varianz von x
t	Zeit
T	Abtastperiode
U_B	Überlappungsgrad zwischen Objekt und Objekthypothese in bezug auf das umschreibende Rechteck
U_P	Überlappungsgrad zwischen Objekt und Objekthypothese in bezug auf Bildpunkte
u	Steuergrößen in der Kalmanfilterung
V_X, V_Y, V_Z	Translationsgeschwindigkeit in bezug auf das Kamerakoordinatensystem
v	In Kapitel 2: Optischer Fluß In Kapitel 3: Meßrauschen in der Kalmanfilterung
w	Systemrauschen in der Kalmanfilterung
w	Gewichtung der Positionsmerkmale im Farbclusterfluß
X, Y, Z	Koordinatensystem der Kamera
x, y	Koordinaten in der Bildebene
x	Systemzustandsgrößen in der Kalmanfilterung
$\Delta x, \Delta y$	Bild-zu-Bild Verschiebung in der Bildebene
y	Meßgrößen in der exponentiellen Mittelung und in der Kalmanfilterung
$\Omega_X, \Omega_Y, \Omega_Z$	Rotationsgeschwindigkeit in bezug auf das Kamerakoordinatensystem
$\arg \min$	Minimierendes Argument
$\delta(\cdot)$	Kontinuierliche Diracfunktion
\det	Determinante
∇	Nablaoperator in bezug auf Bildkoordinaten: $\nabla = (\frac{\partial}{\partial x}, \frac{\partial}{\partial y})^T$
$\ \cdot \ $	Euklidische Norm

1 Einleitung

1.1 Bildverarbeitungssysteme zur Verkehrsszenenanalyse

Mit der Zunahme des Straßenverkehrs sind die Anforderungen an die Verkehrsteilnehmer gestiegen. In der Automobilbranche wird deshalb seit einigen Jahren die Forschung an Fahrerassistenzsystemen vorangetrieben. Ziel dieser Systeme ist die Entlastung des Fahrers durch Übernahme bestimmter Teilaufgaben der Fahrzeugführung, wie z.B. Abstandshaltung und Spurführung. Um adäquat auf die Geschehnisse im Straßenverkehr reagieren zu können, sind Fahrerassistenzsysteme auf Sensoren zur Erfassung der Umwelt angewiesen. Von besonderem Interesse sind dabei optische Sensoren, da optische Information *die* Grundvoraussetzung zur Fahrzeugführung ist. Hinzu kommt, daß durch die vielseitigen Einsatzmöglichkeiten optischer Sensoren ein Markt vorhanden ist, der niedrige Stückpreise und eine kontinuierliche technische Weiterentwicklung von Kameras sicherstellt. Einer weiten Verbreitung von optischen Fahrerassistenzsystemen stehen bisher jedoch die Anforderungen an die Bildinterpretation und Systemzuverlässigkeit, in Verbindung mit der Komplexität des Anwendungsszenarios, im Wege. So stellen widrige Witterungsverhältnisse, Variationen in der Beleuchtung und die Vielfalt der in natürlichen Szenen vorhandenen Objekte höchste Ansprüche an die Bildverarbeitung.

Ein wesentliches Leistungsmerkmal von Bildverarbeitungssystemen für Anwendungen im Straßenverkehr ist deren Fähigkeit zur Interpretation der durch die Kamera erfaßten Verkehrsszenen. Die Schwierigkeit besteht dabei in der Reduktion einer großen Menge von sensorischen Daten auf wenige relevante Informationen, z.B. Verlauf der Fahrbahn oder die Position der anderen Verkehrsteilnehmer.

1.2 Objektdetektion in Verkehrsszenen

Einen wichtigen Beitrag zur Bildanalyse und Datenreduktion leistet die Detektion relevanter Objekte. Unter der Detektion eines Objekts versteht man die Extraktion einer Gruppe von Bildpunkten im Rasterbild, die das Abbild des realen Objekts in der Bildebene darstellt. Im folgenden wird das zu einem Objekt gehörige Abbild auch als Bildobjekt bezeichnet. Die Datenreduktion in der Detektionsstufe findet in zweierlei Hinsicht statt. Zum einen werden irrelevante Bildbereiche ausgesondert, zum andern erfolgt eine Zusammenfassung von Bildpunkten zu Bildobjekten und damit ein erster Schritt in Richtung Interpretation der Bilddaten. Eine genaue Untersuchung der detektierten Bildobjekte bleibt jedoch meist einer nachgeschalteten Erkennungsstufe vorbehalten.

Relevante Objekte für Fahrerassistenzsysteme sind Straße, Verkehrszeichen und Verkehrsteilnehmer; die vorliegende Arbeit beschäftigt sich mit der Detektion von Verkehrsteilnehmern. Die typische Vorgehensweise bei der Lösung einer solchen Detektionsaufgabe ist die Suche nach gemeinsamen Eigenschaften (z.B. Form- oder Farbmerkmale) der zu detektierenden Objekte, durch die sie in den Bilddaten lokalisiert werden können. Das Zusammenfügen dieser Eigenschaften bezeichnet man als Objektmodellierung. Man unterscheidet dabei zwischen der Modellierung der realen Objekte, die zu 3D-Modellen führt und der 2D-Modellierung der Bildobjekte. Schon bei der Einschränkung der zu detektierenden Objekte auf die Gruppe der Kraftfahrzeuge, trifft man jedoch auf eine enorme Vielfalt an Formen und Farben. Erschwerend kommt hinzu, daß Fahrzeuge in Verkehrsszenen unter verschiedensten Blickwinkeln und Beleuchtungsverhältnissen aufgenommen werden. Auf der Autobahn sieht man bei einer nach vorne schauenden Kamera im wesentlichen nur Heckansichten von Fahrzeugen. Auf Landstraßen wird man bei Gegenverkehr auch Fahrzeuge von vorne sehen. Im Bereich einer Kreuzung kommen schließlich auch Seitenansichten von Fahrzeugen vor. Ansätze, basierend auf 2D- bzw. 3D-Objektmodellierung, sind angesichts des Variationsreichtums der Szenarien kompliziert.

Als Alternative soll deshalb ein Ansatz gefunden werden, mit dem sich Verkehrsteilnehmer unabhängig von ihrer Form- und Farbgebung detektieren lassen. Es sollen dabei vor allem Verkehrsteilnehmer mit einer signifikanten Relativbewegung zur Kamera detektiert werden, da von ihnen ein erhöhtes Gefahrenpotential im Hinblick auf mögliche Kollisionen ausgeht. Für die Durchführung von experimentellen Untersuchungen zur Objektdetektion steht eine unkalibrierte 3-Chip CCD-Farbkamera zur Verfügung, die in einem Fahrzeug montiert ist.

1.3 Lösungsansatz zur Objektdetektion durch Bildfolgenanalyse

Robuste Detektionsmöglichkeiten ergeben sich, wenn man statt der Einzelbilder eine zeitliche Folge von Bildern betrachtet. Bewegung zwischen Kamera und der aufgenommenen Szene führt zu zeitlichen Änderungen des Bildinhalts in einer Bildfolge. Durch Analyse dieser Änderungen läßt sich Bewegungsinformation zurückgewinnen. Man unterscheidet dabei zwischen Bewegungsinformation, die sich auf die Bewegung in der 2D-Bildebene (Bildbewegung) bezieht und der Information über die reale 3D-Bewegung in der Szene. Da die Berechnung zuverlässiger 3D-Bewegungsinformation aus einer Bildfolge ohne Kenntnisse über Eigenbewegung der Kamera und 3D-Geometrie der Szene ein weitgehend ungelöstes Problem darstellt, beschränkt sich die vorliegende Arbeit auf Betrachtungen in der 2D-Bildebene. Anhand der Bildbewegung kann in einer Bewegungssegmentierung in jedem Einzelbild eine Gruppierung von Bildpunkten zu Bewegungssegmenten durchgeführt werden. Diese Gruppierung basiert lediglich auf der Bewegung der Objekte und, bedingt durch die Abbildungseigenschaft der Kamera, auf der 3D-Geometrie der Szene. Ein zu detektierendes Objekt wird, soweit es sich in seiner Bewegung von benachbarten Objekten unterscheidet, im bewegungssegmentierten Bild als ein eigenständiges Bewegungssegment repräsentiert – Bildobjekt und Bewegungssegment sind identisch. Damit ist bereits ein wichtiger Schritt in Richtung der Detektion von Verkehrsteilnehmern getan, nämlich die Gruppierung von Bildpunkten zu Bildobjekten. Jetzt fehlt noch die Selektion jener Bewegungssegmente, die den Bildobjekten von Verkehrsteilnehmern entsprechen. Hier macht man sich zu nutze, daß sich die zu detektierenden Verkehrsteilnehmer signifikant gegenüber der Kamera bewegen. Folglich werden nur solche Bewegungssegmente selektiert, die eine deutliche Bildbewegung aufweisen. Nachdem der Weg für eine bewegungsbasierte Detektion vorgezeichnet ist, wird im folgenden auf die Realisierung eingegangen.

In der vorliegenden Arbeit werden zwei als Farbregionenfluß [Heisele & Ritter 95] und Farbclusterfluß [Heisele et al. 97] bezeichnete Verfahren entworfen, realisiert und untersucht. Im Unterschied zur klassischen Vorgehensweise der Bildfolgenanalyse wird in beiden Verfahren die Bestimmung der Bildbewegung nicht punktweise in Grauwertbildern, sondern durch Beobachtung des zeitlichen Verhaltens von Farbgebieten in einer Farbbildfolge durchgeführt. Die Verwendung von Farb- statt Grauwertinformation ist essentiell, da sie eine bessere Trennung benachbarter Bildobjekte erlaubt. Durch eine Farbsegmentierung erfolgt in jedem Einzelbild eine Gruppierung von Bildpunkten ähnlicher Farbe zu Farbgebieten. Der Farbsegmentierung liegt dabei die Annahme zugrunde, daß die Bildpunkte einer farblich homogenen Bildregion zum selben Bildobjekt gehören und somit in einem Farbgebiet zusammengefaßt werden können. Ein Farbgebiet stellt, soweit die getroffene

Annahme erfüllt ist, den Teil *eines* Bildobjekts dar. Beispielsweise werden bei der Heckansicht eines Fahrzeugs die Rückleuchten, die Karosserie und das Nummernschild jeweils durch getrennte Farbgebiete repräsentiert. Durch die Farbsegmentierung erfolgt eine frühe Datenreduktion – von zigtausenden Farbbildpunkten zu einigen hundert Farbgebieten – so daß in der Bewegungsbestimmung das Datenaufkommen und damit der Rechenaufwand gering ist. Statt die Bildbewegung für jeden Bildpunkt zu bestimmen, beschränkt sich die Bewegungsbestimmung auf wenige Farbgebiete. Als Resultat liefert die Bewegungsbestimmung die Trajektorien der Gebietsschwerpunkte in der Bildebene. Entscheidend ist, daß durch die gebietsbasierte Vorgehensweise die Bildbewegung über den gesamten Bildbereich bestimmt wird, wobei Diskontinuitäten in der Bildbewegung, die zur Abgrenzung unterschiedlicher Bildobjekte erforderlich sind, erhalten bleiben. Den letzten, von der Bewegungsbestimmung unabhängigen, Schritt bildet die Bewegungssegmentierung. In ihr werden Gebiete, die ähnliche Trajektorien besitzen und in der Bildebene benachbart sind, zu Bewegungssegmenten zusammengefaßt. Da Objekte mit signifikanter Relativbewegung zur Kamera detektiert werden sollen, werden in der Bewegungssegmentierung nur Farbgebiete berücksichtigt, deren Trajektorien eine gewisse Mindestlänge überschreiten. Nachdem das gemeinsame Konzept des Farbregionen- und Farbclusterflusses vorgestellt wurde, soll im folgenden auf die in Bild 1.1 dargestellten Verarbeitungsschritte der beiden Verfahren näher eingegangen werden.

Prinzipiell unterscheiden sich Farbregionen- und Farbclusterfluß in der Wahl der Farbsegmentierung, die zu unterschiedlich gearteten Farbgebieten führt. Beim Farbregionenfluß entstehen als Resultat einer klassischen Farbsegmentierung Regionen, deren Bildpunkte im segmentierten Bild einheitliche Farbe besitzen und eine zusammenhängende Fläche bilden. Durch eine Farbzusammenhangsanalyse wird das segmentierte Bild ohne Informationsverlust in eine symbolische Beschreibung umgewandelt. Diese Beschreibung erlaubt, im Gegensatz zum Rasterbild, einen effizienten Zugriff auf die Eigenschaften einer Region, wie z.B. Kontur, umschreibendes Rechteck, Schwerpunkt usw. In der Zuordnungsprozedur werden Paare einander entsprechender Regionen in zwei aufeinanderfolgenden Bildern anhand einem auf Regioneneigenschaften definierten Ähnlichkeitsmaß bestimmt. Eine Stabilisierung der Zuordnungsprozedur kann durch Prädiktion der Bewegungen von Farbregionen erzielt werden. Als Resultat der Regionenzuordnung erhält man schließlich die Trajektorien der Farbregionen. Den Abschluß bildet die Bewegungssegmentierung, sie erzeugt die Objekthypothesen.

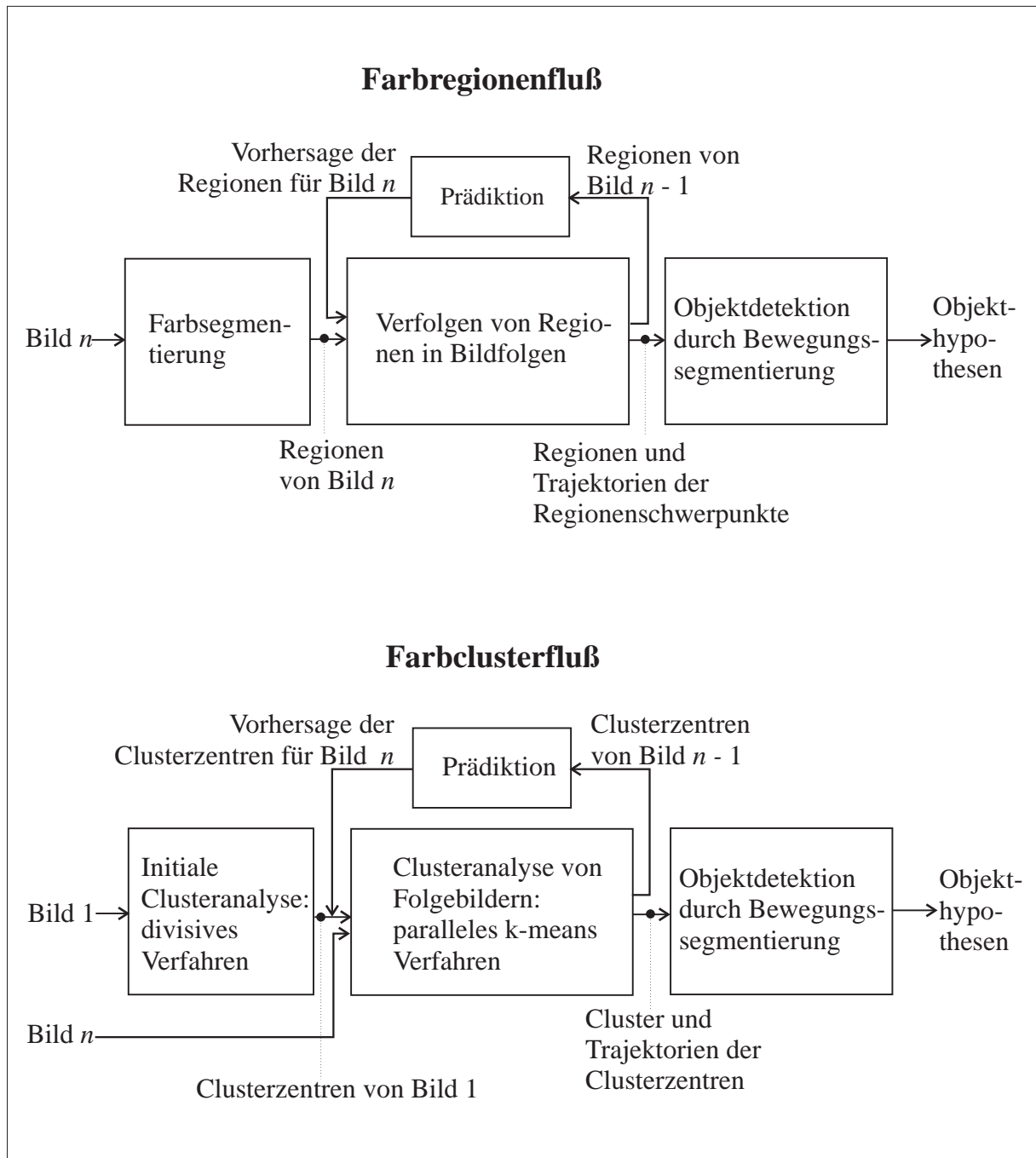


Bild 1.1: Prinzipielle Verarbeitungsschritte des Farbregionen- und Farbclusterflusses

Beim Farbclusterfluß wird durch eine Ballungsanalyse (Clusteranalyse) eine Partitionierung der Bildpunkte bezüglich ihrer Position in der Bildebene und ihrer Farbe vorgenommen. Jede Partition (Cluster) ist durch ihren Schwerpunkt im kombinierten Farb-Positionsraum eindeutig bestimmt. Durch Austauschen der Originalfarben der Bildpunkte mit den Farben der zugehörigen Clusterschwerpunkte entsteht das segmentierte bzw. geclusterte Bild. Entscheidend beim Farbclusterfluß ist die Clusteranalyse von Folgebildern; in ihr werden die Ergebnisse der Clusteranalyse des vorhergehenden Bildes für die Clusteranalyse des aktuellen Bildes verwendet. Durch diese sog. dynamische Segmentierung wird eine zeitliche Konsistenz der Cluster garantiert. Darüberhinaus impliziert die dynamische Segmentierung eine Zuordnung entsprechender Cluster in Folgebildern und damit eine Bewegungsbestimmung der Clusterschwerpunkte. Auch beim Farbclusterfluß können Prädiktionsverfahren zur Vorhersage der Position der Clusterschwerpunkte in der Bildebene eingesetzt werden. Analog zur Vorgehensweise beim Farbregionenfluß erfolgt abschließend eine Bewegungssegmentierung mit der Bildung von Objekthypothesen.

Die Unterschiede in der Wahl der Bildelemente, Regionen beim Farbregionen- und Cluster beim Farbclusterfluß, führen zu unterschiedlichen Charakteristiken der beiden Verfahren. Ohne detaillierte Vergleiche vorwegzunehmen sei hier erwähnt, daß der Farbregionenfluß in der Regel exaktere Bewegungsinformation liefert. Bei formveränderlichen Objekten (z.B. Fußgänger) oder teilweise verdeckten Objekten, erweist sich der Farbclusterfluß als das robustere Verfahren.

1.4 Gliederung

Im zweiten Kapitel wird ein Überblick über bestehende Verfahren zur Objektdetektion in Verkehrsszenen, Bewegungsbestimmung und Bewegungssegmentierung in Bildfolgen gegeben.

Das dritte Kapitel befaßt sich mit dem Farbregionenfluß. Einen Schwerpunkt bildet die Farbsegmentierung. Es werden verschiedene Clusterverfahren zur Farbquantisierung vorgestellt. Weiterhin werden störungsunterdrückende Verfahren zur Nachbearbeitung farbquantisierter Bilder vorgestellt. Den Abschluß der Farbsegmentierung bildet die Zusammenhangsanalyse; sie transformiert das segmentierte Punktbild in eine symbolische Darstellung. Danach wird auf die Zuordnung von Farbregionen in Folgebildern eingegangen. Es wird untersucht, welche Flächenmerkmale sich für die Zuordnung eignen und wie sich die Zuordnung zur Prädiktion der Flächenbewegungen stabilisieren läßt. Das Kapitel wird durch die Bewegungssegmentierung, in der die eigentliche Objektdetektion stattfindet und einer Evaluierung der Detektionsleistung abgeschlossen.

Das vierte Kapitel widmet sich dem Farbclusterfluß. Der erste Teil befaßt sich mit der Motivation für die Entwicklung einer Alternative zum Farbregionenfluß. Im zweiten Teil wird die initiale, ausschließlich auf das erste Bild einer Sequenz angewandte, Clusteranalyse beschrieben. In diesem Zusammenhang wird auf die für das Gesamtverfahren relevante Gewichtung zwischen Farb- und Ortsraum eingegangen. Den zweiten Teil bildet die Folgebildclusteranalyse, deren Aufgabe die Bewegungsbestimmung der Cluster in Bildsequenzen ist. Wie beim Farbregionenfluß wird der Einsatz eines Prädiktionsverfahrens zur Verbesserung der Bewegungsbestimmung untersucht. Die Bewegungssegmentierung, die weitgehend mit der des Farbregionenflusses übereinstimmt, wird kurz beschrieben. Den Abschluß des Kapitels bilden Untersuchungen an Bildfolgen und eine Beurteilung des Verfahrens.

In Kapitel fünf wird die Arbeit mit der Zusammenfassung und einem Ausblick abgeschlossen.

2 Stand der Technik

2.1 Objektdetektion in Verkehrsszenen

In diesem Abschnitt werden die grundlegenden Bildverarbeitungstechniken vorgestellt, die sich zur Detektion von Verkehrsteilnehmern verwenden lassen.

2.1.1 Binokulare Stereoverfahren

Bei binokularen Stereoverfahren (siehe z.B. [Shirai 87]) wird mit einer kalibrierten Anordnung zweier Kameras Tiefeninformation über die im Blickfeld beider Kameras befindliche Szene gewonnen. Durch optische Triangularität ändert sich mit der Entfernung eines Szenenpunktes vom Kamerapaar der Abstand zwischen den Abbildungen des Punktes in den beiden Bildern. Ist dieser Abstand, auch Disparität genannt, erst einmal bestimmt, so können über die Stereogleichung die 3D-Koordinaten des Szenenpunktes bestimmt werden. Das eigentliche Problem besteht somit darin, jeweils Paare von Bildpunkten zu finden, die Abbildungen desselben Szenenpunktes sind. Dies wird auch als Korrespondenzproblem der Stereoskopie bezeichnet. Die Tatsache, daß alle potentiellen Szenenpunkte, die auf genau einen Bildpunkt im *einen* Bild projiziert werden, auf eine berechenbare Epipolarlinie im anderen Bild abgebildet werden, vereinfacht das Korrespondenzproblem. Trotzdem ist eine zuverlässige Paarbildung in der Regel nur an markanten Punkten im Bild möglich, z.B. an Eck- oder Kantenpunkten. Dies hat Lücken im berechneten Tiefenbild zur Folge.

Stereobasierte Verfahren zur Objektdetektion in Verkehrsszenen benötigen zur Trennung zwischen Fahrbahn und den darauf befindlichen Objekten entweder das Wissen über den Verlauf der Fahrbahnoberfläche oder zumindest eine Modellvorstellung davon. Abweichungen des tatsächlichen Tiefenbildes von dem der leeren Fahrbahnoberfläche werden dann als Hinweis auf Objekte interpretiert. Es werden sog. erhabene Objekte detektiert, die sich auf der Straße befinden und eine Ausdehnung vertikal zur Straßenoberfläche besitzen. Neben den Verkehrsteilnehmern gehören zu dieser Objektgruppe auch Hindernisse auf der

Fahrbahn. Durch inverse perspektivische Transformation wird in [Brauckmann et al. 94, Koller et al. 94] eines der beiden Kamerabilder so transformiert, daß die Disparitäten für Szenenpunkte verschwinden, die auf der modellierten Fahrbahnoberfläche liegen. Punkte, an denen eine von null verschiedene Disparität bestimmt wird, werden als Objektpunkte betrachtet. In [Franke & Kutzbach 96] wird auf die inverse perspektivische Transformation verzichtet, statt dessen werden Disparitäten, die durch eine leere Fahrbahn hervorgerufen würden, mit den aus dem Bildpaar bestimmten Disparitäten verglichen. Im Fall signifikanter Abweichungen werden Punkte als Objektpunkte klassifiziert.

Stereobasierte Verfahren benötigen lediglich Wissen über den Verlauf der Fahrbahnoberfläche, um beliebig geformte, erhabene Objekte detektieren und deren Position in der Szene bestimmen zu können. Ein Nachteil stereobasierter Verfahren ist das relativ kleine Sichtfeld, das auf den Überlappungsbereich der Sichtfelder der beiden Kameras beschränkt ist. Die Gruppierung von einzelnen Objektpunkten im Tiefenbild zu Objekthypothesen, die letztlich das Ziel der Objektdetektion ist, wird meist durch Lücken im Tiefenbild erschwert. Beim Einsatz im Fahrzeug bereiten die hohen Anforderungen an die mechanische Stabilität der kalibrierten Kameraanordnung Probleme.

2.1.2 Modellierung des Erscheinungsbildes von Objekten in Einzelbildern

Diese Verfahren führen eine Detektion von Objekten durch, basierend auf dem Erscheinungsbild der Objekte in der Bildebene. Man unterscheidet zwischen Verfahren, die anhand von 3D-Objektmodellen auf das Aussehen der Bildobjekte schließen und Verfahren, die lediglich die 2D-Eigenschaften der Bildobjekte modellieren.

Zur ersten Gruppe zählen die in [Schmid 94, Dickmanns 93, Dickmanns 94, Graefe 93] vorgestellten Verfahren. Dort werden 3D-Drahtgittermodelle von Fahrzeugen generiert, deren Projektionen in die Bildebene zu 2D-Kantenmodellen führen. In der Detektionsstufe wird dann im kantensegmentierten Grauwertbild nach Anordnungen von Kanten gesucht, die Ähnlichkeiten mit den 2D-Kantenmodellen von Fahrzeugrückansichten aufweisen. In [Zielke et al. 92, Schwarzinger et al. 92] wird auf eine 3D-Modellierung verzichtet. Für die Detektion von vorausfahrenden Fahrzeugen werden lediglich die Symmetrieeigenschaften von Fahrzeugrückansichten im Grauwertbild genutzt.

Die oben genannten Verfahren haben sich in einfachen Verkehrsszenarien (z.B. Autobahn-szenario) bewährt, in denen die Vielfalt der Objekte und der Betrachtungswinkel, unter dem sie von der Kamera gesehen werden, eingeschränkt sind. Szenarien, in denen starke Variationen im Erscheinungsbild von Objekten auftreten, z.B. innerstädtisches Szenario,

erfordern aufwendige Modelle. Mit der Komplexität der Modelle steigt der Rechenaufwand und sinkt die Zuverlässigkeit der Detektion.

2.1.3 Bewegungsbasierte Verfahren

Bewegungsbasierte Verfahren nutzen die in einer monokularen Bildfolge vorhandene Information, um Objekte zu detektieren. Im einfachen Fall der stationären Kamera lassen sich bewegte Objekte durch Differenzbildung zwischen Grauwertbildern gewinnen. In [Koller et al. 93] wird das aktuelle Bild von dem aus der Bildfolge generierten Hintergrundbild abgezogen. Durch Binarisierung des Differenzbildes werden anschließend die bewegten Objekte extrahiert. In [Jain & Nagel 79, Chen & Chang 92, Gil et al. 96, Dubuisson & Jain 95] werden die Differenzen zwischen aufeinanderfolgenden Bildern zur Objektdetektion verwendet. Bei Bewegung der Kamera sind solche Differenzverfahren nicht mehr anwendbar. Statt dessen muß die 2D-Bildbewegung anhand der zeitlichen Bildänderungen geschätzt werden. Aus dieser läßt sich theoretisch Tiefen- und Bewegungsinformation über die in der Szene vorhandenen Objekte ableiten. In [Smith & Brady 94, Smith & Brady 95] erfolgt die Bewegungsbestimmung im Bild mit einem punktbasierten Korrespondenzverfahren an Eckpunkten des Grauwertbildes. In der anschließenden Bewegungssegmentierung erfolgt die Gruppierung von Eckpunkten, die ähnliche Bewegungsvektoren aufweisen, zu Bewegungssegmenten. Es wird nur dann ein Bewegungssegment in die Liste der Objekthypothesen aufgenommen, wenn sich seine Form in der Bildebene über einen Beobachtungszeitraum als stabil erweist. Dadurch lassen sich jedoch Falschdetektionen durch Strukturen auf der Straße (z.B. Fahrbahnmarkierungen) und Objekte am Straßenrand (z.B. Verkehrszeichen) nicht vermeiden.

Durch Modellierung der Fahrbahn als planare, horizontale Ebene und Kenntnis der Eigengeschwindigkeit, wird in [Enkelmann et al. 94, Krüger et al. 95] der durch die leere Fahrbahn hervorgerufene optische Fluß berechnet. An Grauwertecken der aufgenommenen Bildfolge werden die optischen Flußvektoren durch ein differentielles Verfahren zweiter Ordnung bestimmt und mit den an diesen Punkten erwarteten Flußvektoren der leeren Fahrbahn verglichen. Bildpunkte, an denen sich deutliche Abweichungen ergeben, werden als Objektpunkte klassifiziert und in einem weiteren Schritt zu Objektkanten zusammengefaßt. Damit lassen sich sowohl erhabene, stationäre Objekte, als auch beliebig geformte, bewegte Objekte auf der Fahrbahn detektieren.

Durch Differenzbilder kann man bei stehender Kamera bewegte Objekte vom stationären Hintergrund trennen. Komplizierter gestalten sich die Verfahren bei bewegter Kamera. Zeitliche Bildänderungen werden jetzt nicht nur durch bewegte Objekte in der Szene hervorgerufen, sondern auch durch die Eigenbewegung der Kamera. Allein durch die Bild-

bewegung ist eine Aufteilung des Bildes in Bildobjekte möglich. Durch Modellannahmen über den Straßenverlauf und Kenntnis der Eigenbewegung der Kamera lassen sich jene Bildobjekte selektieren, die erhabenen stationären Objekten oder bewegten Objekten auf der Fahrbahn entsprechen.

2.1.4 Bewertung der Verfahren

In Tabelle 2.1 sind die Vor- und Nachteile von Detektionsverfahren dargestellt. Ohne kalibrierte Kamera und ohne Straßenmodell lassen sich nur mit einem bewegungsbasierten Ansatz, wie er in [Smith & Brady 94] vorgeschlagen wird, Objekte unabhängig von ihrer Form und Farbe detektieren. Ein Nachteil des in [Smith & Brady 94] genannten Verfahrens ist, daß die Bildbewegung lediglich an Grauwertecken bestimmt wird – in weiten Teilen des Bildes liegt also keine Bewegungsinformation vor. Es stellt sich daher die Frage nach alternativen Methoden zur Bewegungsbestimmung. Im folgenden Abschnitt werden zu diesem Thema die in der Literatur bekannten Verfahren vorgestellt.

Tabelle 2.1: Vor- und Nachteile von Verfahren zur Objektdetektion

Verfahren	Vorteile	Nachteile
Stereo	+ Entfernung der Objekte + unabhängig von Objektform und -farbe	- zwei kalibrierte Kameras - Straßenmodell - kleines Sichtfeld
Modellbasiert	+ effektiv in einfachen Szenarien	- sehr aufwendige Modelle in komplexen Szenarien
Bewegungsbasiert ohne Eigenbewegung, ohne Straßenmodell	+ unabhängig von Objektform und -farbe	- Bildfolge nötig - viele Falschdetektionen
mit Eigenbewegung und Straßenmodell	+ unabhängig von Objektform und -farbe + Entfernung und Geschwindigkeit der Objekte	- Bildfolge nötig - kalibrierte Kamera - Messung der Eigengeschwindigkeit - Straßenmodell

2.2 Bewegungsbestimmung in Bildfolgen

Relative Bewegung zwischen Kamera und Szene kann durch ein 3D-Geschwindigkeitsvektorfeld beschrieben werden. Jeder Vektor eines Punktes der Szene gibt dessen Relativbewegung zur Kamera an. Als optischer Fluß wird in der vorliegenden Arbeit das Vektorfeld¹ bezeichnet, das durch Abbildung der 3D-Geschwindigkeitsvektoren aller sichtbaren Punkte der aufgenommenen Szene auf die 2D-Sensorfläche entsteht. Die Änderung der Position eines Punktes in zwei aufeinanderfolgenden Bildern wird durch den Verschiebungs- bzw. Disparitätsvektor angegeben. In der ursprünglichen, biologisch motivierten Definition des optischen Flusses [Nakayama & Loomis 74] erfolgte die Projektion auf eine kugelförmige Sensorfläche. Im Bereich der maschinellen Bildverarbeitung wird jedoch die zentralperspektivische Projektion auf eine planare Fläche als adäquate Modellierung der Abbildungseigenschaften gebräuchlicher Kameras verwendet. Da optischer Fluß keine meßbare Größe im Ausgangssignal der Kamera ist, führen Verfahren zur Bewegungsanalyse lediglich eine Schätzung des Flusses durch, basierend auf dem orts-zeitveränderlichen Bildsignal. Laut [Mitiche & Bouthemy 96] gibt es drei wesentliche Schwierigkeiten bei der Schätzung des Flusses:

- Unterschiedliche optische Flüsse können zum selben zeitlichen Bildsignal führen. Deutlich wird dies am Beispiel einer homogen beleuchteten Kugel mit idealen Lambert'schen Eigenschaften, die in der einen Aufnahme ruht und in der zweiten um die eigene Achse rotiert. Beide aufgenommenen Bildsignale weisen keinerlei zeitliche Änderungen auf. Sie sind identisch, obwohl ihr zugehöriger optischer Fluß verschieden ist.
- Relative Bewegung zwischen Kamera und Umwelt ist nicht die einzige Quelle für zeitliche Änderungen im Bildsignal. So können z.B. Änderungen in der Oberflächencharakteristik oder der Beleuchtung zu zeitlichen Änderungen im Bildsignal führen.
- Physikalische Modelle der Beleuchtungsquelle und der in der Szene vorhandenen Objekte, die zur exakten Herleitung der Bildänderungen notwendig wären, sind für die natürlichen Szenen meist nicht verfügbar oder zu komplex.

Bevor die wichtigsten Verfahren zur Schätzung des optischen Flusses vorgestellt werden, soll noch kurz auf Kriterien zur Beurteilung der Qualität der Verfahren eingegangen werden. Ein Maß für die Genauigkeit des Verfahrens bildet der Fehler zwischen geschätztem

¹In folgenden Textstellen der Arbeit, in denen eine Verwechslungsgefahr zwischen Bewegung in der 3D-Szene und Bewegung in der Bildebene besteht, werden die Bezeichnungen Szenenbewegung resp. Bildbewegung verwendet. Dasselbe gilt zur Unterscheidung zwischen den durch die Kamera gesehenen realen Objekten (Szenenobjekte) und deren Abbildungen in die Bildebene (Bildobjekte).

und tatsächlichem Fluß. In [Barron et al. 94] wird eine entsprechende Untersuchung an mehreren Verfahren durchgeführt. Probleme bereitet allerdings die Bestimmung des realen Flusses in natürlichen Videosequenzen, bei der die 3D-Geometrie der Szene bekannt sein muß. Umfangreiche quantitative Aussagen beziehen sich deshalb meist auf künstlich generierte Sequenzen und büßen dadurch erheblich an Aussagekraft ein. Neben diesem objektiven Gütemaß entscheiden anwendungsspezifische Kriterien über die Wahl eines Verfahrens. In der vorliegenden Arbeit soll die aus Bildsequenzen gewonnene Bewegungsinformation zur Detektion von Objekten in Verkehrsszenen herangezogen werden. Daraus lassen sich folgende Anforderungen an die Verfahren ableiten:

- Bei der Gruppierung von Bildpunkten zu Objekten ist eine hohe Dichte der Bewegungsinformation erwünscht. Im Idealfall existiert für jeden Bildpunkt ein Vektor des optischen Flusses.
- Diskontinuitäten des optischen Flusses befinden sich u.a. an der Grenze zwischen Szenenobjekten unterschiedlicher Bewegung. Sie sind für die Abgrenzung der Objekte in der Bewegungssegmentierung notwendig. Im geschätzten optischen Fluß müssen deshalb die Diskontinuitäten erhalten bleiben.
- Da in Verkehrsszenen schnell bewegte Objekte auftreten können, muß das Verfahren auch bei relativ großen Verschiebungsvektoren funktionsfähig sein.
- Das Verfahren muß als Echtzeitsystem auf einem Standardrechner realisierbar sein.

2.2.1 Differentielle Verfahren

Differentielle Verfahren bestimmen den optischen Fluß aus den Orts-Zeitableitungen der Intensität. Die Intensität läßt sich als Funktion $f(x, y, t)$ der Bildkoordinaten (x, y) und der Zeit t schreiben. Ein beliebiger physikalischer Punkt P der aufgenommenen Szene habe die Projektion p in der Bildebene mit den Koordinaten $(x(t), y(t))$. Bei einer Bewegung von P im Raum kann die Trajektorie von p in parametrisierter Form als $c(t) = (x(t), y(t), t)$ geschrieben werden. Die Intensität des Punktes P im Bild ist entsprechend $g(t) = f(x(t), y(t), t)$. In [Horn & Schunck 81] wird nun angenommen, daß die Intensität der Projektion p eines aufgenommenen Punktes P bei einer Bewegung im Raum konstant bleibt. Sie ist allerdings nur dann gültig, wenn konstante Beleuchtung, ideale Lambertstrahler und ausschließlich translatorische Szenenbewegungen vorliegen [Verri & Poggio 89]. Aus der Annahme der Konstanz der Intensität entlang einer Trajektorie läßt sich die sog. Gradientengleichung für den optischen Fluß ableiten:

$$\frac{dg}{dt} = 0 \quad \text{oder} \quad (2.1)$$

$$\frac{\partial f}{\partial x} \frac{dx}{dt} + \frac{\partial f}{\partial y} \frac{dy}{dt} + \frac{\partial f}{\partial t} = 0$$

Die Gradientengleichung wird häufig auch in der Form

$$\nabla f^T \cdot \mathbf{v} + f_t = 0 \quad (2.2)$$

geschrieben. Wobei $\mathbf{v} = (u, v)^T = (\frac{dx}{dt}, \frac{dy}{dt})^T$ der optische Fluß, $\nabla f = (\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y})^T$ der Ortsgradient und $f_t = \frac{\partial f}{\partial t}$ die partielle zeitliche Ableitung der Intensität ist.

Aus der Gradientengleichung werden zwei wesentliche Probleme differentieller Verfahren deutlich:

- Sensordaten gebräuchlicher Kameras bestehen aus rasterförmig angeordneten Bildpunkten (Pixeln), deren Daten in konstanten Zeitabständen abgetastet werden. Unterabtastung des Intensitätssignals im Orts- bzw. Zeitraum oder Störungen in den Bilddaten können, bei zu kleiner Wahl der zur numerischen Differentiation verwendeten lokalen Operatoren, zu erheblichen Fehlern im geschätzten Fluß führen. Aus diesem Grund wird vor der Differentiation meist eine zeitaufwendige Tiefpaßfilterung im Orts-Zeitraum durchgeführt.
- Das zweite Problem ist die Mehrdeutigkeit bei der Bestimmung des optischen Flusses nach Gl. (2.2). Es ist mit dem Aperturproblem verwandt, das besagt, daß nur die Bewegungskomponente senkrecht zu einer durch eine Apertur beobachteten, translatorisch bewegten Kante eindeutig bestimmt werden kann (siehe Bild 2.1). Entsprechend kann aus Gl. (2.2) lediglich die Projektion des optischen Flusses in Richtung des Ortsgradienten von f eindeutig ermittelt werden:

$$\mathbf{v}_\perp = \frac{-f_t}{\|\nabla f\|^2} \nabla f \quad (2.3)$$

Zur Einschränkung der Lösungen der Gradientengleichung müssen weitere bzw. andere Forderungen an den zu schätzenden optischen Fluß gestellt werden. In [Mitiche & Bouthemy 96] werden Forderungen unterschieden, die zu lokalen bzw. globalen Schätzungen des Flusses führen. Im globalen Fall sind die punktuellen Berechnungen von Flußvektoren untereinander gekoppelt. In lokalen Verfahren wird für jeden Vektor ein von den anderen Flußvektoren unabhängiges Gleichungssystem gelöst.

Globale Verfahren

In [Horn & Schunck 81] wurde die sog. globale Glattheitsforderung zur Bestimmung des optischen Flusses eingeführt. Sie besagt, daß jenes Vektorfeld am plausibelsten ist, das in

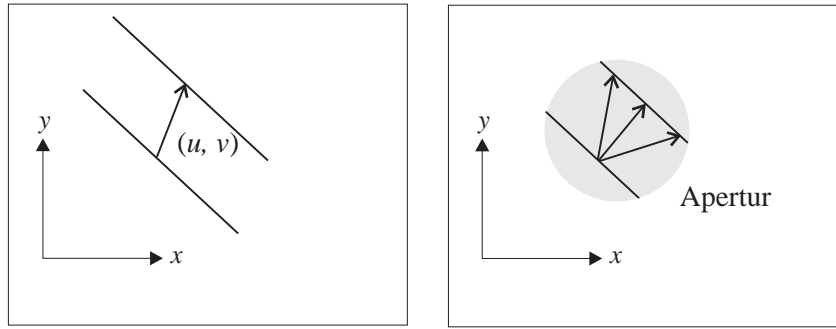


Bild 2.1: Mehrdeutigkeiten bei der Bestimmung des optischen Flusses bei einer lokalen Betrachtung durch eine kreisförmige Apertur

der Nähe einer Lösung der Gradientengleichung liegt, gleichzeitig aber möglichst „glatt“ verläuft. Die mathematische Formulierung führt zu einer über den gesamten Bildbereich D definierten Fehlerfunktion [Horn & Schunck 81]:

$$\iint_D (\nabla f^T \cdot \mathbf{v} + f_t)^2 + \lambda^2 (|\nabla u|^2 + |\nabla v|^2) dx dy \stackrel{!}{=} \min \quad (2.4)$$

Der linke Teil des Integrals berücksichtigt die Abweichung von (2.2); der rechte Teil bewertet die Glattheit der Verschiebungsvektoren. Aus der Vielzahl veröffentlichter Glattheitsforderungen [Schnörr 91, Tekalp 95] ist die in [Nagel & Enkelmann 86] untersuchte, sog. orientierte Glattheitsforderung, hervorzuheben. Dort wird die Glattheitsforderung in Richtung starker Intensitätsgradienten abgeschwächt. Der Grundgedanke ist, daß sich Objektgrenzen, an denen Unstetigkeiten des realen Flusses auftreten können, als Grauwertkanten im Intensitätsbild manifestieren. In [Snyder 91] erfolgt die allgemeine mathematische Formulierung einer globalen Glattheitsforderung, die bestimmten Bedingungen genügt, wie z.B. Unabhängigkeit von der Lage des x - y -Koordinatensystems. Das Problem der unerwünschten Glättung an Bewegungskanten hat zu einer Gruppe von regionenbasierten Verfahren geführt. Sie gehen davon aus, daß sich das Bild in Regionen aufteilen läßt, in denen der Verlauf des optischen Flusses durch einfache mathematische Modelle beschrieben werden kann. Die Globalität dieser Verfahren leitet sich aus der a priori nicht begrenzten Ausdehnung von Regionen innerhalb des Bildes ab. In [Black & Jepson 96] wird zur Bestimmung der Regionen eine Grauwertsegmentierung des Bildes vorgenommen. Statistische Ansätze, in denen die Aufteilung des Bildes in Regionen durch Markov-Random-Fields [Geman & Geman 84] modelliert wird, sind in [Bouthemy & François 93, Heitz & Bouthemy 93] zu finden.

Gemeinsam ist globalen Verfahren, daß sie zu einer dichten, gegenüber lokalen Störungen im Intensitätsbild unempfindlichen, Schätzung des optischen Flusses führen. Durch regio-

nenbasierte Vorgehensweise ist auch eine Bewahrung von Diskontinuitäten im geschätzten Fluß möglich. Die globalen Kopplungen bei der Bestimmung der Flußvektoren führen jedoch zu extrem zeitaufwendigen Optimierungsaufgaben.

Lokale Verfahren

Bei lokalen Verfahren sind die Bestimmungsgleichungen der Flußvektoren entkoppelt. Die punktuelle Schätzung eines Flußvektors wird somit nicht durch die Schätzungen des Flusses in der Umgebung des jeweiligen Punktes beeinflusst.

Eine Möglichkeit der lokalen Einschränkungen der Lösungen der Gradientengleichung ist die Verwendung von Multispektralbildern. Durch Anwendung von Gl. (2.2) auf die Intensitäten in den einzelnen Spektralkanälen erhält man einen Satz von Gleichungen für den optischen Fluß. In [Markandey & Flinchbaugh 90] werden beispielsweise dreikanalige Farbbilder verwendet.

Eine andere Vorgehensweise ist die Verwendung von Ableitungen zweiter Ordnung der Intensität. In [Tretiak & Pastor 84, Verri et al. 90] wird nicht die Konstanz der Intensität, sondern die des Intensitätsgradienten entlang einer Trajektorie vorausgesetzt. Diese Forderung ist allerdings nur dann erfüllt, wenn bestimmte Bewegungsformen im Bild, z.B. Rotation und Dilatation, nicht auftreten. Die mathematische Formulierung der Konstanzforderung lautet:

$$\begin{aligned} \frac{d}{dt} \nabla g = 0 \quad \text{oder} & \quad (2.5) \\ f_{xx}u + f_{xy}v + f_{xt} = 0 \\ f_{xy}u + f_{yy}v + f_{yt} = 0 \end{aligned}$$

Doppelte Indizes kennzeichnen lokale Ableitungen zweiter Ordnung. Offensichtlich besitzt Gl. (2.5) eine eindeutige Lösung für (u, v) , wenn die aus den lokalen Ableitungen zweiter Ordnung bestehende 2×2 Matrix eine von null verschiedene Determinante hat. Flußberechnungen, die nur auf Gl. (2.5) basieren, liefern aufgrund der hohen Störanfälligkeit der numerischen Ableitungen zweiter Ordnung in der Regel ungenauere Ergebnisse als Verfahren nach Gl. (2.2) [Barron et al. 94]. Weitere Probleme bereiten Gebiete, in denen die Intensität annähernd linear von x und y abhängt – dort verschwinden sämtliche Ableitungen zweiter Ordnung; der Vektor (u, v) kann beliebige Werte annehmen.

Eine dritte Verfahrensgruppe modelliert den Verlauf des optischen Flusses in kleinen Bildbereichen [Kearney et al. 87, Lucas 84]. Eine Fehlergleichung gibt die Abweichungen der Schätzung von der Gradientengleichung und dem Modell an. In [Lucas 84] wird der Fluß in der Nachbarschaft eines Punktes als konstant modelliert. Der Flußvektor \mathbf{v}_p an der

Stelle p läßt sich aus der Minimierung von

$$\iint_{D_p} w(x - x_0, y - y_0) \left(\nabla f^T \cdot \mathbf{v}_p + f_t \right)^2 dx dy \quad (2.6)$$

bestimmen. D_p ist eine um den Punkt p mit den Koordinaten (x_0, y_0) definierte Nachbarschaft in der Bildebene; $w(x, y)$ eine mittenbetonte Gewichtsfunktion. Es können auch Modelle mit linearem bzw. polynomialem [Kearney et al. 87] Verlauf eingesetzt werden. In allen Modellen wird eine Stetigkeit des Flusses gefordert, die eine lokale Glattheit impliziert. Die Regularisierung der Bestimmung des optischen Flusses mit lokalen Modellen ist nicht an die Gradientengleichung gebunden und kann durchaus auch in anderen Verfahren [Singh 90, Fleet & Jepson 90] eingesetzt werden.

Zusammenfassend kann man sagen, daß mit lokalen Verfahren in schwach strukturierten Bildbereichen keine zuverlässige Aussage über den optischen Fluß möglich ist. In der Regel wird in solchen Bereichen erst gar keine Schätzung durchgeführt. Im Vergleich zu globalen Verfahren entstehen somit weniger dichte Vektorfelder, die indes genauer sind und einen deutlich geringeren Rechenaufwand erfordern.

2.2.2 Frequenzbasierte Verfahren

Die in [Adelson & Bergen 86, Fleet 92, Heeger 88] vorgestellten Verfahren bestimmen den optischen Fluß durch Fouriertransformation der orts-zeitabhängigen Intensitätsfunktion. Betrachtet wird das Intensitätssignal $f(x, y, t)$, das durch Verschiebung eines zeitinvarianten Intensitätsmusters $g(x, y)$ mit konstanter Geschwindigkeit (u, v) erzeugt wird. Durch die Bewegung entsteht:

$$f(x, y, t) = g(x - ut, y - vt) \quad (2.7)$$

Die Anwendung der Fouriertransformation auf beide Seiten ergibt:

$$F(\omega_x, \omega_y, \omega_t) = G(\omega_x, \omega_y) \cdot \delta(u\omega_x + v\omega_y - \omega_t) \quad (2.8)$$

wobei Großbuchstaben die fouriertransformierten Funktionen

$$\begin{aligned} F &\bullet \longrightarrow f \\ G &\bullet \longrightarrow g \end{aligned}$$

symbolisieren. Die Frequenzen sind durch ω und die Diracfunktion durch $\delta(\cdot)$ dargestellt. Außerhalb der Ebene $u\omega_x + v\omega_y - \omega_t = 0$ verschwindet das Spektrum. In [Heeger 88] werden Paare von frequenzsensitiven Gaborfiltern zur Bestimmung des Verlaufs der oben

genannten Ebene eingesetzt:

$$h_{\sin(x,y,t)} = \frac{1}{\sqrt{2\pi^3}\sigma_x\sigma_y\sigma_t} \cdot e^{-\left(\frac{x^2}{2\sigma_x^2} + \frac{y^2}{2\sigma_y^2} + \frac{t^2}{2\sigma_t^2}\right)} \cdot \sin(2\pi\omega_{x_0}x + 2\pi\omega_{y_0}y + 2\pi\omega_{t_0}t) \quad (2.9)$$

$$h_{\cos(x,y,t)} = \frac{1}{\sqrt{2\pi^3}\sigma_x\sigma_y\sigma_t} \cdot e^{-\left(\frac{x^2}{2\sigma_x^2} + \frac{y^2}{2\sigma_y^2} + \frac{t^2}{2\sigma_t^2}\right)} \cdot \cos(2\pi\omega_{x_0}x + 2\pi\omega_{y_0}y + 2\pi\omega_{t_0}t) \quad (2.10)$$

Die Faltung von $f(x, y, t)$ mit den Gaborfilterfunktionen entspricht einer Bandpaßfilterung der Intensitätsfunktion um die Mittenfrequenz $(\omega_{x_0}, \omega_{y_0}, \omega_{t_0})$. Durch Integration über die Leistungsspektren des mit verschiedenen Mittenfrequenzen gefilterten Signals kann (u, v) ermittelt werden. Die in [Heeger 88] benutzten 12 Gaborfilterpaare, mit Filterkernen von 23 Bildpunkte \times 23 Bildpunkte \times 7 Bilder, lassen ein sehr zeitaufwendiges Verfahren vermuten. Durch Verwendung von Gaußfiltern zur numerischen Integration der quadrierten Gaborfilterausgänge entstehen relativ glatte Vektorfelder, ähnlich wie bei differentiellen Verfahren mit einer globalen Glattheitsforderung.

2.2.3 Korrespondenzverfahren

Korrespondenzverfahren führen keine Schätzung des optischen Flusses durch. Statt dessen werden die Verschiebungen $(\Delta x, \Delta y)$ von Bildelementen zweier im Abstand Δt aufgenommenen Bilder bestimmt. Der Zusammenhang mit dem optischen Fluß lautet demnach:

$$\begin{aligned} u &= \lim_{\Delta t \rightarrow 0} \frac{\Delta x}{\Delta t} \\ v &= \lim_{\Delta t \rightarrow 0} \frac{\Delta y}{\Delta t} \end{aligned} \quad (2.11)$$

Die Bestimmung der Verschiebungen erfolgt durch paarweise Zuordnung (Korrespondenzbildung) von Bildelementen in Folgebildern. Die Schwierigkeit einer korrekten Paarbildung, bei der zwei einander zugeordnete Bildelemente dasselbe physikalische Objekt in der aufgenommenen Szene darstellen müssen, taucht in der Literatur unter dem Schlagwort „Korrespondenzproblem“ [Ullman 79] auf. Hier wird festgestellt, daß es kein Korrespondenzverfahren gibt, das eine stets korrekte Zuordnung der Bildelemente unabhängig von der Komplexität der Szene gewährleistet. Eine Ursache dafür sind Verdeckungen, bei denen Bildelemente im Folgebild verschwinden bzw. neue Elemente auftauchen. Je nach Wahl der Bildelemente wird im folgenden zwischen statischen und adaptiven Korrespondenzverfahren unterschieden. In statischen Verfahren wird das Bild, unabhängig vom Bildinhalt, in stets gleichförmige Bildelemente unterteilt. Im Unterschied dazu ist bei adaptiven Verfahren die Anzahl und Position der in einem Vorverarbeitungsschritt gewonnenen Elemente vom jeweiligen Bild abhängig.

Statische Korrespondenzverfahren

Die bedeutendste Gruppe innerhalb der statischen Korrespondenzverfahren bilden die sog. Blockzuordnungsverfahren [Jain & Jain 81, Anandan 89, Burt et al. 83, Saito & Komatsu 94]. Als Bildelemente werden quadratische Bereiche (Blöcke) verwendet, die um jeden Bildpunkt des Intensitätsbilds ausgeschnitten werden. Im Folgebild wird jedem dieser Ausschnitte das am besten passende Gegenstück zugeordnet. Die Güte der Zuordnung wird über ein Ähnlichkeitsmaß bestimmt. Häufig wird dazu die gewichtete Summe der quadratischen Fehler (Sum of Squared Errors) zwischen zwei $m \times m$ großen Intensitätsmustern $g_1(x, y)$ und $g_2(x, y)$ verwendet:

$$SSE(x, y, \Delta x, \Delta y) = \sum_{j=-m}^{+m} \sum_{i=-m}^{+m} w(i, j) \cdot [g_1(x + i, y + j) - g_2(x + i + \Delta x, y + j + \Delta y)]^2 \quad (2.12)$$

wobei $w(i, j)$ die Fehler gewichtet. Der Verschiebungsvektor $(\Delta x, \Delta y)$ am Punkt (x, y) wird durch Minimierung des SSE bestimmt. Da die Minimierung des SSE für alle zulässigen Verschiebungen zu zeitaufwendig ist, muß der Suchbereich für $(\Delta x, \Delta y)$ eingeschränkt werden. Um dennoch große Verschiebungen erfassen zu können, wird in [Anandan 89] das Bild in einer Auflösungsrampe verarbeitet. Ein weiteres Problem von Gl. (2.12) ist, daß Verschiebungsvektoren nur als ganzzahlige Vielfache von Pixelabständen bestimmt werden können. Diese Einschränkung wird in [Anandan 89] durch Minimierung einer in x - und y -Richtung kontinuierlichen, quadratischen Approximation des SSE aufgehoben.

Alle Blockzuordnungsverfahren sind auf Kosten eines hohen Rechenaufwands in der Lage, an jedem Bildpunkt einen Verschiebungsvektor zu berechnen. Die durch Blockgröße und Suchbereiche vorgegebene Lokalität des Verfahrens führt zu denselben Einschränkungen wie bei lokalen differentiellen Verfahren: Mehrdeutigkeiten in der Bestimmung der Verschiebungsvektoren ergeben sich, wenn die Intensität in einem die Größe des Suchbereichs überschreitenden Bildabschnitt linear oder gar konstant verläuft. Weitere Probleme ergeben sich bei starker Variation der realen Verschiebungsvektoren innerhalb eines Blocks, wenn z.B. nicht nur translatorische Bildbewegung auftritt oder mehrere unterschiedlich bewegte Objektteile in einem Block liegen.

Adaptive Korrespondenzverfahren

In der Vorverarbeitung von adaptiven Korrespondenzverfahren werden in jedem Einzelbild Bildelemente extrahiert und durch Attribute beschrieben [Ullman 79]. Ziel der Vorverarbeitung ist es, die Elemente so zu bestimmen, daß sie sich leicht im Folgebild wiederfinden

lassen und damit zu einer zuverlässigen Verschiebungsschätzung führen. Prinzipiell kommen punktuelle, linienförmige und flächenhafte Bildelemente in Frage.

Punktzuordnungen Die Frage, an welchen Punkten eines Bildes eine eindeutige Bewegungsschätzung aus lokaler Bildinformation möglich ist, stellte sich bereits bei differentiellen Verfahren. Offenbar sind dies Punkte, die sich von ihren Nachbarpunkten durch bewegungsinvariante, lokale Eigenschaften abheben. Zu ihnen zählen beispielsweise Eckpunkte [Weng et al. 92] oder Punkte, an denen die Intensität ein lokales Extremum [Koller et al. 91] aufweist. In [Förstner & Gülch 87, Kahl et al. 78, Levine & O’Handley 73] werden Operatoren zur Extraktion solcher Punkte untersucht. Mit der Zuordnung bereits selektierter Punkte in zwei aufeinanderfolgenden Bildern beschäftigt sich [Barnard & Thompson 80]. Dort wird ein iteratives Verfahren vorgeschlagen, das die Zuordnung eines Punktes mit der Zuordnung von Punkten in dessen Umgebung koppelt. Die Koppelung ist so gewählt, daß sie zu einer lokalen Ausrichtung der Verschiebungsvektoren führt. Das als Relaxationszuordnung bekannte Verfahren weist Parallelitäten mit den globalen Glattheitsforderungen differentieller Verfahren auf.

Verfahren, basierend auf selektierten Punkten, sind dann von Nutzen, wenn eine präzise Bewegungsschätzung bei großen Bildbewegungen stattfinden soll. In der Regel erlaubt die relativ geringe Anzahl der selektierten Punkte keine dichte Bewegungsschätzung. Dies wird an dem einfachen Beispiel einer translatorisch bewegten, quadratischen Fläche klar (siehe Bild 2.2). Bei Betrachtung lokaler Eigenschaften von Punkten kann nur an den Ecken die Verschiebung eindeutig bestimmt werden (siehe Bild 2.3) – eine Folge des bereits erwähnten Aperturproblems.

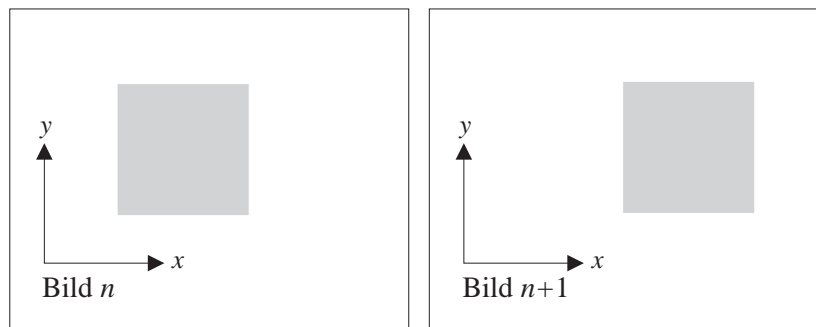


Bild 2.2: Laterale Bewegung eines Quadrates homogener Intensität

Linienzuordnungen In allen bisher vorgestellten Verfahren wird mit zunehmender Verkleinerung des zur punktuellen Bewegungsbestimmung benutzten Bildbereichs die Aperturproblematik verschärft. Andererseits führt eine bilddatenunabhängige Ausdehnung des

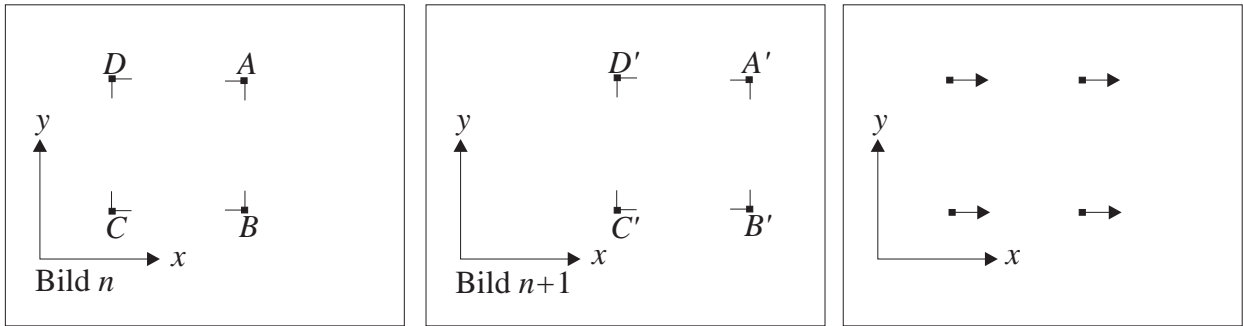


Bild 2.3: Bestimmung der Verschiebungsvektoren durch Zuordnung von Eckpunkten

Betrachtungsbereichs zu ungenauen Schätzungen an Diskontinuitäten des realen Flusses, also dort, wo Verdeckungen von Objekten in der Szene auftreten. Notwendig wäre ein Verfahren, das zuerst die Grenzen der Bildobjekte bestimmt und dann, durch Annahme (bzw. Kenntnis) über die Beschaffenheit des jeweils zugehörigen Szenenobjektes, die Bewegungsschätzung an den Punkten des Bildobjektes vereinfacht.

Diese Idee spiegelt sich in Linienzuordnungsverfahren wider. Unter der Annahme, daß benachbarte Kantenpunkte im Grauwertbild zur Kontur desselben Szenenobjektes gehören, erfolgt eine Gruppierung der Kantenpunkte zu linienförmigen Bildobjekten. Zur kompakteren Beschreibung der Bildobjekte werden in [Mitiche et al. 89, Mitiche & Habelrih 89, McIntosh & Mutch 88, Foresti & Murino 94] Linien durch Geradenstücke approximiert, die unabhängig voneinander in der Bildsequenz verfolgt werden. Durch die Bestimmung der Bewegung einer Kante kann dann auf die Bewegungen der einzelnen Kantenpunkte geschlossen werden.

Das Prinzip der Verfahren wird am Beispiel des translatorisch bewegten Quadrates aus Bild 2.2 deutlich. Angenommen, die vier Kanten entsprechen jeweils Konturstücken einer starren, planaren Fläche, dann können die Verschiebungsvektoren für alle Punkte der Seiten des Quadrates bestimmt werden (siehe Bild 2.4).

Regionenzuordnung Bei regionenbasierten Verfahren erfolgt eine Aufteilung der Bilder in Regionen, die bezüglich bestimmter Merkmale, wie z.B. Intensität oder Textur, ein Homogenitätskriterium erfüllen. Anschließend erfolgt eine Bewegungsbestimmung der Regionen durch Zuordnung einander entsprechender Regionen in Folgebildern. Wiederum wird vorausgesetzt, daß eine segmentierte und über eine Bildsequenz verfolgte Region stets dasselbe Szenenobjekt repräsentiert, bzw. dasselbe Teil eines Szenenobjektes. Hat man die Bewegung einer Region bestimmt, kann durch Annahmen über die Beschaffenheit des zugehörigen Szenenobjektes der Verlauf des Flusses innerhalb der Region modelliert werden. In Kapitel 2.3 wird ein affines Bewegungsmodell für starre, planare Flächen hergeleitet.

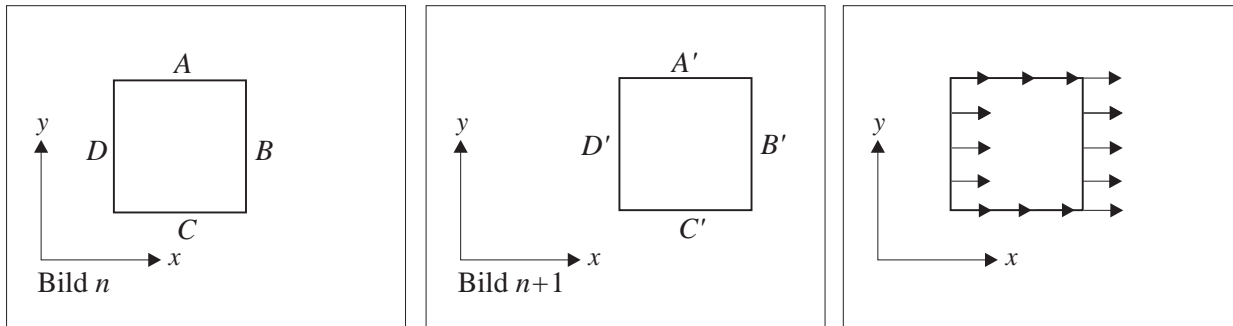


Bild 2.4: Bestimmung der Verschiebungsvektoren durch Kantenzuordnung

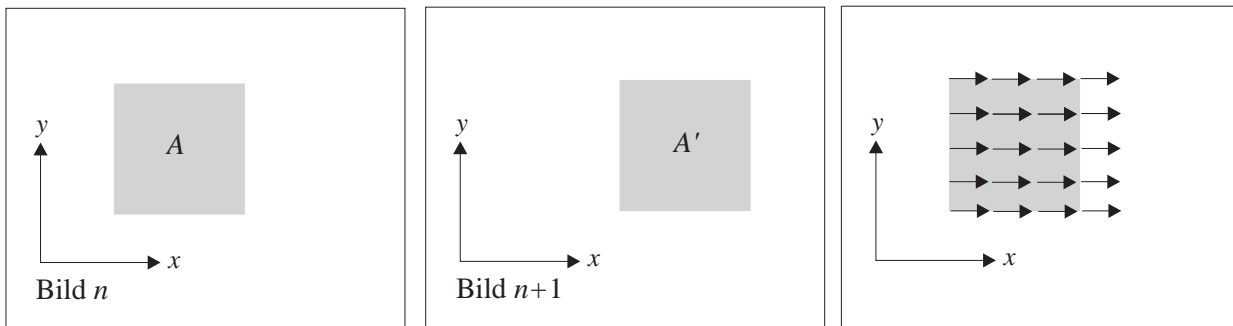


Bild 2.5: Bestimmung der Verschiebungsvektoren durch Regionenzuordnung

Durch Regionenzuordnung können nun auch die Verschiebungsvektoren innerhalb des translatorisch bewegten Quadrates bestimmt werden (siehe Bild 2.5). Dies wird durch die Kenntnis über die Zusammengehörigkeit aller Punkte des Quadrates zum selben Bildobjekt ermöglicht. Allerdings ist die Konstanz der Verschiebungsvektoren innerhalb des Quadrates nur dann gewährleistet, wenn das Quadrat die Abbildung einer planaren, starren Fläche ist.

In [Price & Reddy 77] wird ein Flächenzuordnungsverfahren zur Beschreibung von Unterschieden in Farbbildpaaren vorgestellt. Beide Bilder werden durch eine Farbsegmentierung in Regionen aufgeteilt. Anschließend wird jede Fläche durch einen Satz von Merkmalen wie Flächeninhalt, Farbe, Orientierung, Schwerpunkt usw. beschrieben. Die Zuordnung einander entsprechender Regionen basiert auf Gemeinsamkeiten der Regionenmerkmale. Ein mit Grauwertbildern arbeitendes Verfahren wird in [Nagel 78] untersucht. Zur Berechnung des Ähnlichkeitsmaßes wird die Standardabweichung und der Mittelwert der Intensitätsverteilungen der beiden Regionen in ihrem Überlappungsbereich herangezogen. Mit der Frage nach der Stabilität von Regionen in Bildfolgen beschäftigt sich [Healey 93]. Dort wird jedes Einzelbild in verschiedenen Auflösungen segmentiert. Nur jene Re-

gionen werden als stabil und für die Bewegungsbestimmung als geeignet erachtet, die sich in mehreren Auflösungsstufen des Einzelbildes wiederfinden. In [Chow & Aggarwal 77] wird die zeitliche Verfolgung translatorisch bewegter Regionen in einer Binärbildsequenz untersucht. Als Zuordnungsmerkmale werden der Flächeninhalt und eine rotationsunabhängige Kombination der Hauptachsen genutzt. Um durch Verdeckungen auftretende Verschmelzungen von Regionen erkennen zu können, wird eine Prädiktion der Regionenverschiebungen durchgeführt. Ebenfalls mit dem Thema der Regionenverfolgung in Binärbildern beschäftigen sich [Schalkoff & McVey 82, Legters & Young 82]. Für jede Region wird eine affine Transformation bestimmt, welche die Bild-zu-Bild Bewegung der Region approximiert. Zur Modellierung des zeitlichen Verlaufs der translatorischen und rotatorischen Komponenten der affinen Abbildung wird in [Legters & Young 82] ein Kalmanfilter eingesetzt. In [Meyer & Bouthemy 92, Meyer & Bouthemy 94] werden die Konturen einander entsprechender Flächen in Folgebildern zur Berechnung von affinen Transformationen verwendet. Zur zeitlichen Glättung und Prädiktion der Bewegung erfolgt eine Kalmanfilterung aller affinen Parameter. Affine Transformationen werden auch in [Kalivas & Sawchuk 91] zur Modellierung von Regionenverschiebungen in Grauwertbildern angewandt. Die Parameter der Transformation werden über iterative Maximierung einer Gütefunktion bestimmt. Die Güte wird durch die Größe des Überlappungsbereiches zwischen der aus dem vorhergehenden Bild in das aktuelle Bild transformierten Fläche und der ihr zugeordneten Fläche bestimmt. Ein interessanter Ansatz zur Vereinfachung des Korrespondenzproblems wird in [Bhanu & Burger 88] verfolgt. Indem zur Initialisierung der Segmentierung die Regionen des vorigen Bildes herangezogen werden, wird zum einen die Segmentierung über die Zeit stabilisiert und zum anderen die Zuordnung von Regionen bereits in der Segmentierung durchgeführt. In [Rehrmann & Rothaar 97] wird ein hierarchisches Regionenwachstumsverfahren zur Segmentierung von Farbbildfolgen vorgeschlagen. Die anschließende Bewegungsbestimmung erfolgt über eine n zu m Zuordnung der Regionen auf symbolischer Ebene. In einem abschließenden Schritt erfolgt eine bewegungsbasierte Segmentierung der Szene.

Im Unterschied zu den punkt- und linienbasierten Verfahren decken die Bildmerkmale regionenbasierter Verfahren den gesamten Bildbereich ab und erlauben dadurch eine flächendeckende Bewegungsschätzung. Gegenüber differentiellen und frequenzbasierten Verfahren sind der deutlich geringere Rechenaufwand, die Einsatzmöglichkeit auch bei großen Bildverschiebungen und die Wahrung von Diskontinuitäten als Vorteile zu nennen.

2.2.4 Bewertung der Verfahren

In Tabelle 2.2 sind die Vor- und Nachteile der Verfahren zur Bewegungsbestimmung dargestellt. Die Beurteilung orientiert sich an den auf Seite 13 genannten Anforderungen. Die beste Bewertung erzielen regionenbasierte Verfahren. Erfolgversprechend ist der in [Price & Reddy 77] genannte Ansatz, der durch Verwendung von Farb- statt Grauwertdaten eine bessere Trennung von benachbarten Bildobjekten in der Segmentierung ermöglicht. Die Anwendung in [Price & Reddy 77] ist das Erkennen von Unterschieden in Paaren ähnlicher Bilder und nicht die Bewegungsbestimmung in Bildfolgen. Deshalb werden wichtige Themen der Bildfolgenanalyse, wie z.B. Echtzeitfähigkeit des Verfahrens, zeitliche Konsistenz der Segmentierungsergebnisse und Prädiktionsverfahren, nicht behandelt.

Tabelle 2.2: Vor- und Nachteile von Verfahren zur Bewegungsbestimmung

Verfahren	Vorteile	Nachteile
Differenziell		
Global	+ flächendeckende Bewegungsschätzung	- dichte Bildfolge nötig - zeitaufwendig - Glättung von Bewegungskanten
Lokal	+ präzise Bewegungsschätzung	- dichte Bildfolge nötig - keine Bewegungsschätzung in schwach strukturierten Bild- bereichen
Frequenz- basiert		
	+ flächendeckende Bewegungsschätzung	- zeitaufwendig - Glättung von Bewegungskanten - nur translatorische Bewegung
Statische Korrespondenz- verfahren		
	+ flächendeckende Bewegungsschätzung	- ungenaue Bewegungsschätzung in schwach strukturierten Bild- bereichen
Adaptive Korrespondenz- verfahren		
Punkte oder Linien	+ präzise Bewegungsschätzung + große Verschiebungen erlaubt	- keine Bewegungsschätzung in schwach strukturierten Bild- bereichen
Regionen	+ geringer Rechenaufwand + große Verschiebungen erlaubt + flächendeckende Bewegungsschätzung	

2.3 Bewegungssegmentierung

In Kapitel 2.2 wurde ausführlich auf die Bestimmung der Bildbewegung eingegangen. Im folgenden werden Verfahren zur Auswertung der Bildbewegung vorgestellt; ihre Aufgabe ist eine objektbezogene Segmentierung der Bilder anhand der Bildbewegung. Prinzipiell erfolgt in der Bewegungssegmentierung eine Aufteilung des Bildes in Bereiche, in denen die Bildbewegung einen kontinuierlichen Verlauf aufweist. Entscheidend ist, daß eine solche Aufteilung ausschließlich durch die Geometrie und die Bewegung der Szene beeinflusst wird. Damit lassen sich im Unterschied zur intensitätsbasierten Einzelbildsegmentierung Objektflächen, die in Bereiche unterschiedlicher Intensität zerfallen, als Einheit detektieren.

Die Tatsache, daß der optische Fluß Geometrie- und Bewegungseigenschaften der Szenenobjekte widerspiegelt, hat zu einer Gruppe von Verfahren geführt, die aus dem optischen Fluß die 3D-Struktur der Szene berechnen [Broida & Chellappa 91, Tomasi & Kanade 92, Aggarwal & Nandhakumar 88, Adiv 85]. Aufgrund von Ungenauigkeiten in der Schätzung der 3D-Daten haben sich diese Verfahren bei Anwendungen, in denen die Eigenbewegung der Kamera unbekannt ist, bisher nicht durchgesetzt. Die im folgenden vorgestellten Verfahren arbeiten deshalb ausschließlich in der 2D-Bildebene. Trotzdem ist auch hier der Zusammenhang zwischen Geometrie und Bewegung der Szene und dem entsprechenden optischen Fluß in der 2D-Bildebene essentiell. Dieser Zusammenhang wird in den zwei nächsten Abschnitten beleuchtet.

2.3.1 Optischer Fluß als Folge der Bewegung starrer Objekte

Ein kartesisches Koordinatensystem, dessen Ursprung im Brennpunkt der Kamera fixiert ist, sei durch (X, Y, Z) gegeben. Das kartesische Koordinatensystem der planaren Bildebene sei (x, y) (siehe Bild 2.6). Die Brennweite der Kamera kann ohne Verlust der Allgemeingültigkeit auf 1 normiert werden. Der Zusammenhang zwischen den Koordinaten eines Bildpunktes (x, y) mit den Koordinaten des zugehörigen Szenenpunktes (X, Y, Z) lautet bei zentralperspektivischer Projektion:

$$\begin{aligned} x &= X/Z \\ y &= Y/Z \end{aligned} \tag{2.13}$$

Die Relativbewegung zwischen einem starren Objekt und der Kamera kann durch 6 Parameter beschrieben werden: 3 Komponenten der Translationsgeschwindigkeit (V_X, V_Y, V_Z) und 3 Komponenten der Rotationsgeschwindigkeit ($\Omega_X, \Omega_Y, \Omega_Z$) des Objektes in bezug auf das ruhende Kamerakoordinatensystem. Der optische Flußvektor eines Objektpunktes

(X, Y, Z) , der auf den Bildpunkt (x, y) abgebildet wird, berechnet sich folgendermaßen [Longuet-Higgins & Prazdny 80]:

$$\begin{aligned} u &= -\Omega_X xy + \Omega_Y(1 + x^2) - \Omega_Z y + (V_X - V_Z x)/Z \\ v &= -\Omega_X(1 + y^2) + \Omega_Y xy + \Omega_Z x + (V_Y - V_Z y)/Z \end{aligned} \quad (2.14)$$

Aus Gl. (2.14) werden drei wesentliche Eigenschaften des optischen Flusses deutlich:

- Der durch die Rotationsbewegung verursachte Anteil des optischen Flußvektors an einem Bildpunkt (x, y) ist unabhängig von der Entfernung Z des zugehörigen Szenenpunktes. Damit beinhaltet der Rotationsanteil keine 3D-Information über die aufgenommene Szene.
- Das durch die Translationsbewegung hervorgerufene Vektorfeld verläuft radial aus dem Fluchtpunkt heraus. Die Koordinaten des Fluchtpunktes sind:

$$\begin{aligned} x_F &= V_X/V_Z \\ y_F &= V_Y/V_Z \end{aligned} \quad (2.15)$$

In [Thompson & Pong 90] wird von dieser Eigenschaft Gebrauch gemacht. Bei translatorischer Bewegung der Kamera erfolgt die Detektion bewegter Objekte anhand von Flußvektoren, die nicht in Richtung des durch Kamerabewegung erzeugten Fluchtpunktes laufen.

- Selbst wenn die Relativbewegungen aller Szenenpunkte in bezug auf die Kamera identisch sind, können durch die Abhängigkeit der Translationskomponenten des optischen Flusses von Z Diskontinuitäten im optischen Fluß entstehen. Sie treten an sog. Tiefengrenzen auf, wo Szenenpunkte, die auf benachbarte Bildpunkte abgebildet werden, einen Sprung in der Z -Komponente aufweisen.

Analog zur Berechnung des optischen Flusses läßt sich ein Gleichungssystem für die Disparitäten $(\Delta x, \Delta y)$ angeben. Unter der Annahme, daß im Zeitintervall $[n, n + 1]$ nur kleine Rotationswinkel (Φ_X, Φ_Y, Φ_Z) auftreten, lassen sich die Disparitäten wie folgt berechnen [Adiv 85]:

$$\begin{aligned} \Delta x &= \frac{-\Phi_X xy + \Phi_Y(1 + x^2) - \Phi_Z y + (S_X - S_Z x)/Z}{1 + \Phi_X y - \Phi_Y x + S_Z/Z} \\ \Delta y &= \frac{-\Phi_X(1 + y^2) + \Phi_Y xy + \Phi_Z x + (S_Y - S_Z y)/Z}{1 + \Phi_X y - \Phi_Y x + S_Z/Z} \end{aligned} \quad (2.16)$$

Hierbei gibt (S_X, S_Y, S_Z) die Verschiebung im Zeitraum $[n, n + 1]$ an. Falls $|S_Z/Z| \ll 1$ und das Sichtfeld der Kamera relativ klein ist, ergibt sich für die Verschiebung:

$$\begin{aligned} \Delta x &= -\Phi_X xy + \Phi_Y(1 + x^2) - \Phi_Z y + (S_X - S_Z x)/Z \\ \Delta y &= -\Phi_X(1 + y^2) + \Phi_Y xy + \Phi_Z x + (S_Y - S_Z y)/Z \end{aligned} \quad (2.17)$$

womit die Analogie zu Gl. (2.14) hergestellt ist.

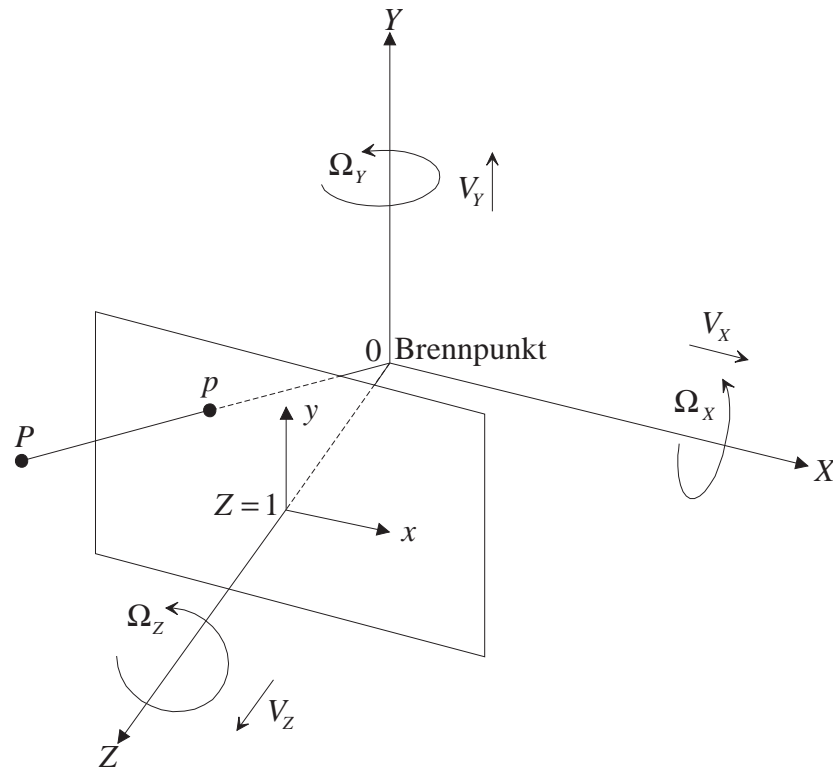


Bild 2.6: (Nach [Longuet-Higgins & Prazdny 80]) Die 3D-Koordinaten beziehen sich auf das mit der Kamera fest verbundene Koordinatensystem (X, Y, Z) . Das Koordinatensystem in der Bildebene ist durch (x, y) gegeben. Durch zentralperspektivische Projektion wird der Szenenpunkt P auf den Bildpunkt p abgebildet. (V_x, V_y, V_z) und $(\Omega_x, \Omega_y, \Omega_z)$ repräsentieren die Relativbewegung zwischen Objekt und Kamera.

2.3.2 Modellierung des optischen Flusses für starre, planare Flächen

Eine planare Fläche in der Szene läßt sich durch die Gleichung

$$k_X X + k_Y Y + k_Z Z = 1 \quad (2.18)$$

beschreiben. Wobei die Koeffizienten k_X , k_Y und k_Z beliebige reelle Werte, mit Ausnahme des Falles $k_X = k_Y = k_Z = 0$, annehmen können. Die Gleichung der in die Bildebene projizierten Fläche lautet:

$$k_X x + k_Y y + k_Z = 1/Z \quad (2.19)$$

Setzt man Gl. (2.19) in Gl. (2.14) ein, erhält man den optischen Fluß für eine starre, planare Fläche [Adiv 85]:

$$\begin{aligned} u &= a_1 + a_2x + a_3y + a_7x^2 + a_8xy \\ v &= a_4 + a_5x + a_6y + a_7xy + a_8y^2 \end{aligned} \quad (2.20)$$

mit

$$\begin{aligned} a_1 &= \Omega_Y + k_Z V_X \\ a_2 &= k_X V_X - k_Z V_Z \\ a_3 &= -\Omega_Z + k_Y V_X \\ a_4 &= -\Omega_X + k_Z V_Y \\ a_5 &= \Omega_Z + k_X V_Y \\ a_6 &= k_Y V_Y - k_Z V_Z \\ a_7 &= \Omega_Y - k_X V_Z \\ a_8 &= -\Omega_X - k_Y V_Z \end{aligned} \quad (2.21)$$

Unter der Annahme, daß das Sichtfeld der Kamera relativ klein ist oder das Objekt sehr weit entfernt, gilt: $x \ll 1$ und $y \ll 1$. Damit kann der Verlauf des optischen Flusses einer starren, planaren Fläche näherungsweise durch eine affine Transformation beschrieben werden:

$$\begin{aligned} u &= a_1 + a_2x + a_3y \\ v &= a_4 + a_5x + a_6y \end{aligned} \quad (2.22)$$

2.3.3 Bewegungssegmentierung anhand affiner Bewegungsmodelle

Gl. (2.22) bildet die Grundlage für eine Gruppe von Verfahren, die eine Aufteilung des Bildes in Regionen vornehmen, in denen der Verlauf des aus der Bildfolge geschätzten optischen Flusses durch eine affine Transformation hinreichend genau modelliert werden kann. In [Adiv 85] werden durch eine Hough-Transformation optische Flußvektoren zusammengefaßt, die derselben affinen Transformation genügen. Der multidimensionale, diskrete Hough-Raum wird dabei durch die Parameter der affinen Transformation aufgespannt. Mit jedem Punkt des Hough-Raums ist genau eine affine Transformation verknüpft. Die aus der Bildfolge geschätzten optischen Flußvektoren bewerten nun jeden Punkt im Hough-Raum hinsichtlich der Konsistenz der zugehörigen Transformation mit dem jeweiligen Vektor. Die Punkte, die eine hohe Bewertung erhalten, weisen auf relevante Segmente im optischen Fluß hin. Die Globalität der Hough-Transformation und die 6 Dimensionen des Parameterraums führen zu einem erheblichen Rechenaufwand. In [Bouthemy & François 93] wird ein auf Markov-Random-Fields basierendes statistisches Modell zur Segmentierung verwendet. In die zu minimierende globale Energiefunktion gehen die Abweichungen der optischen Flußvektoren vom affinen Modell ein. Ein zweiter Term erhöht die Energie im Falle der Zuordnung benachbarter Bildpunkte zu unterschiedlichen Segmenten. Neben

der Initialisierung der Segmentierung ist auch hier der hohe Rechenaufwand problematisch.

Eine zweite Gruppe von Verfahren verwendet affine Transformationen zur Modellierung des Verlaufs der Disparitäten innerhalb von Abbildungen starrer, planarer Flächen. Die Koordinaten eines Bildpunktes zum Zeitpunkt $n + 1$ lassen sich aus den Koordinaten des entsprechenden Bildpunktes zum Zeitpunkt n durch eine affine Transformation berechnen:

$$\begin{aligned}x(n + 1) &= b_1 + b_2x(n) + b_3y(n) \\y(n + 1) &= b_4 + b_5x(n) + b_6y(n)\end{aligned}\tag{2.23}$$

Wobei die Parameter der affinen Transformation durch den Verlauf der Ebene sowie die Translation und Rotation der Ebene im Zeitraum $[n, n + 1]$ gegeben sind. Die Konstanz der Intensität entlang einer Trajektorie angenommen, gilt für die Intensitätsfunktionen zweier aufeinanderfolgender Bilder:

$$f(b_1 + b_2x + b_3y, b_4 + b_5x + b_6y, n + 1) = f(x, y, n)\tag{2.24}$$

Damit ist die Basis für Verfahren geschaffen, die aus der Differenz der beiden Seiten von Gl. (2.24) auf die Güte einer affinen Transformation zur Approximation einer gegebenen Bildänderung schließen. Durch Maximierung der Gütefunktion in Abhängigkeit der affinen Parameter kann dann die optimale affine Transformation bestimmt werden. Im Unterschied zu den oben genannten Verfahren ist somit keine Berechnung des optischen Flusses erforderlich. Meist wird im ersten Schritt dieser Verfahren eine sog. dominante affine Transformation für die Verschiebungen im gesamten Bildbereich bestimmt. Die Parameter dieser Transformation berücksichtigen beispielsweise das durch rotatorische Eigenbewegung der Kamera hervorgerufene, von der Szenengeometrie unabhängige, Verschiebungsvektorfeld. In einem zweiten Schritt werden dann Gebiete selektiert, in denen die dominante Transformation eine schlechte Approximation der Verschiebungen liefert. Für diese Gebiete wird erneut eine affine Transformation bestimmt. So kann das Bild sukzessive in Bereiche unterteilt werden, denen eine affine Transformation zugeordnet wird. In [Irani & Anandan 96] wird das Bild auf diese Weise in mehrere große Bereiche unterteilt. Als bewegte Bildobjekte werden die verbleibenden Bildregionen interpretiert, in denen die Verschiebungen durch die berechneten affinen Transformationen nur schlecht approximiert werden. In [Ayer et al. 94] wird die Aufteilung des Bildes durch eine Grauwertsegmentierung gestützt. Das in [Odobez & Bouthemy 94] vorgestellte Verfahren begnügt sich mit der Berechnung einer affinen Transformation für das gesamte Bild. Die anschließende Bewegungssegmentierung erfolgt mit Hilfe von Markov-Random-Fields. In Anlehnung an das bereits erwähnte Verfahren aus [Adiv 85] erfolgt in [Bober & Kittler 94] die Berechnung der Parameter der affinen Transformation im Hough-Raum.

2.3.4 Modellfreie Bewegungssegmentierung

Die bisher beschriebenen Verfahren zur Bewegungssegmentierung gehen davon aus, daß sich die Szene durch wenige starre, planare Flächen modellieren läßt. Tatsächlich ist diese Annahme in Szenen, die weitgehend aus künstlichen Objekten bestehen, meist zulässig. Probleme ergeben sich jedoch, wenn natürliche Objekte auftreten, die beispielsweise ihre Form ändern oder deren Oberflächen sich schlecht durch planare Flächenstücke beschreiben lassen. In diesem Fall sind Verfahren erforderlich, die von Modellen absehen und lediglich eine Ähnlichkeit der Bewegung von Bildpunkten voraussetzen, die zu einem Objekt gehören. In [Smith & Brady 94] wird aufgrund der Vektordifferenz zweier im Bild benachbarter optischer Flußvektoren entschieden, ob die entsprechenden Bildpunkte zum selben Bildobjekt gehören. Durch eine auf optische Flußvektoren angewandte Clusteranalyse erfolgt in [Bennamoun et al. 94] die Detektion bewegter Objekte in einfachen Szenen. Eine interessante Idee der Kombination von Bewegungs- und Farbsegmentierung wird in [Etoh & Shirai 93] vorgestellt. Jedem Bildpunkt wird ein 8-dimensionaler Merkmalsvektor zugeordnet, bestehend aus den drei RGB-Farbwerten, der Position des Punktes im Bild und den Orts-Zeit ableitungen der Intensität. Anschließend erfolgt die Segmentierung des Bildes durch eine Clusteranalyse der Bildpunkte im 8-dimensionalen Merkmalsraum.

2.3.5 Bewertung der Verfahren

In Tabelle 2.3 sind die Vor- und Nachteile der modellbasierten und modellfreien Bewegungssegmentierung dargestellt. Da in Verkehrsszenen häufig natürliche Objekte auftreten und die Rotationsbewegungen um die optische Achse der Kamera relativ gering sind, ist ein modellfreier Ansatz in der vorliegenden Anwendung vorzuziehen. Ein Schwachpunkt aller in diesem Kapitel genannten Verfahren ist, daß die Segmentierung nur auf der momentanen Bildbewegung beruht. Information über die zeitliche Änderung der Bildbewegung, die bei einer Bewegungsbestimmung durch adaptive Korrespondenzverfahren in Form von Trajektorien vorliegt, bleibt ungenutzt. Eine Verbesserung der Segmentierungsergebnisse könnte durch Auswertung von Trajektorien anstelle von Verschiebungsvektoren erzielt werden.

Tabelle 2.3: Vor- und Nachteile von Verfahren zur Bewegungssegmentierung

Verfahren	Vorteile	Nachteile
affines Bewegungsmodell	+ Aussagen über Szenengeometrie und -bewegung möglich	- zeitaufwendig - nur künstliche Objekte
modellfrei	+ natürliche Objekte	- Probleme bei starker Rotation um optische Achse

3 Farbregionenfluß

Der Farbregionenfluß besteht aus drei wesentlichen Bestandteilen: Der Farbsegmentierung, der Bewegungsbestimmung und der Objektdetektion. Entsprechend gliedert sich das Kapitel. Im ersten Unterkapitel wird die Farbsegmentierung beschrieben. Ihre Aufgabe ist die Aufteilung der einzelnen Farbbilder in Regionen homogener Farben. Im Anschluß daran wird die Bewegungsbestimmung von Farbregionen in Bildfolgen erläutert. Im nächsten Unterkapitel wird das Zusammenfassen von benachbarten Regionen mit ähnlichem Bewegungsverhalten zu Objekthypothesen dargestellt. Am Schluß des Kapitels werden Ergebnisse der Objektdetektion präsentiert.

3.1 Farbsegmentierung

3.1.1 Farbe

Farbe ist eine Sinnesempfindung des Menschen, die durch Auftreffen von Licht im Spektralbereich von etwa 380 nm – 760 nm auf die Netzhaut hervorgerufen wird. Auf der Netzhaut befinden sich drei verschiedene Typen von Farbzeptoren (Zapfen). Jeder Typ weist eine wellenlängenabhängige Empfindlichkeitskurve auf, aus der sich die Erregungsstärke des Zapfentyps für beliebige Spektralverteilungen des einfallenden Lichtes bestimmen läßt. Die Empfindlichkeitskurven sind in Bild 3.1 dargestellt. Ihre Maxima liegen in Wellenlängenbereichen des roten, grünen und blauen Lichtes. Die Erregungsstärken P , D , und T der drei Zapfentypen lauten bei einem gegebenen Farbreiz φ_λ [Richter 81]:

$$\begin{aligned} P &= k \int_{\lambda} \varphi_{\lambda} \bar{p}(\lambda) d\lambda \\ D &= k \int_{\lambda} \varphi_{\lambda} \bar{d}(\lambda) d\lambda \\ T &= k \int_{\lambda} \varphi_{\lambda} \bar{t}(\lambda) d\lambda \end{aligned} \tag{3.1}$$

dabei sind $\bar{p}(\lambda)$, $\bar{d}(\lambda)$ und $\bar{t}(\lambda)$ die Empfindlichkeitskurven; k ist eine Konstante.

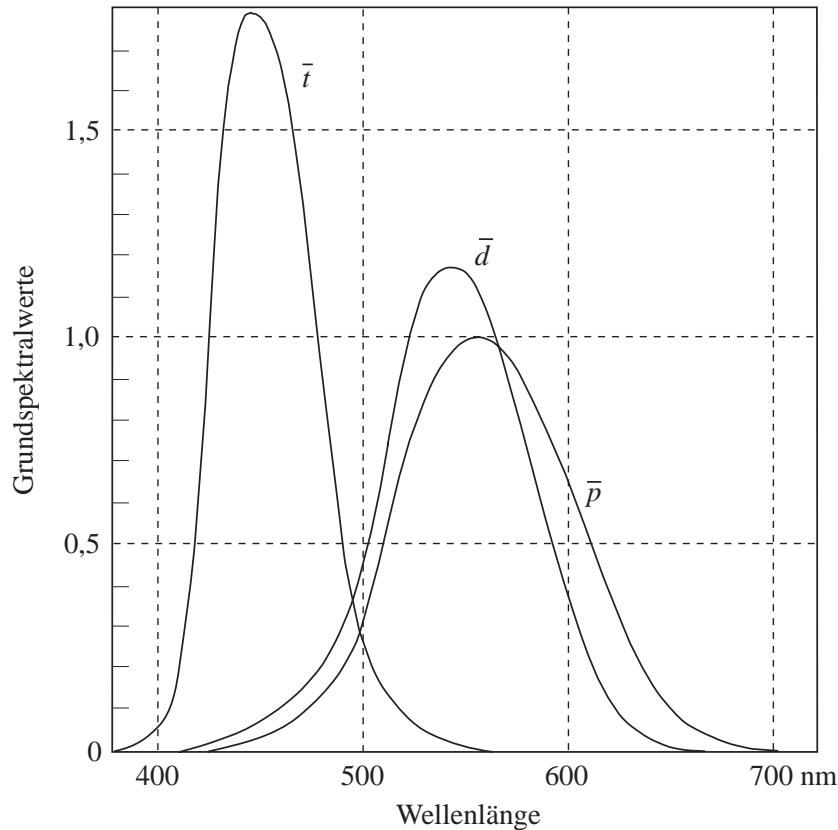


Bild 3.1: Die Empfindlichkeitskurven der drei Zapfenarten nach [Richter 81]

Die an das Gehirn weitergeleitete Information besteht nur noch aus den drei Anteilen, die von den verschiedenen Zapfentypen herrühren. Die Kodierung der Spektralverteilung in drei Komponenten ist der Grund, weshalb jede durch den Menschen wahrnehmbare Farbe als additive Mischung genau dreier Primärfarben erzeugt werden kann. In der Bildverarbeitung nutzt man diese Tatsache, um Farben in einem 3-dimensionalen Koordinatensystem darzustellen. Am gebräuchlichsten ist der RGB-Raum (siehe Bild 3.2). Der Ursprung dieses Koordinatensystems stellt die Farbe Schwarz dar. Die Primärfarben Rot, Grün und Blau liegen auf der R - G - bzw. B -Achse. Die durch den Ursprung und den Weißpunkt verlaufende Raumdiagonale wird als Unbuntachse bezeichnet. Da man sich auf Farben beschränkt, die aus der additiven Mischung positiver Anteile von Primärfarben entstehen, nutzt man vom Farbraum lediglich den durch die 3 Primärfarben aufgespannten Quader. In technischen Systemen werden die Vektorkomponenten meist als 8 Bit Werte mit einem ganzzahligen Wertebereich von 0 bis 255 dargestellt. Alle folgenden Zahlenangaben im Zusammenhang mit Farbvektoren beziehen sich deshalb auf den Wertebereich $[0, 255]$.

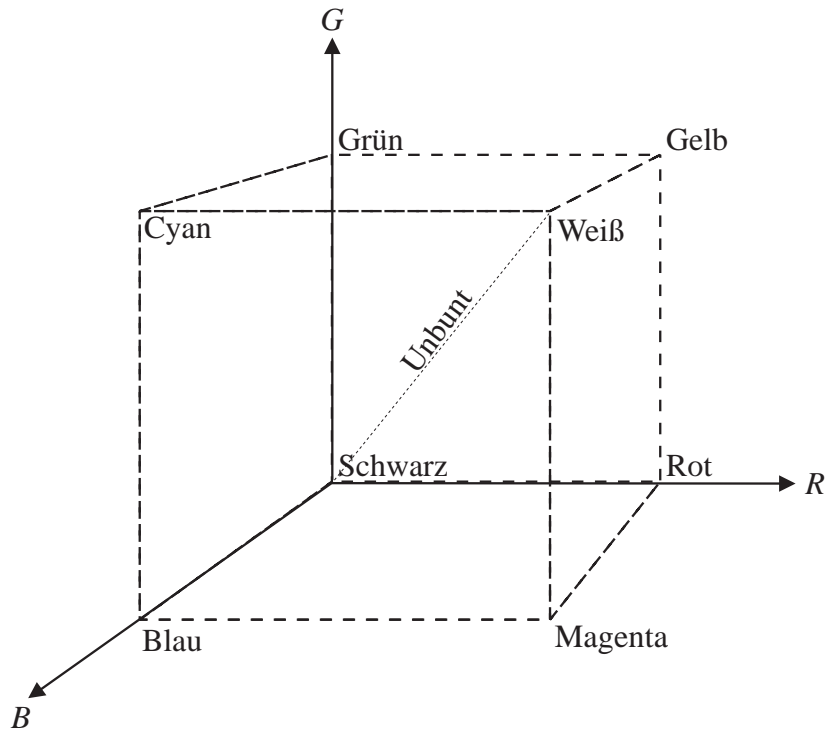


Bild 3.2: Der RGB-Farbraum

3.1.2 Aufgabe der Farbsegmentierung

Die Aufgabe der Farbsegmentierung ist die Aufteilung des Bildes in farblich homogene Regionen, die sich nicht überlappen. Die Vielfalt der in der Literatur vorgeschlagenen Farbsegmentierungen [Pal & Pal 93, Skarbek & Koschan 94] hat ihre Ursache im Fehlen eines allgemeingültigen Maßes für die Qualität der Segmentierungsergebnisse. Nur im Zusammenhang mit einer Anwendung kann eine Beurteilung erfolgen. Für die vorliegende Anwendung sind zwei wichtige Forderungen an die Farbsegmentierung zu stellen:

- Um Regionen in einer Bildsequenz verfolgen zu können, müssen die Segmentierungsergebnisse für Folgebilder konsistent sein.
- Der Rechenaufwand der Segmentierung soll eine Demonstrator-Realisierung der Objektdetektion auf einer Workstation zulassen. So scheidet Segmentierungen, die mehrere Minuten pro Bild benötigen, von vornherein aus.

Die Farbquantisierung mittels Clusterverfahren erfüllt diese Forderungen. Sie reduziert die im Bild auftretenden Farben auf eine kleine Anzahl signifikanter Farben. Durch die Farbreduktion erreicht man, daß große Regionen mit Bildpunkten gleicher Farbe entstehen. Die Eingangsdaten der Farbquantisierung sind die Farbvektoren aller Bildpunkte.

Als Ausgang liefert das Verfahren einen Satz von Referenzfarben, mit denen man die Originalfarben im Bild ersetzt und somit das segmentierte Bild erhält. Eine Konsistenz der Segmentierungsergebnisse in Bildfolgen kann durch Verwendung derselben Referenzfarben zur Quantisierung mehrerer aufeinanderfolgender Bilder erzielt werden.

3.1.3 Farbquantisierung durch Clusteranalyse

Da Clusterverfahren bei unterschiedlichsten Problemen der multivariaten Datenanalyse eingesetzt werden, soll im folgenden eine allgemeine Beschreibung ihrer Aufgabenstellung erfolgen. Anschließend werden einige Clusterverfahren zur Quantisierung von Farbbildern untersucht.

Das Ziel der Clusteranalyse ist eine unüberwachte Aufteilung einer Menge von Daten, so daß ein gegebenes Fehlermaß minimiert wird. Als Fehlermaß wird meist der Mittelwert der quadratischen euklidischen Abstände (Mean Squared Error) der Datenpunkte zu den zugehörigen Referenzvektoren verwendet. Der Referenzvektor ist der Mittelwert der Datenpunkte des jeweiligen Clusters. Bei Quantisierungsaufgaben werden die Originaldaten nach der Partitionierung durch die zugehörigen Referenzvektoren ersetzt; der *MSE* wird deshalb auch als Quantisierungsfehler bezeichnet.

Die formale Aufgabenbeschreibung der Clusteranalyse bei Verwendung des *MSE* als Fehlermaß lautet wie folgt: Bestimme für eine gegebene Datenmenge $\mathcal{S} = \{\mathbf{s}_1, \dots, \mathbf{s}_N\}$ mit $\mathcal{S} \subseteq \mathbb{R}^D$ einen Satz von Q ($N \gg Q > 1$) Referenzvektoren $\mathcal{R} = \{\mathbf{r}_1, \dots, \mathbf{r}_Q\}$ mit $\mathcal{R} \subseteq \mathbb{R}^D$, so daß der mittlere quadratische Abstand:

$$MSE = \frac{1}{N} \sum_{i=1}^N \|\mathbf{s}_i - \mathbf{r}_{q(\mathbf{s}_i)}\|^2 \quad (3.2)$$

minimiert wird. Dabei ist $\mathbf{r}_{q(\mathbf{s}_i)}$ der Referenzvektor des Clusters, in dem der Datenvektor \mathbf{s}_i liegt. Offensichtlich ist eine notwendige Voraussetzung für ein lokales Minimum des *MSE*, daß $\mathbf{r}_{q(\mathbf{s}_i)}$ der dem Datenvektor \mathbf{s}_i am nächsten liegende Referenzvektor ist:

$$\mathbf{r}_{q(\mathbf{s}_i)} = \arg \min_{1 \leq n \leq Q} \|\mathbf{s}_i - \mathbf{r}_n\|^2 \quad (3.3)$$

Durch die Festlegung der Referenzvektoren ist damit auch die Aufteilung der Daten gegeben. Zum Cluster \mathcal{C}_q gehören alle Datenvektoren, die am nächsten beim Referenzvektor \mathbf{r}_q liegen:

$$\mathcal{C}_q = \{\mathbf{s}_n \mid \|\mathbf{s}_n - \mathbf{r}_q\|^2 \leq \|\mathbf{s}_n - \mathbf{r}_i\|^2 \forall i = 1, \dots, Q\} \quad (3.4)$$

Umgekehrt ist durch die Cluster der Referenzvektor als Mittelwert aller Daten des Clusters eindeutig gegeben:

$$\mathbf{r}_q = \frac{1}{S_q} \sum_{\mathbf{s}_i \in \mathcal{C}_q} \mathbf{s}_i \quad (3.5)$$

Wobei S_q die Anzahl der Daten im Cluster \mathcal{C}_q ist. Da die Minimierung des Quantisierungsfehlers ein NP-vollständiges Problem ist, werden in der Praxis meist suboptimale Techniken zur Clusteranalyse eingesetzt. Sie versuchen entweder ein lokales Minimum des Quantisierungsfehlers zu finden oder begnügen sich sogar mit Näherungslösungen, die nicht zwangsläufig lokale Minima sind. Im folgenden sollen einige für die Arbeit relevante Clustertechniken vorgestellt werden.

Iterative Verfahren

Iterative Verfahren sind in der Lage, nach einer endlichen Anzahl von Iterationsschritten ein lokales Minimum des Quantisierungsfehlers zu finden. Am bekanntesten ist das k-means Verfahren, das in seiner ursprünglichen Form [MacQueen 67] zur Gruppierung zeitlich aufeinanderfolgender Daten konzipiert wurde. Da die Daten der Reihe nach zur Adaption der Referenzvektoren herangezogen werden, bezeichnet man das Verfahren als sequentielles k-means Verfahren. Beim parallelen k-means Verfahren [Duda & Hart 73] werden die Referenzvektoren in jeder Iteration anhand aller Daten neu berechnet. Die Reihenfolge der Daten spielt keine Rolle. Für Anwendungen, in denen die Daten bereits zu Beginn der Clusteranalyse vorliegen, ist es geeigneter als das sequentielle Verfahren.

Paralleles k-means Verfahren Beim parallelen k-means Verfahren erfolgt zuerst eine Initialisierung der Referenzvektoren, beispielsweise durch eine Auswahl von Q Datenvektoren. Die nachfolgenden Iterationen gliedern sich jeweils in zwei Schritte. Im ersten Schritt wird jeder Datenvektor dem am nächsten liegenden Referenzvektor zugeordnet. Im zweiten Schritt werden dann die Referenzvektoren als Mittelwerte der ihnen zugeordneten Datenvektoren neu berechnet. Ein lokales Minimum des Quantisierungsfehlers ist dann erreicht, wenn sich die Referenzvektoren in zwei aufeinanderfolgenden Iterationen nicht geändert haben. Beim k-means Verfahren ist die Initialisierung der Referenzvektoren problematisch. Eine ungünstige Initialisierung kann zum einen viele Iterationen zum Erreichen eines lokalen Minimums zur Folge haben, zum anderen kann sie in ein lokales Minimum führen, bei dem der Quantisierungsfehler immer noch relativ groß ist.

Divisive Verfahren

Statt der lokalen Minimierung des Quantisierungsfehlers versuchen divisive Verfahren den Rechenaufwand möglichst gering zu halten, gleichzeitig aber eine akzeptable Näherungslösung der Minimierungsaufgabe zu bestimmen. Zu Beginn der Verfahren werden alle Daten in einem Cluster zusammengefaßt. In den folgenden Schritten wird jeweils eines der Cluster ausgewählt und in zwei neue Cluster aufgeteilt. Dies geschieht so oft, bis

die vorgegebene Clusterzahl erreicht ist. Je nach Verfahren ist die Wahl des im nächsten Schritt zu teilenden Clusters und dessen Aufteilung unterschiedlich.

Median-cut Verfahren In [Heckbert 80] wird das median-cut Verfahren zur Quantisierung von RGB-Farbbildern vorgeschlagen. Beginnend bei einem Cluster, das die gesamte Datenmenge umspannt, wird der Datenraum schrittweise in quaderförmige Cluster unterteilt. Für jedes Cluster wird die Anzahl der in ihm enthaltenen Datenpunkte berechnet, die unterschiedliche Farbwerte besitzen. Anschließend wird das Cluster mit der größten Anzahl von verschiedenen Farbwerten¹ aufgeteilt. Dazu wird die Farbachse bestimmt, entlang derer die Daten des Clusters die größte Spreizung (Differenz zwischen Maximum und Minimum) aufweisen. Das Cluster wird dann durch eine Ebene geteilt, die senkrecht zur vorher bestimmten Farbachse läuft und gleichzeitig den Median der auf diese Achse projizierten Clusterdaten schneidet. Damit enthalten die zwei neu gebildeten Cluster gleichviele Datenpunkte. Ist die vorgegebene Clusterzahl erreicht, werden die Referenzfarben als Mittelwerte der in den Partitionen befindlichen Datenpunkte berechnet. Die Wahl des Medians als Schnittpunkt zwischen Farbachse und Trennebene bezweckt in erster Linie eine Aufteilung des Datenraums in Cluster, die etwa gleich viele Daten enthalten.

Wu Verfahren Das zur Quantisierung von RGB-Farbbildern entwickelte Wu Verfahren [Wu 91] erzeugt wie das median-cut Verfahren quaderförmige Cluster, deren Trennebenen senkrecht zu den Achsen verlaufen. Es wird jeweils das Cluster mit der höchsten Streuung aufgeteilt. Die Trennebene wird so gewählt, daß die Summe der Streuungen in den zwei neu gebildeten Partitionen minimiert wird. Um den Rechenaufwand niedrig zu halten, erfolgt eine Reduzierung der Auflösung der Farbachsen von ursprünglich 8 Bit auf 5 Bit.

Eigenvektor Verfahren Das im Daimler-Benz Forschungszentrum in Ulm entstandene Eigenvektor Verfahren teilt wie das Wu Verfahren jeweils das Cluster mit der größten Streuung. Über die Eigenvektoren der Kovarianzmatrix eines Clusters wird die Richtung der größten Streuung innerhalb des Clusters bestimmt. Die Aufteilung erfolgt durch eine Ebene, die senkrecht zu dieser Richtung verläuft und das Cluster im Zentrum schneidet.

¹Ursprünglich wurde in [Heckbert 80] das Cluster mit der größten Ausdehnung in bezug auf eine der Achsen aufgeteilt. In der hier verwendeten Implementierung des median-cut Verfahrens aus dem Jahre 1989 von J. Poskanzer wurde jedoch die Anzahl der verschiedenen Farben als Auswahlkriterium verwendet.

Tabelle 3.1: Mittlerer Quantisierungsfehler und mittlere CPU-Zeit pro Bild auf einer Sun SPARC-20 bei 10 RGB-Farbbildern $(360 \times 288)^2$. Es wurden 16 Cluster pro Bild berechnet. Die Varianz der Farben³ in den 10 Bildern lag im Mittel bei $\sigma^2 = 8361$. Die Bilder stammen aus 10 verschiedenen Autobahnsequenzen; auf Seite 42 in Bild 3.5 ist ein Beispiel zu sehen.

Verfahren	<i>MSE</i>	CPU-Zeit [ms]
Referenz	184	> 10000
Eigenvektor	224	1866
Wu	246	196
Median-cut	345	2176

Experimentelle Untersuchung von Clusterverfahren zur Segmentierung von Farbbildern

Zur Wahl einer geeigneten Farbquantisierung wurden die oben genannten divisiven Clusterverfahren an zehn, 360×288 großen, RGB-Farbbildern untersucht. Pro Bild wurden 16 Cluster (Referenzfarben) bestimmt. Als Referenz wurde ein iteratives Verfahren herangezogen, das Bestandteile des Eigenvektor Verfahrens mit denen des parallelen k-means Verfahrens kombiniert. Es garantiert die Auffindung eines lokalen Minimums des Quantisierungsfehlers bei gleichzeitiger Umgehung der Initialisierungsproblematik des parallelen k-means Verfahrens. Zur Beurteilung der Verfahren wurde der Quantisierungsfehler und die Rechenzeit gemessen. Die Resultate sind in Tabelle 3.1 zu sehen. Wie erwartet, weist das iterative Referenzverfahren den niedrigsten Quantisierungsfehler auf, scheidet aber aufgrund seines hohen Rechenaufwands für die vorliegende Anwendung aus. Da beim Eigenvektor Verfahren beliebige Richtungen der Trennebenen möglich sind, führt es im Vergleich zu den beiden anderen divisiven Verfahren zu einem geringeren Quantisierungsfehler. Ausgewählt wurde das Wu Verfahren, das trotz des geringen Rechenaufwands einen akzeptablen Quantisierungsfehler liefert.

In Bild 3.3 ist der Quantisierungsfehler und die Anzahl der Regionen im quantisierten Bild in Abhängigkeit der Clusterzahl exemplarisch für das Wu Verfahren aufgetragen. Ein kritischer Punkt der Farbquantisierung ist die Festlegung der Anzahl der Cluster. Fällt sie zu klein aus, entstehen nur wenige Regionen im segmentierten Bild. Als Folge lassen sich

²Die Größenangaben beziehen sich – auch in folgenden Textstellen – auf Bildpunkte.

³Die Varianz der Farben berechnet sich aus der Summe der Varianzen in den 3 Farbkanälen: $\sigma^2_R + \sigma^2_G + \sigma^2_B$. Die Farbwerte liegen im ganzzahligen Wertebereich von 0 bis 255.

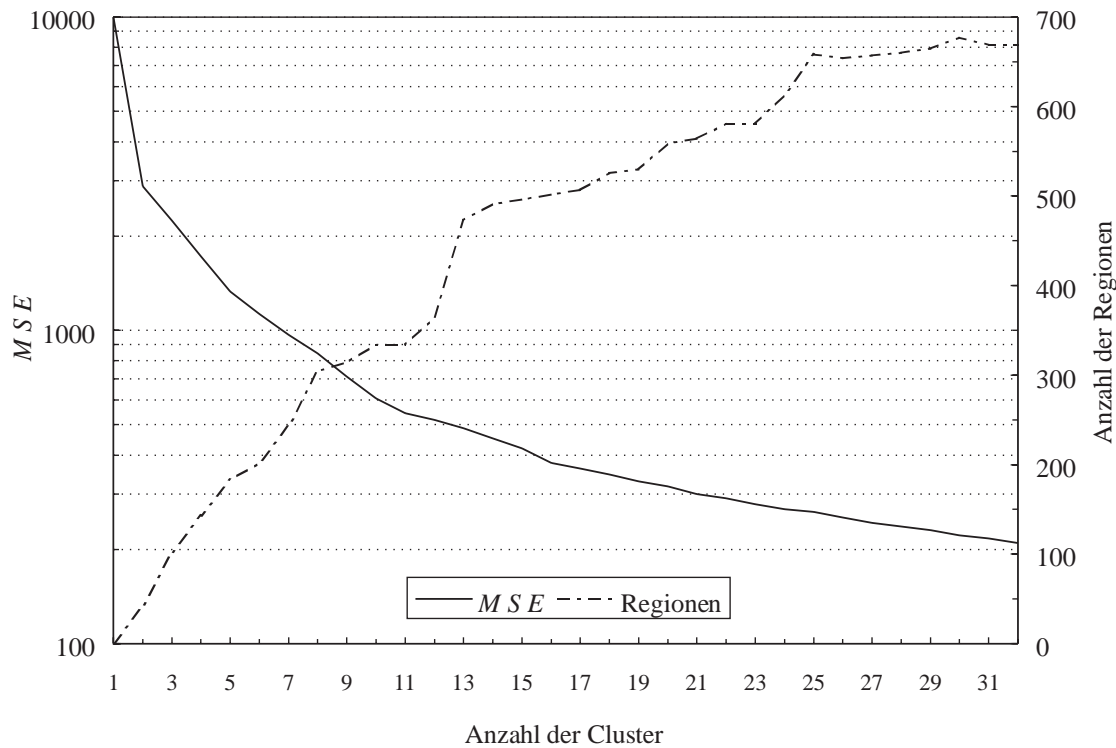


Bild 3.3: Quantisierungsfehler und Regionenzahl im quantisierten Bild in Abhängigkeit von der Anzahl der Cluster. Zur Quantisierung wurde das Wu Verfahren eingesetzt. Im quantisierten Bild wurden nur Farbregionen gezählt, die aus mindestens 10 Bildpunkten und maximal 50000 Bildpunkten bestanden. Die Varianz der Farben des untersuchten Bildes lag bei $\sigma^2 = 11959$.

die Bildobjekte nicht mehr trennen, man spricht von einer Untersegmentierung. Zu viele Cluster führen zu einer Übersegmentierung des Bildes, es entstehen viele kleine, instabile Regionen, die sich nur schlecht in einer Bildsequenz verfolgen lassen. Bei der Untersuchung von Bildfolgen mit Autobahn- und Landstraßenszenen hat sich eine Quantisierung auf 16 Farben bewährt. Sie führt im Durchschnitt zu etwa 400 Farbregionen und einem Quantisierungsfehler von etwa 250.

Nachbearbeitung der quantisierten Farbbilder

In farbquantisierten Bildern trifft man auf zerklüftete Regionenkonturen und eine Vielzahl kleiner Regionen, bedingt durch Störungen in den Bilddaten (z.B. Meßrauschen der Farbkamera). Zur Eliminierung dieser unerwünschten Effekte wurden zwei Verfahren, die

Medianfilterung und die Relaxation, zur Nachbearbeitung der quantisierten Bilder untersucht. Beide Verfahren führen eine Glättung des segmentierten Bildes durch, indem sie Bilddaten in der Umgebung des jeweils betrachteten Punktes nutzen, um diesem eine Farbe zuzuweisen. Von Bedeutung ist außerdem, daß durch die beiden Nachbearbeitungsverfahren die im Bild auftretenden Farben nicht geändert werden.

Medianfilter Beim Medianfilter wird eine $m \times m$ Maske punktweise über das Bild geschoben. Die Bilddaten innerhalb der Maske werden nach einem gegebenen Kriterium sortiert. In den Untersuchungen wurden die Farben entsprechend der Summe ihrer Vektorkomponenten geordnet. Im gefilterten Bild wird dann dem Bildpunkt im Zentrum der Maske die zum Median gehörende Farbe zugewiesen. Die Medianfilterung kann mehrfach hintereinander auf ein Bild angewandt werden. Als Eingangsdaten der aktuellen Iteration werden dabei die in der vorhergehenden Iteration erzeugten gefilterten Bilddaten verwendet. Die gestrichelte Kurve in Bild 3.4 zeigt die Anzahl der pro Iteration geänderten Farben. Angefangen bei über 10000 Änderungen in der ersten Iteration sinkt nach 3 Iterationen die Zahl der Änderungen bereits unter 600. Tabelle 3.2 zeigt, daß die Rechenzeit für 3 Iterationen mit einer 3×3 Maske in einem akzeptablen Rahmen liegt. Die Glättungswirkung des Medianfilters wird aus Bild 3.5 deutlich.

Tabelle 3.2: Mittlere CPU-Zeit pro Bild auf einer Sun SPARC-20. Es wurden zehn 360×288 große, auf 16 Farben quantisierte, Bilder verwendet. Pro Bild wurden drei Iterationen durchgeführt.

Nachbearbeitung	Maskengröße	Iterationen pro Bild	CPU-Zeit [ms]
Medianfilter	3×3	3	175
Relaxation	3×3	3	5754

Relaxation Im Unterschied zum Medianfilter basiert die Relaxation [Zucker et al. 78] auf einem probabilistischen Modell. In der Initialisierung des Verfahrens wird für jeden Bildpunkt ein Q -dimensionaler Wahrscheinlichkeitsvektor berechnet. Er gibt an, mit welcher Wahrscheinlichkeit einem Punkt n eine der Referenzfarben zuzuordnen ist. Die Komponenten des Vektors berechnen sich folgendermaßen:

$$p_{n,q}^{(0)} = \frac{K R_q}{1 + \|\mathbf{s}_n - \mathbf{r}_q\|^2} \quad (3.6)$$

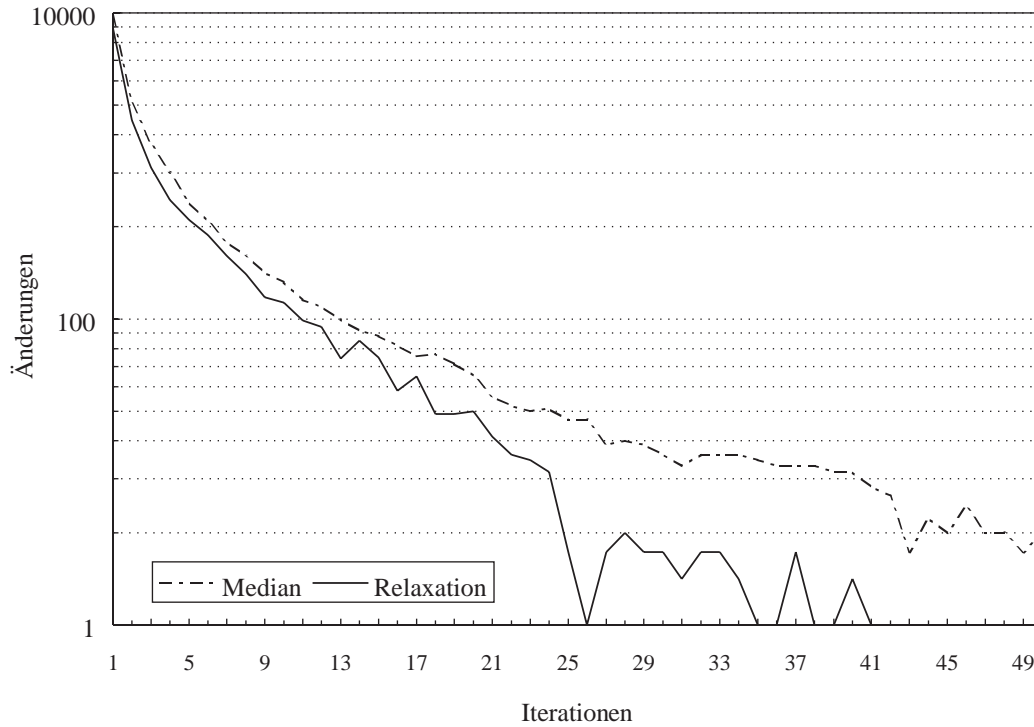


Bild 3.4: Anzahl der durch Medianfilterung bzw. Relaxation geänderten Farben pro Iteration bei einem 360×288 großen Bild mit 16 Farben. In beiden Verfahren wurde eine 3×3 Maske verwendet.

\mathbf{s}_n ist die Originalfarbe, \mathbf{r}_q eine der Referenzfarben. Die Normierungsgröße K wird so gewählt, daß $\sum_{i=1}^Q p_{n,i}^{(0)} = 1$. Die initiale Wahrscheinlichkeit eignet sich zur zeitlichen Glättung segmentierter Bildfolgen. So kann die Annahme, daß die meisten Bildpunkte ihre Farben aus dem vorhergehenden Bild beibehalten, durch den Faktor R_q berücksichtigt werden. Für die Referenzfarbe aus dem vorhergehenden Bild wird $R_q \geq 1$ gewählt, für alle anderen Referenzfarben ist $R_q = 1$. Nach Bestimmung der initialen Wahrscheinlichkeiten wird jedem Bildpunkt die Referenzfarbe mit der höchsten Wahrscheinlichkeit zugeordnet. In den folgenden Iterationsschritten der Relaxation wird eine $m \times m$ Maske über das Bild geschoben. Für den Bildpunkt n in der Mitte der Maske werden nun die Wahrscheinlichkeiten neu berechnet:

$$p_{n,q}^{(k)} = K p_{n,q}^{(k-1)} \sin^2 \left(\frac{r_q \pi / 2}{m^2} \right) \quad (3.7)$$

r_q gibt die Anzahl der dem Referenzvektor \mathbf{r}_q zugeordneten Bildpunkte innerhalb der Maske an, k ist der Iterationsindex. K dient wiederum der Normierung der Wahrscheinlichkeiten auf $\sum_{i=1}^Q p_{n,i}^{(k)} = 1$. Die \sin^2 -Funktion ist eine typische Wahl zur Modellierung des Einflusses der Nachbarpunkte. Denkbar sind aber auch andere Funktionen, die in Abhängigkeit von r_q stetig und streng monoton steigend verlaufen. Am Ende einer Iteration wird erneut jedem Bildpunkt die wahrscheinlichste Referenzfarbe zugewiesen. Die durchgezogene Kurve in Bild 3.4 zeigt die Anzahl der pro Iteration geänderten Farben. Sie weist einen ähnlichen Verlauf wie beim Medianfilter auf. Auch bei den in Bild 3.5 dargestellten segmentierten Bildern sind die Unterschiede zwischen Medianfilter und Relaxation geringfügig. Aus Tabelle 3.2 wird jedoch deutlich, daß der Rechenaufwand bei der Relaxation erheblich über dem bei der Medianfilterung liegt. Trotzdem ist die Relaxation für das Gesamtsystem nicht uninteressant, da Relaxationsverfahren in analoger Hardware⁴ realisierbar sind – der Rechenaufwand liegt dort im μs -Bereich.

⁴Ein Themenschwerpunkt im Projekt „Elektronisches Auge“, in dessen Rahmen die vorliegende Arbeit durchgeführt wurde, war die Entwicklung analoger Hardware zur Bildvorverarbeitung. In [Ranke et al. 97] werden u.a. hardwarenahe Relaxationsverfahren zur Glättung von quantisierten Farbbildern untersucht.

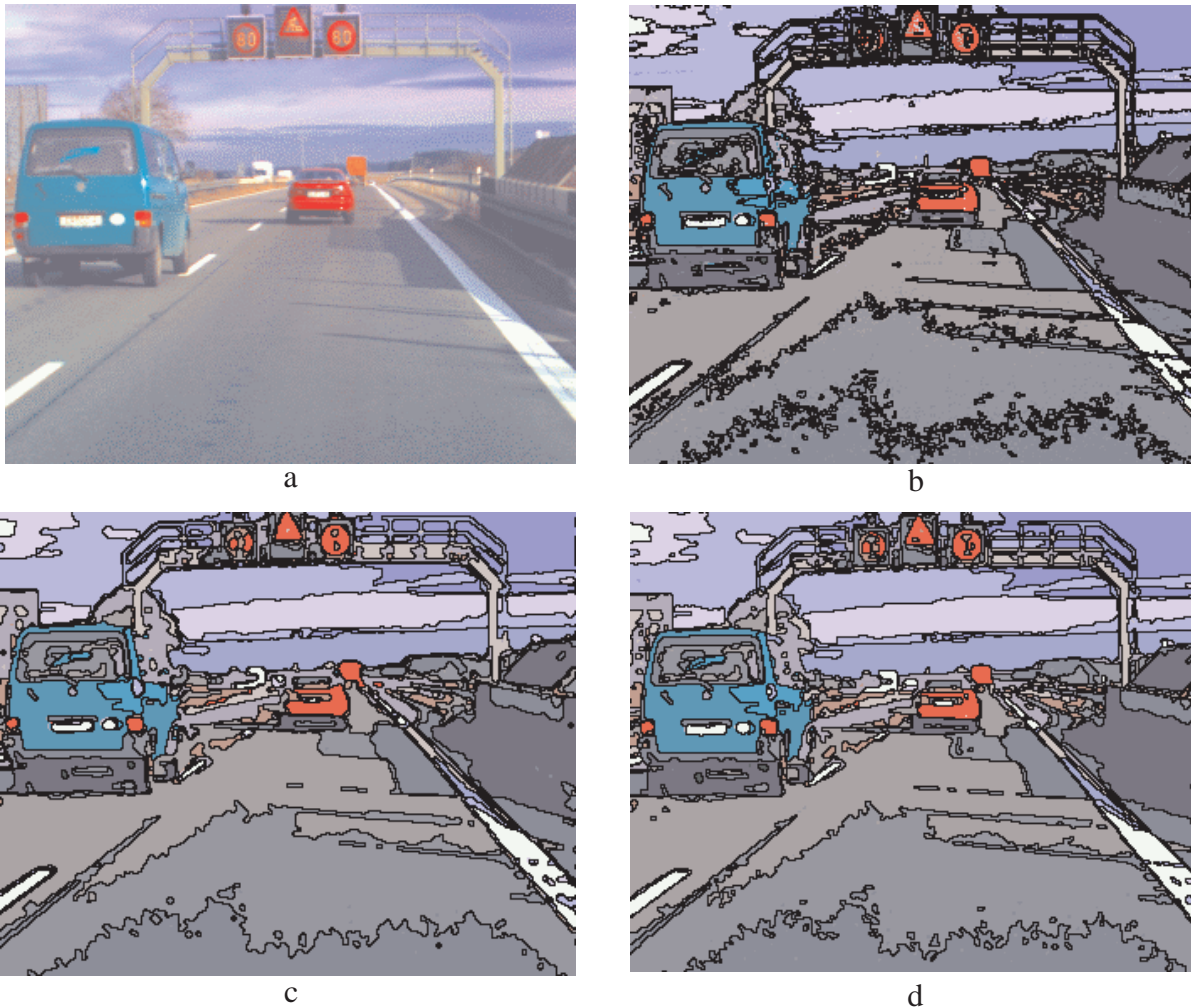


Bild 3.5: Beispiel der Segmentierung durch Farbquantisierung. In a) ist das Originalbild, in b) das mit dem Wu Verfahren auf 16 Farben quantisierte Bild dargestellt. In c) ist das mit dem Medianfilter nachbearbeitete Bild und in d) das Ergebnis der Nachbearbeitung durch Relaxation zu sehen. Zur Verdeutlichung der Glättung durch die Nachbearbeitung sind die Regionengrenzen in den segmentierten Bildern dunkel hervorgehoben.

3.1.4 Zusammenhangsanalyse

Die Zusammenhangsanalyse hat die Aufgabe, das segmentierte Punktbild ohne Informationsverlust in eine für die folgenden Verarbeitungsstufen besser handhabbare Darstellungsform zu transformieren. Da die Segmentierung eine Aufteilung des Bildes in Regionen durchführt, bietet sich statt der punkt- eine regionenbasierte Darstellung des segmentierten Bildes an. Wie bereits erwähnt, bezeichnet man als Farbregion eine zusammenhängende Fläche aus Bildpunkten gleicher Farbe. Zusammenhängend bedeutet, daß jeder Punkt der Region zumindest einen weiteren Punkt derselben Region in seiner Nachbarschaft hat. Zwei Punkte gelten als benachbart, wenn sie entweder horizontal oder vertikal im Raster unmittelbar nebeneinander liegen. Basierend auf dieser Regionendefinition liefert die in [Bartneck 87, Mandler & Oberländer 90] beschriebene Zusammenhangsanalyse eine Darstellung des segmentierten Bildes in Form eines Baumes (siehe Bild 3.6). Die Regionen sind die Knoten des Baumes; Lagebeziehungen der Regionen werden durch Kanten repräsentiert. Diese Datenstruktur bietet effiziente Möglichkeiten zur Auswahl von Regionen nach bestimmten Eigenschaften und zur Abfrage der Eigenschaften einer selektierten Region. Bleibt anzumerken, daß der Rechenaufwand der Zusammenhangsanalyse bei der Bearbeitung der Referenzsequenz etwa bei 160 ms auf einer Sun SPARC-20 lag.

3.2 Verfolgung von Farbregionen in Bildfolgen

Der nächste Schritt ist die Bewegungsbestimmung der Regionen in einer Bildfolge. Das Kernproblem ist die Zuordnung einander entsprechender Regionen in zwei aufeinanderfolgenden segmentierten Bildern. Bei einer korrekten Zuordnung sind die einander zugeordneten Regionen die Abbildungen desselben Szenenobjektes. Folgende Probleme treten bei der Zuordnung auf:

- Durch Bewegung der Kamera und der Objekte können neue Regionen auftauchen bzw. bisher vorhandene Regionen verschwinden.
- Störungen in den Bilddaten, Änderungen der Beleuchtung oder Teilverdeckungen können dazu führen, daß eine Region in mehrere zerfällt oder umgekehrt mehrere Regionen zu einer verschmelzen.

Die in dieser Arbeit gewählte Zuordnungsstrategie beruht auf einer 1 zu 1 Zuordnung (Paarbildung) von Regionen. Es wird ausgeschlossen, daß dieselbe Region in mehreren Paaren auftritt. Die Zuordnungskriterien für die Paarbildung basieren auf den durch die Zusammenhangsanalyse generierten Regionenmerkmalen. Lagerrelationen der Regionen werden nicht berücksichtigt. Zur quantitativen Bewertung von Paarbildungen wird

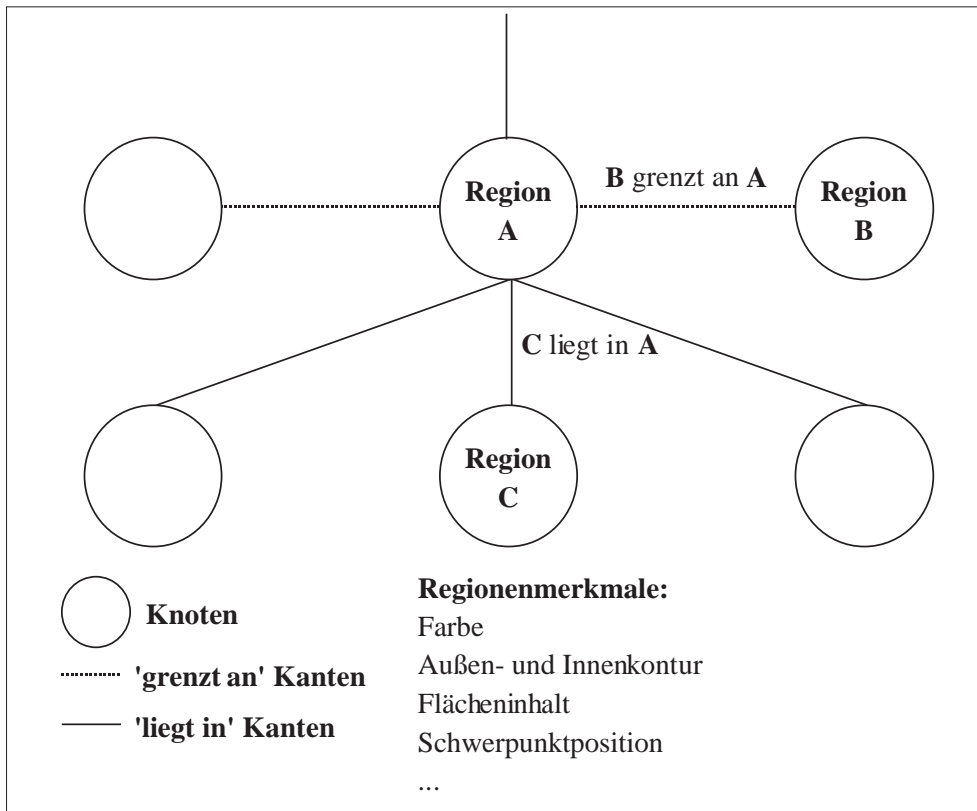


Bild 3.6: Schematische Darstellung der regionenbasierten Bildbeschreibung mittels Zusammenhangsanalyse. Regionen werden durch Knoten im Baum symbolisiert. Jeder Knoten enthält einen Satz von Merkmalen, durch den die zugehörige Region charakterisiert wird. Kanten repräsentieren die Lagebeziehungen der Regionen.

ein Fehler definiert, der die Abweichung zwischen den Merkmalen beider Regionen angibt. Um falsche Zuordnungen zu vermeiden, wird eine Obergrenze für den Fehler festgelegt. Die Zuordnungen werden so durchgeführt, daß jeder Region im zweiten Bild die Region mit minimalem Fehler zugewiesen wird. Ein Struktogramm der Zuordnungsprozedur ist in Anhang A.1 zu finden. Eine alternative Vorgehensweise ist die in [Rehrmann & Rothaar 97] beschriebene Anwendung von Graphzuordnungsverfahren (siehe z.B. [Deo 74]); damit können sowohl die Lagerrelationen als auch Aufteilungen bzw. Verschmelzungen von Regionen berücksichtigt werden. Allerdings stößt man bei Graphzuordnungsverfahren sehr schnell an die Grenze der verfügbaren Rechenzeit. Darüberhinaus kompliziert eine n zu m Zuordnung, bei der einer Gruppe aus n Regionen des ersten Bildes eine Gruppe von m Regionen des zweiten Bildes zugeordnet wird, die Bestimmung des Bewegungsverhaltens einzelner Regionen.

3.2.1 Festlegung des Zuordnungsfehlers für Regionen

Wie bereits angesprochen, ist für die Zuordnung von Regionen in Folgebildern ein Fehler zu definieren, durch den sich die Ähnlichkeit zweier Regionen⁵ quantitativ angeben läßt. Folgende Merkmale wurden zur Fehlerberechnung herangezogen:

- In der Gradientengleichung des optischen Flusses wurde die Invarianz der Intensität eines Punktes gegenüber seiner Bewegung vorausgesetzt. Analog kann die Farbe als von der Bewegung weitgehend unbeeinflusstes Merkmal betrachtet werden. Das Maß für die Abweichung der Farben zweier Regionen im RGB-Raum wird wie folgt definiert:

$$E_C = \sqrt{(R_A - R_B)^2 + (G_A - G_B)^2 + (B_A - B_B)^2} \quad (3.8)$$

Dabei ist (R, G, B) die Farbe der Region. Die Indizes A und B kennzeichnen die zwei zu vergleichenden Regionen.

- Der Flächeninhalt einer Region ist unabhängig von translatorischer Bewegung parallel zur Bildebene und rotatorischer Bewegung um die optische Achse. Bei den restlichen drei Bewegungsvarianten ändert sich jedoch der Flächeninhalt. Diese Änderungen sind bei den vorliegenden Bildfolgen – hohe zeitliche Abtastung und geringe Rotationsbewegungen – relativ klein und erlauben die Verwendung des Flächeninhalts als Zuordnungskriterium. Die Abweichung des Flächeninhalts zweier Regionen wird wie folgt berechnet:

$$E_F = \frac{|F_A - F_B|}{F_A + F_B} \quad (3.9)$$

wobei F den Flächeninhalt gemessen in Bildpunkten angibt.

- Das Seitenverhältnis des umschreibenden Rechtecks einer Region ist ein leicht zu berechnendes Formmerkmal einer Region. Es ist unabhängig von translatorischer, nicht jedoch von rotatorischer Bewegung. Berücksichtigt man, daß in den meisten Verkehrsszenen nur geringe Rotationsbewegungen auftreten, dann stellt das Seitenverhältnis ein brauchbares Ähnlichkeitsmerkmal dar. Das verwendete Fehlermaß lautet:

$$E_S = \frac{|S_A - S_B|}{S_A + S_B} \quad (3.10)$$

wobei S das Verhältnis aus Breite und Höhe (in Bildpunkten) des umschreibenden Rechtecks einer Region ist. Das Rechteck wird so gewählt, daß seine Seiten parallel zu den Bildkoordinatenachsen verlaufen.

⁵Wie in Kapitel 2.3.1 wird in den folgenden Betrachtungen davon ausgegangen, daß eine Region im segmentierten Bild einer starren, annähernd planaren Szenenfläche entspricht.

- Da in den vorliegenden Sequenzen die Verschiebungen von Bildobjekten meist im Bereich $< 1/25$ der Bildbreite liegen, kann die Schwerpunktposition als Ähnlichkeitsmerkmal verwendet werden. Das Fehlermaß für die Position lautet:

$$E_P = \sqrt{(x_A - x_B)^2 + (y_A - y_B)^2} \quad (3.11)$$

wobei x die horizontale und y die vertikale Komponente der in Bildpunkten gemessenen Schwerpunktposition ist.

Der Fehler für eine Paarbildung wird nun als Summe der gewichteten Einzelfehler berechnet:

$$E = w_C E_C + w_F E_F + w_P E_P + w_s E_s \quad (3.12)$$

Die Gewichte sind zum einen aufgrund der unterschiedlichen Größenordnungen der Einzelfehler erforderlich, zum andern ermöglichen sie eine Bewertung der Merkmale bezüglich ihrer Relevanz für die Zuordnung. Für die Bestimmung der Gewichte wären segmentierte Bildsequenzen ideal, in denen die korrekten Regionenpaarbildungen bekannt sind. Solche Lernsequenzen müßten für jedes Segmentierungsverfahren durch manuelle Bearbeitung erstellt werden. Der Aufwand wäre allerdings erheblich, da sich in einem Bild mehrere hundert Farbregionen befinden, und bei Änderungen in der Farbsegmentierung die Lernsequenzen jeweils erneut erstellt werden müßten. In der vorliegenden Arbeit wird eine Alternative gewählt. Die Gewichte werden mit Hilfe eines zweistufigen Verfahrens bestimmt, das auf eine ausgewählte Referenzsequenz⁶ angewandt wird. Es erfordert keine Kenntnisse über die korrekten Regionenzuordnungen und läßt sich auf beliebige Segmentierungsverfahren anwenden.

Im ersten Schritt werden die Bilder der Referenzsequenz unabhängig voneinander untersucht. Um festzustellen, in welchen Größenordnungen die Einzelfehler bei einer *falschen* Zuordnung liegen, wird jede Region mit den sie umgebenden Regionen im selben Bild verglichen. Für jeden Vergleich werden die Einzelfehler bestimmt und festgehalten. Betrachtet werden ausschließlich Regionen, deren Flächeninhalt > 20 und < 50000 Bildpunkte ist. Damit schließt man sehr kleine bzw. sehr große Regionen, die für die Objektdetektion ohne Bedeutung sind, frühzeitig von der Untersuchung aus. Die Umgebung einer Region ist als kreisförmiger Bildbereich mit dem Schwerpunkt der jeweiligen Region als Mittelpunkt definiert. Alle Regionen, deren Schwerpunkte innerhalb dieses Kreises liegen, werden zum

⁶Die Referenzsequenz wurde bei einer Geradeausfahrt auf der Autobahn aufgenommen, die Geschwindigkeit des Versuchsfahrzeugs lag etwa bei 80 km/h. Aufgenommen wurde mit einer 3 Chip CCD-Farbkamera (Sensorfläche $4,8 \times 3,7$ mm) und einem 16 mm Objektiv. Als Aufzeichnungsgerät diente ein digitaler Videorecorder, der im YUV 4:2:2 Format aufzeichnet. Ausgewertet wurde jedes zweite Halbbild mit jeweils halber Spaltenzahl (360×288) und 24 Bit RGB-Werten pro Bildpunkt. Die Sequenz umfaßt 230 Bilder.

Vergleich herangezogen (siehe Bild 3.7). Der Radius des Umgebungsbereichs wird auf 15 Bildpunkte festgelegt. Tabelle 3.3 zeigt einige Ergebnisse für die Referenzsequenz bei unterschiedlichen Segmentierungsverfahren. Die Gewichte werden in Tabelle 3.5 so gewählt, daß die Mittelwerte der Einzelfehler auf 1 normiert werden.

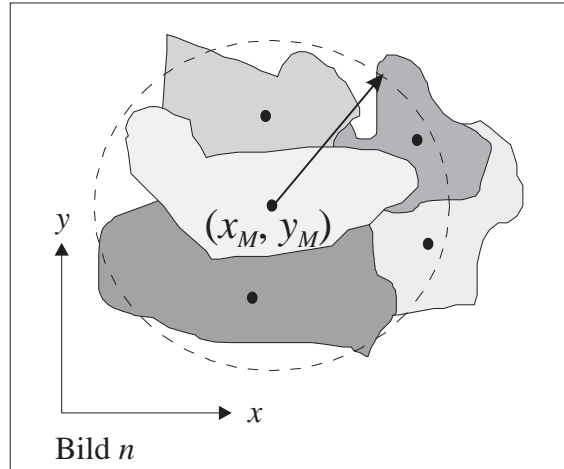


Bild 3.7: Zur Bestimmung der Größenordnungen der Einzelfehler wird jede Region mit den sie umgebenden Regionen verglichen. Die Umgebung ist kreisförmig und hat den Flächenschwerpunkt (x_M, y_M) der jeweiligen Region als Mittelpunkt. Alle Regionen, deren Schwerpunkte innerhalb dieses Kreises liegen, werden zum Vergleich herangezogen.

Tabelle 3.3: Mittlere Einzelfehler beim Vergleich zwischen benachbarten Regionen (Distanz der Schwerpunkte < 15 Bildpunkte) eines Bildes. Es wurden ausschließlich Regionen mit einem Flächeninhalt > 20 und < 50000 Bildpunkte betrachtet. Die Farbpfeferenzvektoren wurden für jedes zehnte Bild der Sequenz mit dem Wu Verfahren neu berechnet. Bei Medianfilterung und Relaxation wurden jeweils 3 Iterationen mit einer 3×3 Maske durchgeführt. Bei der zeitlichen Glättung in der Relaxation wurde die Farbe aus dem vorhergehenden Bild entsprechend Gl. (3.6) mit $R_q = 2$ gewichtet.

Segmentierung	\bar{E}_C	\bar{E}_F	\bar{E}_S	\bar{E}_P
Farbquantisierung + Medianfilter	81	0,40	0,27	10
Farbquantisierung + Relaxation	85	0,40	0,28	10
Farbquantisierung + Relaxation mit zeitlicher Glättung	88	0,41	0,27	10

Im zweiten Schritt soll nun untersucht werden, wie sich die Merkmale in aufeinanderfolgenden Bildern verhalten. Im Unterschied zum ersten Schritt, wird eine Region mit Regionen des Folgebilds verglichen. Die Menge der zum Vergleich herangezogenen Regionen wird wieder durch einen kreisförmigen Umgebungsbereich, jetzt aber im darauffolgenden Bild, eingeschränkt (siehe Bild 3.8). Die Einzelfehler werden mit den im ersten Schritt bestimmten Gewichten multipliziert. Für jede Region und jedes Merkmal X wird der minimale Einzelfehler E_X^{min} und die dabei auftretende Summe der vier gewichteten Einzelfehler $E(X)$ festgehalten. Der erste Wert gibt die Untergrenze für die Bild-zu-Bild Änderung des Merkmals X an. Je kleiner der Wert, desto robuster das Merkmal. Der zweite Wert sagt aus, wie sich der Gesamtfehler bei einer Zuordnung aufgrund des Merkmals X verhält. Auch hier gilt, je kleiner der Gesamtfehler, desto eher eignet sich das Merkmal für die Zuordnung. Die Ergebnisse sind in Tabelle 3.4 dargestellt. Bei der Farbe ist nicht nur der minimale Fehler am geringsten, sondern auch der damit verbundene Gesamtfehler kleiner als bei den anderen Merkmalen. Dies deckt sich mit der bereits getroffenen Feststellung, daß die Farbe eines Bildobjekts weitgehend unabhängig von dessen Bewegung ist. Die Schwerpunktposition weist gegenüber der Fläche und dem Seitenverhältnis einen kleineren Gesamtfehler auf. Als Konsequenz wurden die Gewichte so verändert, daß der Farbe und der Position eine erhöhte Bedeutung im Zuordnungsprozeß zufallen (siehe Tabelle 3.5). Aus Tabelle 3.4 ist außerdem zu erkennen, daß die Relaxation im Vergleich zur Medianfilterung zu einer zeitlich stabileren Segmentierung führt. Eine weitere Verbesserung der Segmentierung kann durch die in die Relaxation integrierte zeitliche Glättung verzeichnet werden.

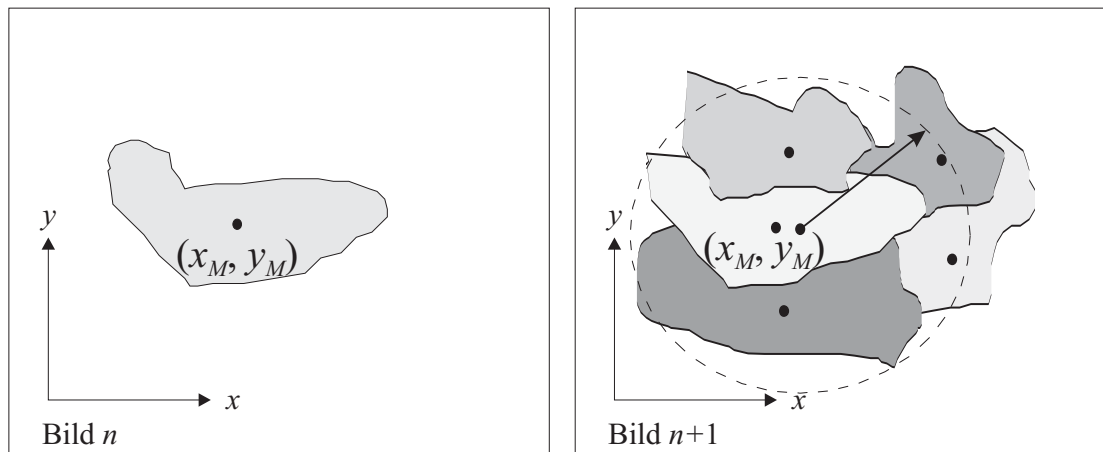


Bild 3.8: Zur Bestimmung der Fehlergewichtung wird jede Region mit Regionen des Folgebilds verglichen. Die Flächenschwerpunkte der zum Vergleich herangezogenen Regionen des Folgebilds müssen in einer kreisförmigen Umgebung um den Flächenschwerpunkt (x_M, y_M) liegen.

Tabelle 3.4: Mittlere Zuordnungsfehler beim Vergleich zwischen Regionen in zwei aufeinanderfolgenden Bildern. Der Wert $w_X \bar{E}_X^{min}$ ist der mittlere minimale Einzelfehler für das Merkmal X . Zusätzlich ist der mittlere gewichtete Gesamtfehler $\bar{E}(X)$ angegeben, der sich bei der Zuordnung aufgrund des Merkmals X ergibt. Es gelten dieselben Bedingungen wie in Tabelle 3.3.

Segm.	$w_C \bar{E}_C^{min}$	$\bar{E}(C)$	$w_F \bar{E}_F^{min}$	$\bar{E}(F)$	$w_S \bar{E}_S^{min}$	$\bar{E}(S)$	$w_P \bar{E}_P^{min}$	$\bar{E}(P)$
F.+M.	0,10	1,46	0,25	1,73	0,28	1,65	0,35	1,46
F.+R.	0,09	1,41	0,23	1,70	0,27	1,59	0,32	1,38
F.+R.	0,07	1,18	0,19	1,45	0,23	1,35	0,27	1,13
m. Gl.								

Tabelle 3.5: Festlegung der Gewichte für die Zuordnungsfehler

Schritt	w_C	w_F	w_S	w_P
1	0,012	2,5	3,6	0,10
2	0,048	2,5	3,6	0,20

3.2.2 Festlegung der Obergrenzen für die Zuordnungsfehler

In Bildsequenzen können Regionen verschwinden bzw. neue Regionen auftauchen. Um solche Effekte zu berücksichtigen und Fehlzuordnungen zu vermeiden, ist die Einführung von Obergrenzen für die Einzelfehler sinnvoll. Zuordnungen werden nur erlaubt, wenn alle Einzelfehler unterhalb der jeweiligen Obergrenze liegen. Anhaltspunkte für die Wahl der Obergrenzen bei Positions- und Flächenfehlern können am Beispiel eines schnell überholenden Fahrzeugs abgeleitet werden. Wenn das Fahrzeug mit einer Relativgeschwindigkeit von 50 km/h und einem Seitenabstand von 1 m überholt, tritt bei der gegebenen Aufnahmeeinrichtung eine maximale Verschiebung von 12,5 Bildpunkten und ein maximaler Flächenfehler von 0,08 auf. Dabei wird eine planare Szenenfläche angenommen, deren Normale parallel zur Bewegungsrichtung verläuft. Die durch Rotationsbewegungen verursachten Änderungen im Seitenverhältnis sind bei den untersuchten Szenen in der Regel vernachlässigbar. Die Einstellung der Obergrenze für den Farbfehler erfolgte anhand einer Begutachtung der Farbbänderungen in der Referenzsequenz. In Tabelle 3.6 sind die Obergrenzen angegeben.

Tabelle 3.6: Obergrenzen für die Zuordnungsfehler

\bar{E}_C	\bar{E}_F	\bar{E}_S	\bar{E}_P
100	0,2	0,2	15

3.2.3 Prädiktion von Bewegungen

Um das Wiederauffinden einer Farbregion im Folgebild zu stabilisieren, kann die während der Beobachtung angesammelte Information über die Bewegung einer Region genutzt werden, um deren Position im nächsten Bild vorherzusagen. Bei der Berechnung des Positionsfehlers in der Zuordnungsprozedur wird dann die prädizierte Schwerpunktposition verwendet. Besonders hilfreich ist die Prädiktion, wenn bei der Verfolgung von Regionen Lücken auftreten. Beispielsweise kann eine Region durch Verdeckungen oder Bildstörungen verschwinden und kurze Zeit später wieder auftauchen. Die während dieses Zeitraums erfolgte Schwerpunktverschiebung überschreitet häufig die im Zuordnungsverfahren definierte Obergrenze. Durch geeignete Prädiktion kann nun die Position zum Zeitpunkt des Wiederauftauchens so vorhergesagt werden, daß eine korrekte Zuordnung bei gleicher Obergrenze möglich ist.

Exponentielle Mittelung

Die exponentielle Mittelung erfordert keine Kenntnisse über den datenerzeugenden Prozeß. Anhand einer Reihe von zeitdiskreten Messungen, $\mathbf{y}(1) \dots \mathbf{y}(k-1)$, erfolgt die Prädiktion $\hat{\mathbf{y}}$ zum Zeitpunkt kT , wobei T die Abtastperiode ist, als gewichtete Summe aller bisherigen Meßwerte:

$$\hat{\mathbf{y}}(k) = \sum_{n=1}^{k-1} (1 - \alpha)^{(k-1-n)} \alpha \mathbf{y}(n) \quad 0 \leq \alpha \leq 1 \quad (3.13)$$

Je größer der Parameter α gewählt wird, desto stärker wird die Prädiktion durch die aktuelle Messung bestimmt. Obige Gleichung läßt sich in rekursiver Form schreiben, so daß nur der letzte Wert der Datenreihe und der letzte prädizierte Wert zur Berechnung von $\hat{\mathbf{y}}(k)$ erforderlich sind:

$$\hat{\mathbf{y}}(k) = (1 - \alpha)\hat{\mathbf{y}}(k-1) + \alpha \mathbf{y}(k-1) \quad (3.14)$$

Bei der Verfolgung von Regionen wird die exponentielle Mittelung auf die Verschiebungen der Schwerpunkte angewandt. Um Verwechslungen mit dem Meßvektor $\mathbf{y}(k)$ zu vermeiden, wird die Schwerpunktposition einer Region in den folgenden beiden Gleichungen mit

(ξ_x, ξ_y) bezeichnet. Der Meßvektor beinhaltet die Bild-zu-Bild Verschiebung des Regionschwerpunkts und lautet somit:

$$\mathbf{y}(k) = (\Delta\xi_x(k), \Delta\xi_y(k))^T \quad (3.15)$$

Die prädiizierte Position eines Schwerpunktes ist demnach:

$$(\hat{\xi}_x(k), \hat{\xi}_y(k))^T = (\xi_x(k-1), \xi_y(k-1))^T + (\Delta\hat{\xi}_x(k-1), \Delta\hat{\xi}_y(k-1))^T \quad (3.16)$$

Kalmanfilter

Bei der Bildfolgenanalyse stellt der aus der Regelungstechnik stammende Kalmanfilter ein Standardverfahren zur Prädiktion von Bewegungen dar (siehe z.B. [Koller et al. 93, Dickmanns 94, Meyer & Bouthemy 92]). Der Kalmanfilter ermöglicht bei stochastischen, linearen Systemen eine rekursive, optimale Schätzung der Systemgrößen. Im Unterschied zur exponentiellen Mittelung fließen beim Kalmanfilter Kenntnisse über den datenerzeugenden Prozeß in Form eines Systemmodells ein. Der allgemeine Aufbau eines zeitdiskreten, linearen Systems ist in Bild 3.9 dargestellt. System- und Meßgleichung lauten:

$$\mathbf{x}(k+1) = \mathbf{A}(k)\mathbf{x}(k) + \mathbf{B}(k)\mathbf{u}(k) \quad (3.17)$$

$$\mathbf{y}(k) = \mathbf{C}(k)\mathbf{x}(k) \quad (3.18)$$

Wobei \mathbf{x} die Systemzustandsgrößen, \mathbf{y} die Ausgangs- bzw. Meßgrößen und \mathbf{u} die Steuergrößen sind. Entsprechend bezeichnet man \mathbf{A} als System- bzw. Transitionsmatrix, \mathbf{C} als Meßmatrix und \mathbf{B} als Steuermatrix. Erweitert man die Zustandsraumgleichungen um weiße, gaußverteilte Rauschprozesse so entsteht ein stochastisches, lineares System:

$$\mathbf{x}(k+1) = \mathbf{A}(k)\mathbf{x}(k) + \mathbf{B}(k)\mathbf{u}(k) + \mathbf{w}(k) \quad (3.19)$$

$$\mathbf{y}(k) = \mathbf{C}(k)\mathbf{x}(k) + \mathbf{v}(k) \quad (3.20)$$

Dabei modelliert \mathbf{v} die Meßungenauigkeit und \mathbf{w} die Ungenauigkeit des Modells in bezug auf das reale dynamische System. Im folgenden wird vorausgesetzt, daß die Rauschprozesse unabhängig voneinander sind und die Kovarianzmatrizen $\mathbf{Q}(k)$ von $\mathbf{w}(k)$ und $\mathbf{R}(k)$ von $\mathbf{v}(k)$ bekannt sind. Der Kalmanfilter kann für ein durch die Gleichungen (3.19) und (3.20) beschriebenes System anhand der Meßergebnisse $\mathbf{y}(1) \dots \mathbf{y}(k)$ eine optimale Schätzung der Systemzustandsgrößen \mathbf{x} zum Zeitpunkt kT abgeben. Damit ist auch eine optimale Prädiktion von \mathbf{x} zum Zeitpunkt $(k+1)T$ aufgrund der Messungen $\mathbf{y}(1) \dots \mathbf{y}(k)$ möglich. Der Kalmanfilter ist in zwei Schritte unterteilt – die Prädiktion und die Innovation. Während der Filterung erfolgt abwechselnd ein Prädiktionsschritt zur Vorhersage der Systemgrößen und ein Innovationsschritt zur Verbesserung der vorhergesagten Werte nach Eintreffen eines neuen Meßvektors. In Anlehnung an [Loffeld 90] werden die in der

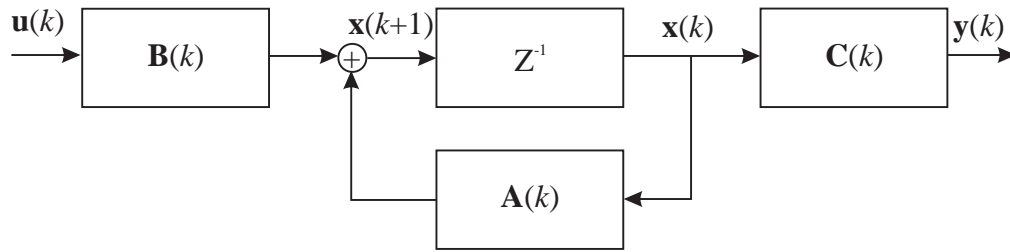


Bild 3.9: Zeitdiskretes, lineares System

Prädiktion geschätzten Größen durch '−', die in der Innovation geschätzten Größen durch '+', gekennzeichnet.

- Im *Prädiktionsschritt* liegen die Messungen $\mathbf{y}(1) \dots \mathbf{y}(k-1)$ vor. Nun wird die gaußförmige Prädiktionsdichte $p(\mathbf{x}(k)|\mathbf{y}(1) \dots \mathbf{y}(k-1))$ bestimmt:

$$p(\mathbf{x}(k)|\mathbf{y}(1) \dots \mathbf{y}(k-1)) = \frac{1}{\sqrt{(2\pi)^N \det \mathbf{P}^-(k)}} e^{\Theta} \quad (3.21)$$

dabei sind:

$$N = \text{Dimension von } \mathbf{x}(k) \quad (3.22)$$

$$\Theta = -1/2[\mathbf{x}(k) - \hat{\mathbf{x}}^-(k)]^T \mathbf{P}^-(k)^{-1} [\mathbf{x}(k) - \hat{\mathbf{x}}^-(k)] \quad (3.23)$$

$$\hat{\mathbf{x}}^-(k) = \mathbf{A}(k-1)\hat{\mathbf{x}}^+(k-1) + \mathbf{B}(k-1)\mathbf{u}(k-1) \quad (3.24)$$

$$\mathbf{P}^-(k) = \mathbf{A}(k-1)\mathbf{P}^+(k-1)\mathbf{A}(k-1)^T + \mathbf{Q}(k-1) \quad (3.25)$$

$\hat{\mathbf{x}}^-(k)$ ist der Erwartungswert von $\mathbf{x}(k)$, basierend auf den Messungen $\mathbf{y}(1) \dots \mathbf{y}(k-1)$. Er wird durch Extrapolation des besten Schätzwertes von \mathbf{x} zum Zeitpunkt $(k-1)T$ berechnet. Das Rauschen \mathbf{w} hat den Erwartungswert 0 und geht damit nicht in die Prädiktion ein. Die Schätzfehlerkovarianzmatrix $\mathbf{P}^-(k)$ besteht aus zwei Summanden. Der erste beschreibt die Wirkung der Systemmatrix \mathbf{A} , der zweite die des Systemrauschens \mathbf{w} .

- Im *Innovationsschritt* wird nun die eingetroffene Messung $\mathbf{y}(k)$ genutzt, um die Filterdichte $p(\mathbf{x}(k)|\mathbf{y}(1) \dots \mathbf{y}(k))$ zu berechnen. Sie lautet:

$$p(\mathbf{x}(k)|\mathbf{y}(1) \dots \mathbf{y}(k)) = \frac{1}{\sqrt{(2\pi)^N \det \mathbf{P}^+(k)}} e^{\Theta} \quad (3.26)$$

dabei sind:

$$N = \text{Dimension von } \mathbf{x}(k) \quad (3.27)$$

$$\Theta = -1/2[\mathbf{x}(k) - \hat{\mathbf{x}}^+(k)]^T \mathbf{P}^+(k)^{-1} [\mathbf{x}(k) - \hat{\mathbf{x}}^+(k)] \quad (3.28)$$

$$\hat{\mathbf{x}}^+(k) = \hat{\mathbf{x}}^-(k) + \mathbf{K}(k)[\mathbf{y}(k) - \mathbf{C}(k)\hat{\mathbf{x}}^-(k)] \quad (3.29)$$

$$\mathbf{P}^+(k) = \mathbf{P}^-(k) - \mathbf{K}(k)\mathbf{C}(k)\mathbf{P}^-(k) \quad (3.30)$$

$$\mathbf{K}(k) = \mathbf{P}^-(k)\mathbf{C}(k)^T[\mathbf{C}(k)\mathbf{P}^-(k)\mathbf{C}(k)^T + \mathbf{R}(k)]^{-1} \quad (3.31)$$

Der optimale Schätzwert $\hat{\mathbf{x}}^+(k)$ ist die Summe aus dem Prädiktionswert $\hat{\mathbf{x}}^-(k)$ und der mit der Kalman-Gain Matrix $\mathbf{K}(k)$ multiplizierten Differenz aus tatsächlichem Meßwert und erwartetem Meßwert. Die Bedeutung der Kalman-Gain Matrix läßt sich im Fall skalarer Größen veranschaulichen. Es gilt dann:

$$K(k) = \frac{C(k)P^-(k)}{C(k)^2P^-(k) + R(k)} \quad (3.32)$$

Im Fall von $R(k) \rightarrow \infty$ bzw. $P^-(k) \rightarrow 0$ ist die Messung sehr stark verrauscht bzw. das Vorwissen sehr genau. In beiden Fällen geht $K(k)$ gegen 0, so daß die Messung $y(k)$ bei der Prädiktion von $x(k)$ vollständig ignoriert wird. Im umgekehrten Fall, von $R(k) \rightarrow 0$ bzw. $P^-(k) \rightarrow \infty$, ist die Messung sehr genau bzw. das Vorwissen sehr ungenau. Die Prädiktion von $x(k)$ beruht dann ausschließlich auf der neuesten Messung $y(k)$.

Der erste Schritt in der Kalmanfilterung ist ein Prädiktionsschritt. Vor Beginn der Filterung müssen somit die Größen $\hat{\mathbf{x}}^+(0)$ und $\hat{\mathbf{P}}^+(0)$ initialisiert werden.

Um den Kalmanfilter auf die Problematik der Regionenverfolgung in Bildsequenzen anwenden zu können, muß ein Bewegungsmodell für die Regionen entworfen werden. Da nicht von vornherein bekannt ist, zu welchem Objekt eine Region gehört, scheiden komplexe objektspezifische Modelle aus (z.B. Fahrzeug- oder Fußgängermodelle). Statt dessen werden zwei einfache Modelle untersucht, die von einer zeitkontinuierlichen Bewegung mit weitgehend konstanter Geschwindigkeit bzw. weitgehend konstanter Beschleunigung ausgehen [Bar-Shalom & Li 93]. In beiden Modellen wird vorausgesetzt, daß die Bewegungen der Flächenschwerpunkte der Regionen in x - und y -Richtung voneinander entkoppelt sind. Für jede Region werden somit zwei Kalmanfilter eingesetzt, die unabhängig voneinander die Bild-zu-Bild Bewegung des Flächenschwerpunkts der jeweiligen Region in x - bzw. y -Richtung vorhersagen. Aufgrund der Entkopplung der Kalmanfilter, beschränken sich die Betrachtungen auf eindimensionale Bewegungen. Zur Unterscheidung vom Systemzustandsvektor \mathbf{x} wird die Positionskoordinate im folgenden mit ξ bezeichnet.

Modellierung der Beschleunigung als weißes Rauschen Diesem Modell wird eine Bewegung mit konstanter Geschwindigkeit zugrunde gelegt. Um in der Realität vorkommende Abweichungen von der konstanten Geschwindigkeit zu berücksichtigen, wird die Beschleunigung als weißes Rauschen modelliert:

$$\frac{d^2\xi(t)}{dt^2} = w(t) \quad (3.33)$$

Man erhält folgendes zeitkontinuierliches Modell:

$$\frac{d\mathbf{x}(t)}{dt} = \mathbf{F}\mathbf{x}(t) + \mathbf{w}(t) \quad (3.34)$$

mit

$$\mathbf{x}(t) = \left(\xi, \frac{d\xi}{dt} \right)^T \quad (3.35)$$

$$\mathbf{F} = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \quad (3.36)$$

$$\mathbf{w}(t) = (0, w(t))^T \quad (3.37)$$

Geht man zu einem zeitdiskreten System mit der Abtastperiode T über, so erhält man (siehe Anhang B.1):

$$\mathbf{x}(k+1) = \mathbf{A}\mathbf{x}(k) + \mathbf{w}(k) \quad (3.38)$$

wobei

$$\mathbf{A} = \begin{pmatrix} 1 & T \\ 0 & 1 \end{pmatrix} \quad (3.39)$$

$$\mathbf{Q} = \begin{pmatrix} \frac{1}{3}T^3 & \frac{1}{2}T^2 \\ \frac{1}{2}T^2 & T \end{pmatrix} \sigma_w^2 \quad (3.40)$$

Gemessen wird die Position. Die Meßgleichung lautet somit:

$$\mathbf{y}(k) = \mathbf{C}\mathbf{x}(k) + v(k) \quad (3.41)$$

mit

$$\mathbf{C} = (1 \ 0) \quad (3.42)$$

$$R = \sigma_v^2 \quad (3.43)$$

Die Initialisierung des Filters und die Wahl der Parameter σ_w^2 und σ_v^2 ist in Anhang B.1 beschrieben.

Modellierung der Ableitung der Beschleunigung als weißes Rauschen Dieses Modell ist weniger restriktiv als das bereits genannte. Es basiert auf einer zeitkontinuierlichen Bewegung mit konstanter Beschleunigung. Auch hier werden Abweichungen der realen Bewegung durch das Systemrauschen berücksichtigt:

$$\frac{d^3\xi(t)}{dt^3} = w(t) \quad (3.44)$$

Die zeitkontinuierliche Systemgleichung lautet:

$$\frac{d\mathbf{x}(t)}{dt} = \mathbf{F}\mathbf{x}(t) + \mathbf{w}(t) \quad (3.45)$$

mit

$$\mathbf{x}(t) = \left(\xi, \frac{d\xi}{dt}, \frac{d^2\xi}{dt^2} \right)^T \quad (3.46)$$

$$\mathbf{F} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix} \quad (3.47)$$

$$\mathbf{w}(t) = (0, 0, w(t))^T \quad (3.48)$$

Der Übergang zu einem zeitdiskreten System mit der Abtastperiode T ist in Anhang B.2 erläutert. Man erhält folgende Systemgleichung:

$$\mathbf{x}(k+1) = \mathbf{A}\mathbf{x}(k) + \mathbf{w}(k) \quad (3.49)$$

wobei

$$\mathbf{A} = \begin{pmatrix} 1 & T & \frac{1}{2}T^2 \\ 0 & 1 & T \\ 0 & 0 & 1 \end{pmatrix} \quad (3.50)$$

$$\mathbf{Q} = E[\mathbf{w}(k)\mathbf{w}(k)^T] = \begin{pmatrix} \frac{1}{20}T^5 & \frac{1}{8}T^4 & \frac{1}{6}T^3 \\ \frac{1}{8}T^4 & \frac{1}{3}T^3 & \frac{1}{2}T^2 \\ \frac{1}{6}T^3 & \frac{1}{2}T^2 & T \end{pmatrix} \sigma_w^2 \quad (3.51)$$

Gemessen wird die Position. Damit lautet die Meßgleichung:

$$\mathbf{y}(k) = \mathbf{C}\mathbf{x}(k) + v(k) \quad (3.52)$$

mit

$$\mathbf{C} = \begin{pmatrix} 1 & 0 & 0 \end{pmatrix} \quad (3.53)$$

$$R = \sigma_v^2 \quad (3.54)$$

Auch hier sei zur Parametereinstellung und Initialisierung des Filters auf den Anhang B.2 verwiesen.

Experimentelle Untersuchung der Prädiktionsverfahren

Zur Beurteilung der Prädiktionstechniken wurde die Referenzsequenz herangezogen. Als Kriterien für die Güte der Verfahren wurden die Zuordnungsfehler bestimmt. Die Ergebnisse sind in Tabelle 3.7 dargestellt. Offenbar erreicht man durch die untersuchten

Prädiktionsverfahren nur eine geringe Reduktion der Zuordnungsfehler. Dies läßt sich dadurch erklären, daß eine Großzahl der Regionen in der Referenzsequenz nur eine schwache Bewegung in der Bildebene aufweist. Damit stellt die Annahme, daß sich der Flächenschwerpunkt an derselben Stelle wie im vorhergehenden Bild befindet, bereits eine relativ gute Vorhersage dar. Bei schnell bewegten Regionen konnte jedoch durch Prädiktion die Genauigkeit der Schwerpunktvorhersage um über 50% gesteigert werden. Die Ergebnisse in Tabelle 3.7 zeigen nur einen geringen Unterschied zwischen der modellfreien Prädiktion durch exponentielle Mittelung und der Kalmanfilterung. Dies liegt daran, daß die Systemmodelle im Kalmanfilter die realen Verhältnisse nur unvollkommen beschreiben. Zum Beispiel werden Nickbewegungen des Versuchsfahrzeugs, die zu abrupten Geschwindigkeitsänderungen der Regionen in der Bildebene führen, nicht im Modell berücksichtigt. Der Kalmanfilter bietet gegenüber der exponentiellen Mittelung jedoch den Vorteil, neben der Prädiktion auch eine Glättung der Meßwerte durchzuführen. Da in der Bewegungssegmentierung entauschte Trajektorienverläufe erwünscht sind, wird im weiteren Verlauf der Arbeit ausschließlich die Kalmanfilterung mit Modell 1 eingesetzt. Es muß allerdings bemerkt werden, daß der Rechenaufwand in der Regionenzuordnung durch den Kalmanfilter von 30 ms auf 80 ms pro Bild erhöht wurde. Die Zeitmessungen wurden auf einer Sun SPARC-20 durchgeführt. In Bild 3.10 sind die Filtergrößen bei diesem Modell für ein überholendes Fahrzeug exemplarisch aufgetragen. Etwa ab Bild 30 setzt eine Nickbewegung des Versuchsfahrzeugs ein. Deutlich sind die dadurch erhöhten Abweichungen zwischen geschätzten und gemessenen Positionswerten zu erkennen. In Bild 3.11 sind die gemessenen und in der Innovationsstufe berechneten Trajektorien einander gegenübergestellt. Die Glättungswirkung des Kalmanfilters wird besonders an den Regionen des schnell überholenden Fahrzeugs auf der linken Fahrbahnseite deutlich.

Tabelle 3.7: Mittlere Zuordnungsfehler für unterschiedliche Prädiktionsverfahren. EM steht für exponentielle Mittelung; KF für Kalmanfilterung. Im Modell 1 wird die Beschleunigung als weißes Rauschen modelliert; im Modell 2 wird die zeitliche Ableitung der Beschleunigung als weißes Rauschen modelliert. Die Fehler wurden für die farbquantisierte und anschließend relaxierte Referenzsequenz bestimmt.

Verfahren	\bar{E}_C	\bar{E}_F	\bar{E}_S	\bar{E}_P	\bar{E}
keine Prädiktion	2,17	0,04	0,03	2,00	0,72
EM mit $\alpha = 0,75$	2,13	0,04	0,03	1,86	0,69
KF Modell 1	2,16	0,04	0,03	1,79	0,68
KF Modell 2	2,17	0,04	0,03	1,90	0,69

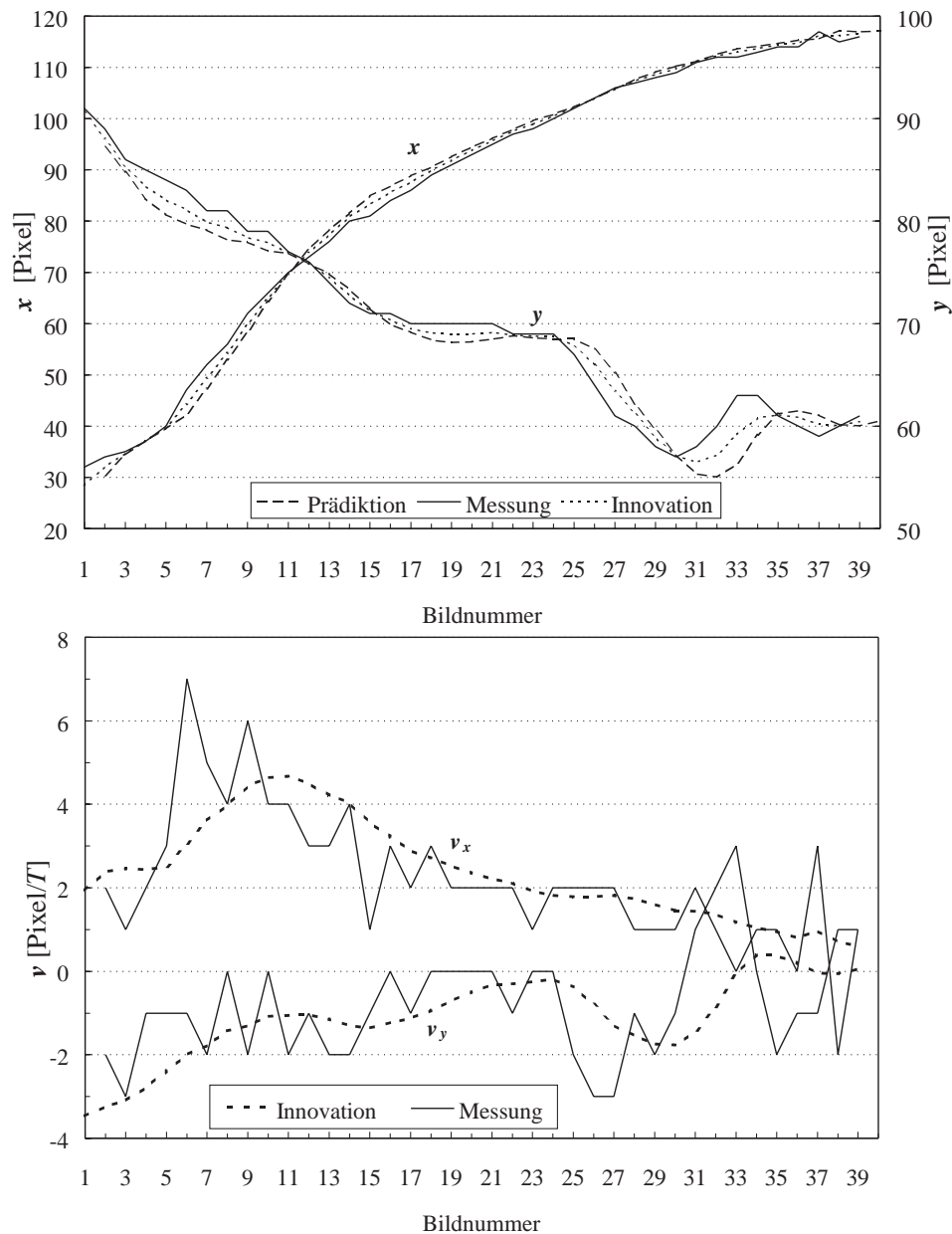


Bild 3.10: Exemplarische Darstellung der Kalmanfiltergrößen bei der Verfolgung eines überholenden Fahrzeugs. Das obere Diagramm zeigt die Verläufe der vorhergesagten, gemessenen und geglätteten Schwerpunktpositionen. Im unteren Diagramm ist der kalmangefilterte Geschwindigkeitsverlauf und der Verlauf der gemessenen Schwerpunktverschiebungen $\frac{\xi(k) - \xi(k-1)}{T}$ dargestellt.

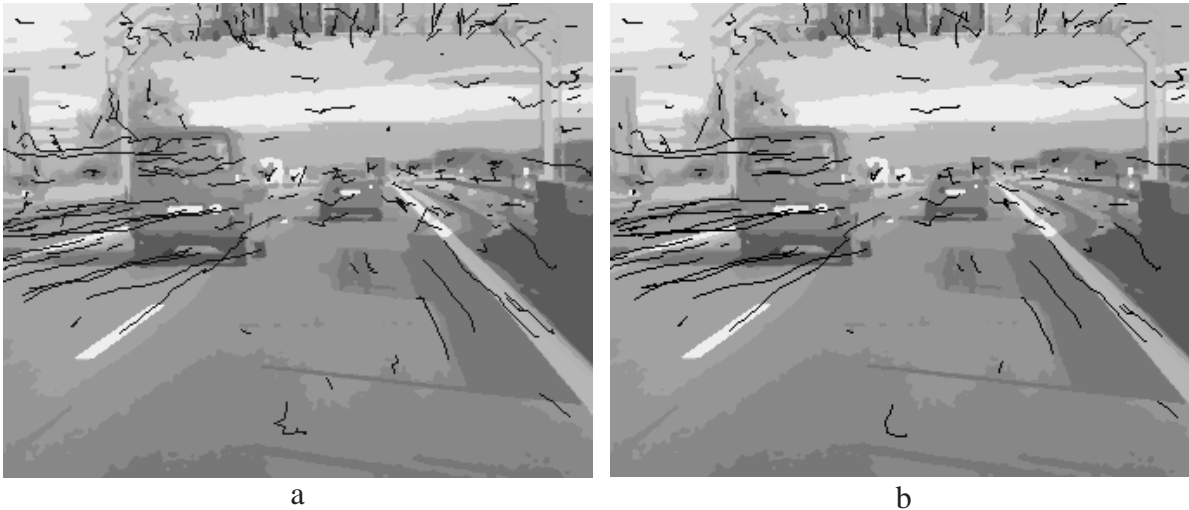


Bild 3.11: Beispiel für die Kalmanfilterung von Regionenbewegungen. In a) sind die aus den segmentierten Bildern bestimmten Trajektorien der Flächenschwerpunkte dargestellt; in b) die durch Kalmanfilterung geglätteten Trajektorien.

3.3 Objektdetektion durch Bewegungssegmentierung

Die Aufgabe der Bewegungssegmentierung im Farbregionenfluß ist die Detektion bewegter Objekte in farbsegmentierten Bildfolgen. Im Unterschied zu den in Kapitel 2.3 vorgestellten Verfahren, bestehen die Eingangsdaten der Bewegungssegmentierung nicht aus optischen Flußvektoren, sondern aus Trajektorien von Flächenschwerpunkten, die sich jeweils aus einem Satz aneinandergeketteter Verschiebungsvektoren zusammensetzen. Während der optische Fluß die momentane Bildbewegung wiedergibt, beinhalten Trajektorien Bewegungsinformation über einen gewissen Zeitraum und ermöglichen somit eine differenziertere Objekttrennung in der Bewegungssegmentierung. Eine weitere Besonderheit der Bewegungssegmentierung im Farbregionenfluß ist die Gruppierung von Regionen, statt der üblichen Gruppierung von Bildpunkten. Der Vorteil dieser Vorgehensweise ist eine deutliche Vereinfachung und damit Beschleunigung des Gruppierungsvorgangs. Der Zeitaufwand für die implementierte Bewegungssegmentierung lag auf einer Sun SPARC-20 unter 10 ms pro Bild.

In der Arbeit müssen die Objekte folgende Voraussetzungen erfüllen, damit sie detektiert werden können:

1. Ein Objekt muß aus mindestens zwei Regionen bestehen. Durch diese Forderung verhindert man, daß einzelne Regionen, deren Trajektorien fehlerhaft bestimmt wurden, zu falschen Objekthypothesen führen. In den untersuchten Sequenzen hat sich gezeigt, daß Objekte, die zur Gruppe der Verkehrsteilnehmer gehören, zumeist in mehr als zwei Farbregionen aufgeteilt werden.
2. Ein Objekt muß sich aus Regionen zusammensetzen, deren umschreibende Rechtecke sich überlappen. Die Oberfläche eines kompakten Szenenobjekts wird auf *eine* Fläche aus zusammenhängenden Bildpunkten projiziert. Damit müssen Regionen, die zum selben Objekt gehören, aneinandergrenzen. Durch Untersegmentierung oder durchsichtige Objektflächen (z.B. Fahrzeugscheiben) kann es jedoch vorkommen, daß ein Objektteil nicht vom Hintergrund getrennt wird – es können Lücken zwischen den verbleibenden Farbregionen eines Objektes entstehen. Um in solchen Fällen trotzdem eine Gruppierung der Farbregionen zu ermöglichen, wird die Forderung nach gemeinsamen Regionengrenzen abgeschwächt und lediglich eine Überlappung der umschreibenden Rechtecke der Regionen gefordert.
3. Die Objekte müssen eine signifikante Bewegung in der Bildebene aufweisen. Es wird gefordert, daß die Trajektorien der Farbregionen eine gegebene Mindestlänge überschreiten. Bei stehender oder langsam bewegter Kamera können damit die Regionen bewegter Szenenobjekte von denen der unbewegten Szenenteile getrennt werden. Bei Eigenbewegung der Kamera wird die Forderung nur von Objekten erfüllt, die eine signifikante Relativbewegung zur Kamera besitzen. In allen Untersuchungen wurde eine minimale Schwerpunktverschiebung von 9 Bildpunkten innerhalb der letzten drei Bilder gefordert. Dies entspricht einer mittleren minimalen Bild-zu-Bild Verschiebung von 3 Bildpunkten.
4. Die Trajektorien der Regionen eines Objektes müssen sich in ihren Verläufen ähneln. Diese für die Bewegungssegmentierung fundamentale Annahme leitet sich aus der Tatsache ab, daß die Punkte eines Szenenobjektes physikalisch gekoppelt sind und somit ihre Bewegungen voneinander abhängen. Die Ähnlichkeit der Bewegungen zweier Regionen A und B wird durch die Korrelation ihrer Trajektorien bestimmt:

$$\rho(A, B) = \left(1 - \frac{|l_A - l_B|}{l_A + l_B} \right). \quad (3.55)$$

$$\frac{\sum_{n=M}^N (\mathbf{x}_A(n) - \bar{\mathbf{x}}_A)^T (\mathbf{x}_B(n) - \bar{\mathbf{x}}_B)}{\sqrt{\sum_{n=M}^N (\mathbf{x}_A(n) - \bar{\mathbf{x}}_A)^T (\mathbf{x}_A(n) - \bar{\mathbf{x}}_A) \sum_{n=M}^N (\mathbf{x}_B(n) - \bar{\mathbf{x}}_B)^T (\mathbf{x}_B(n) - \bar{\mathbf{x}}_B)}}$$

wobei

$$\bar{\mathbf{x}} = \frac{1}{1 + N - M} \sum_{n=M}^N \mathbf{x}(n) \quad (3.56)$$

$$l = \sum_{n=M}^{N-1} \sqrt{(\mathbf{x}(n+1) - \mathbf{x}(n))^T (\mathbf{x}(n+1) - \mathbf{x}(n))} \quad (3.57)$$

Dabei ist $\mathbf{x}(n) = (x(n), y(n))^T$ die Schwerpunktposition einer Region im Bild n . Das Intervall $[M, N]$ gibt den Ausschnitt aus der Bildsequenz vor, der zum Vergleich der Trajektorien verwendet wird. Der erste Faktor in Gl. (3.55) berücksichtigt Abweichungen in der Länge der Trajektorien. Der zweite Faktor ist die lineare Korrelation der Trajektorien. Sie wird 1, wenn die Trajektorien bis auf eine Verschiebung oder einen Skalierungsfaktor gleich sind. Bei Gegenläufigkeit der Trajektorien wird sie zu -1 . Eine Korrelation von 0 ergibt sich, wenn die Verschiebungsvektoren der Regionen zu jedem Zeitpunkt orthogonal zueinander verlaufen. Um zu entscheiden, ob zwei Regionen A und B zum selben Objekt gehören, wird geprüft, ob ρ einen gegebenen Schwellwert überschreitet.

In Bild 3.12 ist die Bewegungssegmentierung schematisch dargestellt. Eine detaillierte Übersicht über die einzelnen Schritte ist dem Struktogramm in Anhang A.2 zu entnehmen.

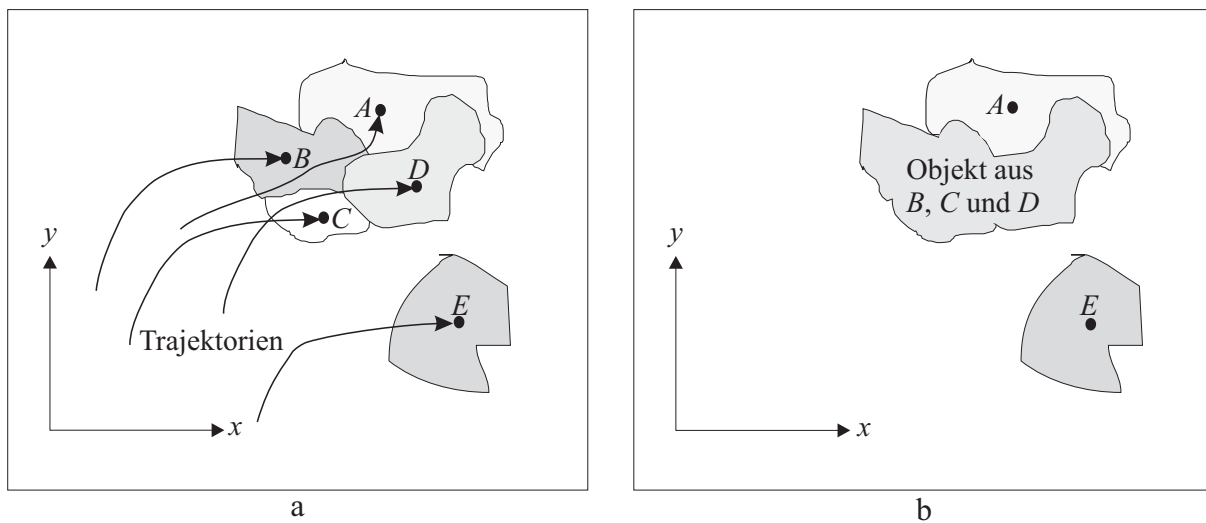


Bild 3.12: In der Bewegungssegmentierung werden Regionen, die ähnliche Trajektorien besitzen und benachbart sind, zu Objekthypothesen gruppiert. In a) sind die Regionen mit ihren Trajektorien dargestellt, Bild b) zeigt das Ergebnis der Bewegungssegmentierung. Die Regionen B , C , und D werden aufgrund der Ähnlichkeit ihrer Trajektorien zu einer Objekthypothese gruppiert. Region E besitzt zwar eine ähnliche Trajektorie wie die Regionen B , C und D , liegt von diesen aber zu weit entfernt.

3.4 Ergebnisse der Objektdetektion anhand des Farbregionenflusses

Die Bewegungssegmentierung wurde an vier verschiedenen Sequenzen untersucht:

1. Diese Aufnahme erfolgte mit einer senkrecht zum Straßenrand ausgerichteten, stehenden Kamera. In der Szene fährt ein Motorradfahrer mit einer Geschwindigkeit von ca. 30 km/h quer zur optischen Achse der Kamera (siehe Bild 3.13 a)). Die Sequenz umfaßt 65 Bilder.
2. Die aus 48 Bildern bestehende Sequenz wurde bei der Anfahrt auf eine Kreuzung aufgenommen. Die Eigengeschwindigkeit betrug etwa 20 km/h. In der Szene ist ein Auto zu sehen, das die Kreuzung quer zur eigenen Fahrtrichtung passiert (siehe Bild 3.13 b)).
3. Die 123 Bilder umfassende Sequenz wurde auf der Autobahn bei bewölktem Himmel aufgenommen. Während der Aufnahme wurde das Versuchsfahrzeug von mehreren PKWs überholt (siehe Bild 3.13 c)). Die Eigengeschwindigkeit betrug etwa 80 km/h.
4. Die letzte Sequenz ist die bereits erwähnte Referenzsequenz. Sie umfaßt 230 Bilder und wurde bei strahlendem Sonnenschein auf der Autobahn aufgenommen. Zu sehen sind mehrere überholende Fahrzeuge (siehe Bild 3.13 d)). Die Eigengeschwindigkeit betrug etwa 80 km/h.

Die Bilder aller Sequenzen wurden auf 16 Farben quantisiert und anschließend durch Relaxation mit zeitlicher Glättung nachbearbeitet. Zur Bewegungssegmentierung wurden die in der Innovationsstufe des Kalmanfilters berechneten Trajektorien verwendet. Alle Fahrzeuge in den Testsequenzen, auch solche mit geringer Relativbewegung zur Kamera, wurden manuell durch ihre umschreibenden Rechtecke markiert, die im folgenden als Referenzrechtecke bezeichnet werden. Zur Bewertung der Detektionsleistung wurden folgende Größen bestimmt:

- Die Anzahl der korrekt detektierten Objekte. Eine Detektion wird als korrekt bewertet, wenn sich Bildpunkte des detektierten Objekts in einem der Referenzrechtecke befinden.
- Die Anzahl der nicht detektierten Objekte. Ein Objekt wird als nicht detektiert gewertet, wenn die Überlappungsgrade U_P der Objekthypothesen mit dem jeweiligen Referenzrechteck alle kleiner als 10% sind.

- Die Anzahl der Falschdetektionen. Eine Detektion wird als falsch bewertet, wenn sich kein Bildpunkt der Objekthypothesen in einem der Referenzrechtecke befindet.
- Der Überlappungsgrad in bezug auf Bildpunkte. Er ist ein Maß dafür, wie stark sich die Regionen in der Objekthypothese mit dem Referenzrechteck überlappen:

$$U_P = 100 \cdot \frac{\text{Größe}(\mathcal{O} \cap \mathcal{R})}{\text{Größe}(\mathcal{O} \cup \mathcal{R})} \quad (3.58)$$

dabei ist \mathcal{O} die Menge aller Punkte in der Objekthypothese und \mathcal{R} die Menge aller Punkte im Referenzrechteck.

- Der Überlappungsgrad in bezug auf das umschreibende Rechteck. Die meisten Objekte füllen nur einen Teil des Referenzrechtecks aus. Zum Beispiel wird bei Seitenansichten von Motorrädern nur etwa 50% des Referenzrechtecks durch das Objekt ausgefüllt. In solchen Fällen ist der Überlappungsgrad U_B des umschreibenden Rechtecks der Objekthypothese mit dem Referenzrechteck ein besseres Maß für die Detektionsleistung als die oben genannte Größe U_P .

$$U_B = 100 \cdot \frac{\text{Größe}(\tilde{\mathcal{O}} \cap \mathcal{R})}{\text{Größe}(\tilde{\mathcal{O}} \cup \mathcal{R})} \quad (3.59)$$

$\tilde{\mathcal{O}}$ ist die Menge aller Punkte im umschreibenden Rechteck der Objekthypothese und \mathcal{R} die Menge aller Punkte im Referenzrechteck.

- Die Anzahl von Verschmelzungen. Eine Verschmelzung liegt vor, wenn verschiedene Objekte oder Teile verschiedener Objekte in derselben Objekthypothese zusammengefaßt werden.

Tabelle 3.8: Ergebnisse der Objektdetektion mit dem Farbregionenfluß

Sequenz	Richtig	Nicht	Falsch	U_P	U_B	Verschm.
1	62	0	0	34	85	0
2	40	5	22	46	95	0
3	158	35	35	32	45	31
4	273	154	72	31	48	13

Da der Farbregionenfluß zur Objektdetektion Informationen aus der Bildfolge benötigt, wurden die ersten drei Bilder jeder Sequenz von der Bewertung der Detektionsleistung ausgeschlossen. In Bild 3.13 ist von jeder Szene ein Beispiel für die Objektdetektion dargestellt. Die numerischen Ergebnisse für die einzelnen Sequenzen sind in Tabelle 3.8 angegeben. Keine Probleme ergaben sich bei der Detektion des Motorradfahrers in der ersten

Sequenz. Auch bei bewegter Kamera und quer zur optischen Achse sich bewegender Fahrzeuge in der zweiten Sequenz ist die Detektionsleistung hervorragend. Da das Fahrzeug zu Beginn der Szene aus dem Stand heraus langsam beschleunigte, konnte es in den ersten 7 Bildern aufgrund seiner geringen Bildbewegung nicht detektiert werden. Die Falschdetektionen sind, wie in Bild 3.14 a) zu erkennen ist, auf die Eigenbewegung der Kamera zurückzuführen. In der dritten Sequenz kommt es zu Situationen, in denen sich Autos mit annähernd gleicher Geschwindigkeit teilweise verdecken. Die Fahrzeuge lassen sich dann durch die Bewegungssegmentierung nicht mehr voneinander trennen. Dies schlägt sich in der Anzahl der Verschmelzungen und einem geringen Überdeckungsgrad U_B nieder. In Bild 3.14 b) ist ein solcher Fall dargestellt. Fahrzeuge mit ähnlicher Geschwindigkeit wie das Versuchsfahrzeug weisen in der Bildebene keine bzw. nur eine geringe Bewegung auf. Sie werden demzufolge auch nicht detektiert. In Sequenz drei und vier lag die Bild-zu-Bild Verschiebung der Referenzrechtecke 35 mal bzw. 106 mal unterhalb der Detektionsschwelle von 3 Bildpunkten. Dies ist der Grund für die hohe Anzahl der nicht detektierten Objekte in diesen beiden Sequenzen. Der geringe Überdeckungsgrad U_B in der dritten Sequenz ist auf die Schlagschatten der Fahrzeuge zurückzuführen. Sie werden, da keine Modellannahmen über das Erscheinungsbild der Objekte verwendet werden, in die Objekthypothesen aufgenommen (siehe Bild 3.14 d)). Ein weiterer Problemfall aus Szene vier ist in Bild 3.14 c) zu sehen. Der weiße PKW, ein Teil der Leitplanken und die beiden entgegenkommenden LKWs werden in einer Farbregion zusammengefaßt.

Im folgenden werden Stärken und Schwächen des Verfahrens zusammengefaßt:

- + Das Verfahren benötigt keine kalibrierte Kamera, kein Straßendmodell, keine Messung der Eigengeschwindigkeit und keine Kenntnisse über das Aussehen der zu detektierenden Objekte.
- + Objekte mit signifikanter Relativbewegung zur Kamera lassen sich zuverlässig detektieren.
- + Die Objekthypothesen liefern eine gute Beschreibung der Objekte. In [Kolb 95] wurde eine Klassifikation des Fahrzeugtyps (PKW, LKW, Motorrad) bei seitlich vorbeifahrenden Fahrzeugen implementiert. Zur Detektion der Objekte und Bestimmung der Objektkonturen wurde der Farbregionenfluß verwendet.
- + Der Rechenaufwand ist für ein Verfahren der Farbbildfolgenanalyse relativ klein. Auf einem PC mit einem zwei-Prozessor System konnte eine Bildwiederholrate von 6 Bildern pro Sekunde, bei Verwendung der Medianfilterung anstelle der Relaxation, erzielt werden.
- Es lassen sich nur Objekte detektieren, die sich relativ zur Kamera bewegen.

- Objekte, die in der Bildebene aneinandergrenzen und dieselbe Bildbewegung besitzen, können nicht voneinander getrennt werden. Genausowenig kann zwischen dem Schatten und dem Objekt selbst unterschieden werden.
- In der Farbsegmentierung kommt es vor, daß verschiedene Objekte ähnlicher Farbe in einer Farbregion zusammengefaßt werden. Insbesondere bei grau gefärbten Fahrzeugen wurden öfters Teile der Straße und des Objekts zu einer Farbregion verschmolzen.
- Trotz lokaler und zeitlicher Glättung in der Relaxation, verursacht das Meßrauschen der Kamera z.T. erhebliche Änderungen in den Segmentierungsergebnissen von Folgebildern. Betroffen sind insbesondere Regionen, die nur einen schwachen Farbkontrast gegenüber ihren Nachbarregionen besitzen. Auch schmale Regionen, wie sie z.B. bei Armen und Beinen von Personen entstehen, erweisen sich häufig als instabil. Dort genügt bereits die Farbänderung weniger Bildpunkte, um eine Region auseinanderzubrechen bzw. zwei Regionen miteinander zu verschmelzen.

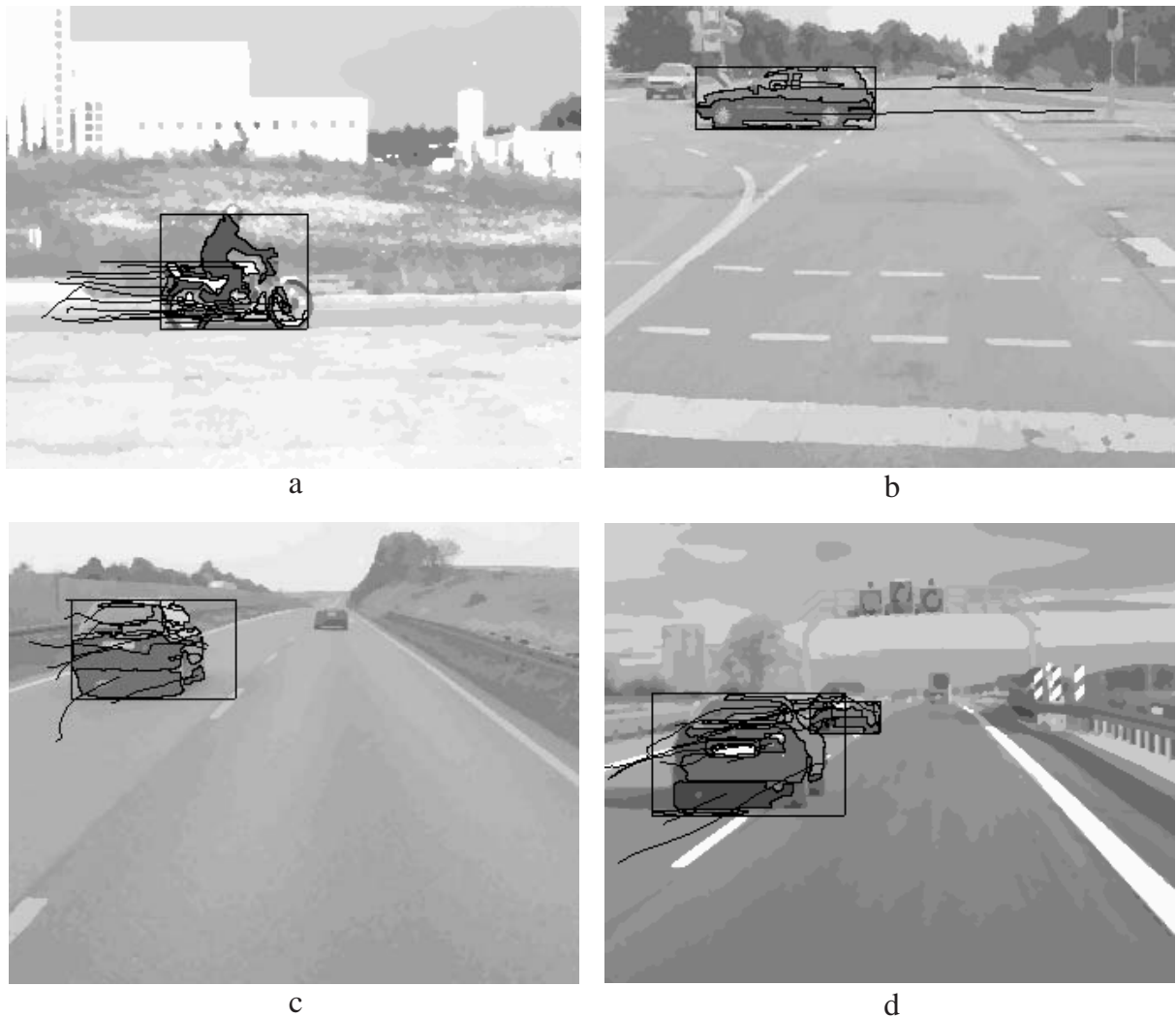


Bild 3.13: Beispiele der Objektdetektion durch den Farbregionenfluß. Die Trajektorien und Konturen der Regionen sowie die umschreibenden Rechtecke der Objekt-hypothesen sind schwarz hervorgehoben. Die Motorradszene in a) wurde mit stehender Kamera aufgenommen. Bild b) entstand bei der Anfahrt auf eine Kreuzung. Die Bilder der unteren Reihe sind zwei verschiedenen Autobahnsequenzen entnommen. Bei Aufnahme c) war der Himmel bewölkt, bei d) herrschte strahlender Sonnenschein.

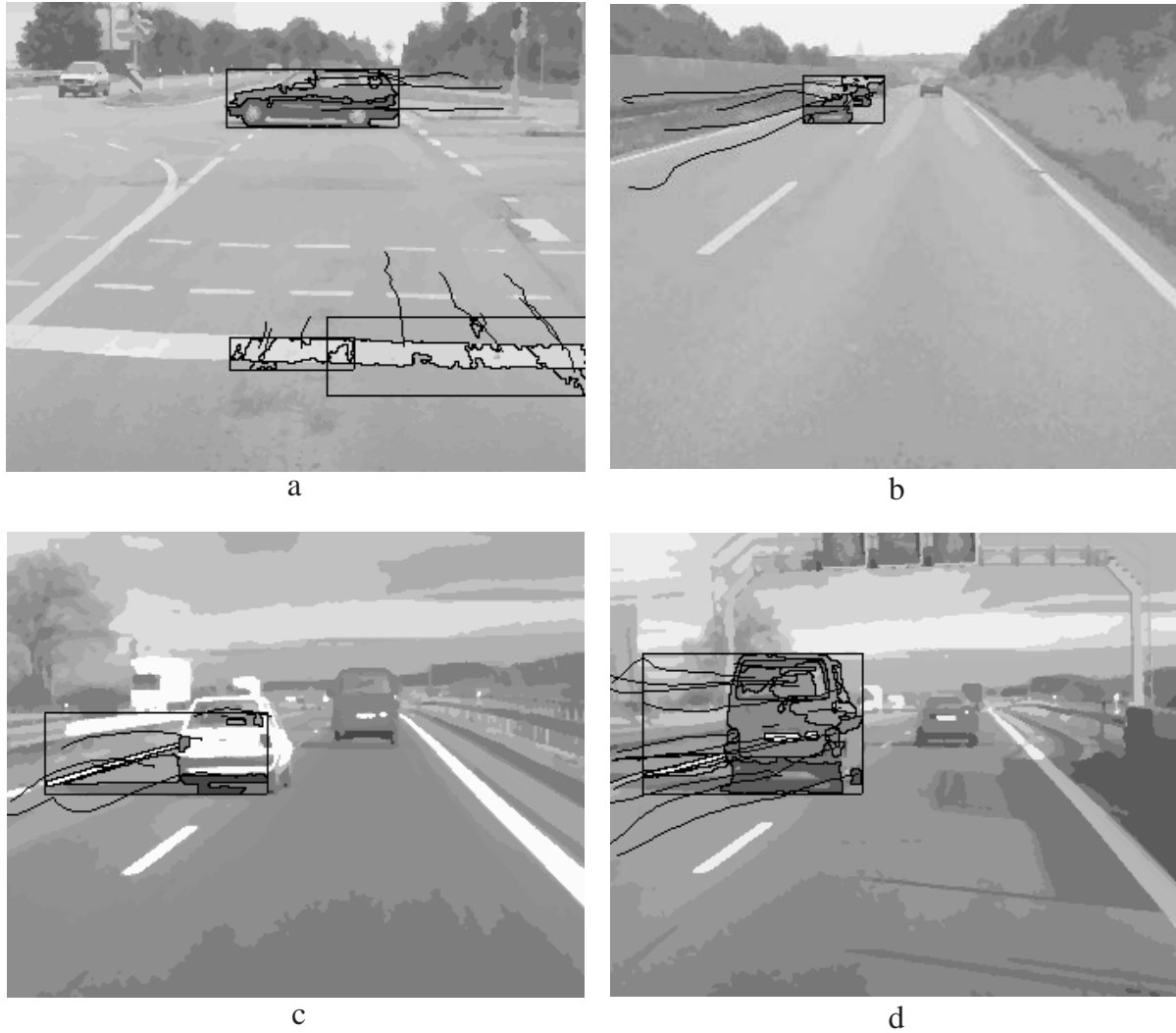


Bild 3.14: Problemfälle bei der Objektdetektion durch den Farbregionenfluß. Die Trajektorien und Konturen der Regionen sowie die umschreibenden Rechtecke der Objekthypothesen sind schwarz hervorgehoben. Die falschen Detektionen in a) sind auf die Eigenbewegung der Kamera zurückzuführen. In b) konnten die drei Fahrzeuge auf der linken Fahrbahn nicht mehr voneinander getrennt werden. In c) wurde bereits in der Farbsegmentierung der weiße PKW, ein Teil der Leitplanken und die beiden entgegenkommenden LKWs in einer Farbregion zusammengefaßt. Die Bildbewegung des Minivans auf der rechten Fahrbahnseite lag unter der Detektionsschwelle. In c) und d) werden durch die Schlagschatten der Fahrzeuge Straßenteile in die Objekthypothesen aufgenommen.

4 Farbclusterfluß

Der Farbclusterfluß gliedert sich in drei Schritte, die sich im Aufbau des Kapitels widerspiegeln: der initialen, divisiven Clusteranalyse des ersten Bildes einer Sequenz, der iterativen Clusteranalyse von Folgebildern und schließlich der Bewegungssegmentierung. Das Kapitel wird durch Ergebnisse der Objektdetektion anhand des Farbclusterflusses abgeschlossen. Bevor jedoch auf die Verfahrensschritte eingegangen wird, soll die Motivation für die Entwicklung einer Alternative zum bereits vorgestellten Farbregionenfluß erläutert werden.

4.1 Motivation

Mit dem Farbregionenfluß wurde ein regionenbasiertes Verfahren zur Bewegungsbestimmung in Bildfolgen entwickelt, das bei geringem Rechenaufwand eine flächendeckende Bewegungsbestimmung ermöglicht. Als Problem hat sich jedoch die zeitliche Instabilität der Farbregionen erwiesen. Eine Glättung der Daten über die Bildfolge konnte zwar auf der Ebene von Bildpunkten realisiert werden, nicht aber auf der Ebene von Regionen. Die Kenntnis, daß eine Region, die stets dasselbe Szenenobjekt repräsentiert, in der Regel nur geringfügigen Bild-zu-Bild Änderungen unterliegt, konnte in der Segmentierung nicht unmittelbar genutzt werden. Man fragt sich, wie eine dynamische Segmentierung aussehen könnte, die eine zeitliche Konsistenz der Regionen impliziert. Aus der Literatur ist zu diesem Thema das in [Bhanu & Burger 88] genannte Regionenwachstumsverfahren zu finden. Als Anfangsgebiete für das Wachstumsverfahren im aktuellen Bild werden Teile der Regionen des vorhergehenden Bildes verwendet. Aufgrund seiner Komplexität und dem damit verbundenen Rechenaufwand scheidet das Verfahren für die vorliegende Anwendung aus. Bei dem in dieser Arbeit entwickelten Verfahren, dem Farbclusterfluß [Heisele et al. 97], macht man sich die Eigenschaften des in Kapitel 3.1.3 erwähnten parallelen k-means Clusterverfahrens zu nutze. Dieses Clusterverfahren führt, ausgehend von einer Anfangspartitionierung einer Datenmenge, eine iterative Optimierung der Partitionierung in bezug auf ein gegebenes Fehlermaß durch. Wenn sich die Farbsegmentierung durch ei-

ne Clusteranalyse realisieren ließe, wäre somit das parallele k-means Verfahren ideal zur dynamischen Segmentierung geeignet. Die Partitionierung bzw. Segmentierung des vorhergehenden Bildes könnte dann zur Initialisierung der Clusteranalyse des aktuellen Bildes herangezogen werden. Eine zeitliche Konsistenz der Cluster wäre somit garantiert. Es muß also noch geklärt werden, wie durch Clusteranalyse eine sinnvolle Aufteilung von Farbbildern in flächenhafte Bildelemente erreicht werden kann. In der klassischen Farbsegmentierung nimmt man an, daß Punkte ähnlicher Farbe, die eine geschlossene Fläche bilden, zum selben Bildobjekt gehören und damit in einer Farbregion zusammengefaßt werden können. Für die Clusteranalyse wird diese Annahme modifiziert; sie lautet jetzt: Bildpunkte, die sich in ihrer Farbe und ihrer Position in der Bildebene ähnlich sind, gehören zum selben Bildobjekt. Legt man diese Annahme für die zu untersuchenden Bilddaten zugrunde, läßt sich durch eine Clusteranalyse im Farb-Positionsraum eine sinnvolle Aufteilung des Bildes vornehmen. Jeder Bildpunkt wird dabei durch fünf Merkmale charakterisiert: die drei Farbwerte R , G und B sowie die horizontale und vertikale Rasterkoordinate in der Bildebene x und y . Da Farbe und Position unterschiedliche Größen mit unterschiedlichen Wertebereichen sind, ist es sinnvoll, einen konstanten Gewichtungsfaktor w einzuführen. Ein Bildpunkt ist dann durch den Merkmalsvektor $\mathbf{s} = (R, G, B, w \cdot x, w \cdot y)^T$ beschrieben. Die Aufgabe der Clusteranalyse besteht nun darin, für ein Bild aus N Bildpunkten, von denen jeder durch einen Merkmalsvektor beschrieben ist, einen Satz von Q Referenzvektoren $\mathbf{r}_1, \dots, \mathbf{r}_Q$ zu bestimmen, so daß der mittlere quadratische Abstand

$$MSE = \frac{1}{N} \sum_{i=1}^N \|\mathbf{s}_i - \mathbf{r}_{q(\mathbf{s}_i)}\|^2 \quad (4.1)$$

minimiert wird. Dabei ist $\mathbf{r}_{q(\mathbf{s}_i)}$ der Referenzvektor des Clusters, in dem der Merkmalsvektor \mathbf{s}_i liegt. Durch die Lage der Referenzvektoren ist auch die Partitionierung festgelegt. Für das Cluster \mathcal{C}_q mit dem Referenzvektor \mathbf{r}_q gilt:

$$\mathcal{C}_q = \{\mathbf{s}_n \mid \|\mathbf{s}_n - \mathbf{r}_q\|^2 \leq \|\mathbf{s}_n - \mathbf{r}_i\|^2 \forall i = 1, \dots, Q\} \quad (4.2)$$

Man beachte den Unterschied in der Wahl der Merkmale im Vergleich zu der in Kapitel 3.1.3 beschriebenen Farbquantisierung. Dort ging es lediglich darum, die im Bild auftretenden signifikanten Farben zu bestimmen – der Merkmalsvektor eines Bildpunktes bestand somit nur aus den drei Farbwerten. Erst in der Zusammenhangsanalyse wurde dann auch die Position eines Bildpunktes bei der Generierung der Farbregionen berücksichtigt.

4.2 Initiale Clusteranalyse

4.2.1 Eigenvektor Verfahren

Im vorigen Abschnitt wurde das parallele k-means Verfahren zur dynamischen Segmentierung von Farbbildern vorgeschlagen. Da es eine Initialisierung in Form einer Anfangspartitionierung der Daten erfordert, ist es für die Segmentierung des ersten Bildes einer Sequenz ungeeignet. Bei einer schlecht gewählten Anfangspartitionierung ist eine erhebliche Anzahl von Iterationen nötig, bis eine akzeptable Segmentierung erreicht ist. Effizienter, im Hinblick auf Rechenzeit, sind die in Kapitel 3.1.3 vorgestellten divisiven Clustertechniken. Von den drei dort genannten Verfahren wurde das Eigenvektor Verfahren gewählt, da es ohne Modifikationen auf beliebig dimensionierte Datensätze anwendbar ist. Das Eigenvektor Verfahren führt eine schrittweise Unterteilung des Merkmalsraums in Cluster durch, bis die vorgegebene Clusteranzahl Q erreicht ist. Begonnen wird mit einem Cluster, das die gesamten Merkmalsvektoren beinhaltet. In jedem Schritt des Verfahrens wird dann das Cluster mit der höchsten Streuung der darin enthaltenen Merkmalsvektoren durch eine Trennebene aufgeteilt. Sie wird so gewählt, daß sie senkrecht zur Richtung maximaler Streuung und durch den Schwerpunkt des Clusters verläuft. In den Untersuchungen wurde mit Werten von Q zwischen 64 und 128 gearbeitet. Die Rechenzeit auf einer Sun SPARC-20 lag für 64 Cluster im Mittel bei 5 s und für 128 Cluster bei etwa 7 s.

4.2.2 Gewichtung zwischen Farbe und Position

Entscheidend im Farbclusterfluß ist die Gewichtung der Farbe gegenüber der Position. Sie muß, wie die Clusteranzahl Q , zu Beginn der initialen Clusteranalyse festgelegt werden und wird dann unverändert für die Clusteranalyse der Folgebilder beibehalten. Die Gewichtung tritt in Form eines Faktors w in den Merkmalsvektoren der Bildpunkte auf: $(R, G, B, w \cdot x, w \cdot y)^T$. Der Einfluß von w auf die Segmentierung wird aus Bild 4.1 deutlich. Dargestellt sind Ergebnisse der Aufteilung eines Bildes in 64 Partitionen in Abhängigkeit von w . In der linken Spalte wurden die Originalfarben der Bildpunkte durch die Farben der zugehörigen Referenzvektoren ersetzt. Zur Illustration der Clusterform in der Bildebene ist in der rechten Spalte ein Cluster schwarz hervorgehoben. Bei $w = 10$ entstehen kompakte Cluster in der Bildebene, die kaum einen Bezug zur Farbverteilung im Originalbild aufweisen. Eine bessere Beschreibung der Szene erhält man für $w = 2$. Bei Bildobjekten, die einen hohen Farbkontrast zum Hintergrund aufweisen (z.B. Verkehrszeichen und Fahrbahnmarkierungen), ist eine deutliche Übereinstimmung der Clusterformen mit den Objektformen zu erkennen. In großen Bildbereichen, mit annähernd gleicher Färbung, wird die Aufteilung hauptsächlich durch die Position der Bildpunkte bestimmt. Dann

ähnelt die Form der Cluster der von konvexen Polygonen, wie im Bereich der Straße deutlich wird. Bei $w = 0,1$ wird die Clusteranalyse durch die Farbverteilung im Bild dominiert. Im Farbraum hat die Kompaktheit der Cluster zugenommen, wogegen sie in der Bildebene abgenommen hat. Dies kann dazu führen, daß Bildobjekte ähnlicher Farbe in einem Cluster zusammengefaßt werden, obwohl sie in der Bildebene weit voneinander entfernt liegen. In den untersuchten Bildsequenzen hat sich $w = \frac{10}{6}$ als brauchbare Gewichtung erwiesen.

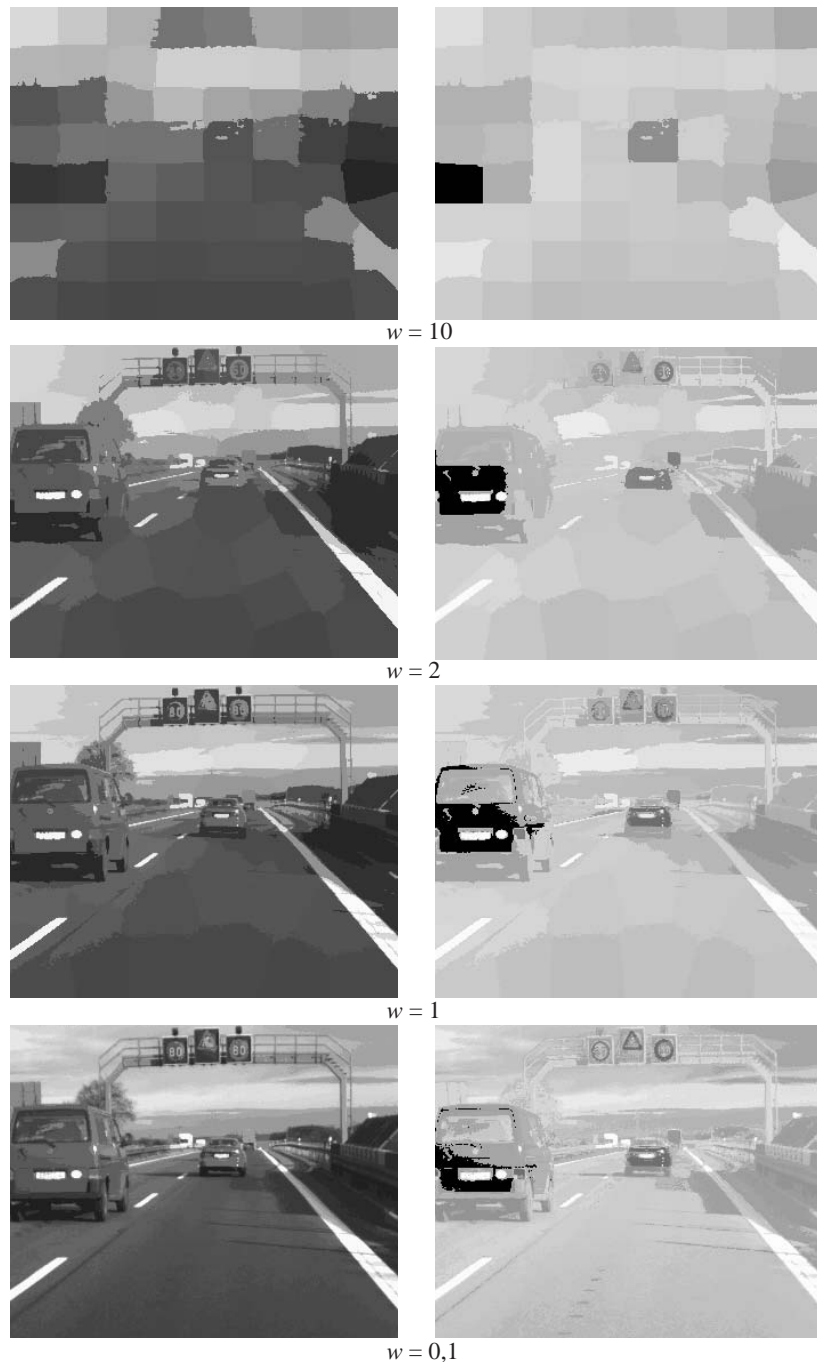


Bild 4.1: Ergebnisse der Clusteranalyse im Farb-Positionsraum in Abhängigkeit der Gewichtung w der Position gegenüber der Farbe. Es wurde ein 360×288 großes Bild mit dem Eigenvektor Verfahren in 64 Cluster aufgeteilt. In der linken Spalte sind die Originalfarben der Bildpunkte durch die Farben der Referenzvektoren ersetzt. Um die Auswirkung von w auf die Clusterform zu verdeutlichen, ist in der rechten Spalte ein am Heck eines Fahrzeugs plaziertes Cluster schwarz hervorgehoben.

4.3 Clusteranalyse von Folgebildern

4.3.1 Paralleles k-means Verfahren

Zur Clusteranalyse von Folgebildern wird das parallele k-means Verfahren eingesetzt. Es führt eine iterative Minimierung des Quantisierungsfehlers durch. Eine Iteration besteht dabei aus zwei Schritten:

- Berechnung der Partitionierungen

$$\mathcal{C}_q(k) = \{\mathbf{s}_n \mid \|\mathbf{s}_n - \mathbf{r}_q(k)\|^2 \leq \|\mathbf{s}_n - \mathbf{r}_i(k)\|^2 \forall i = 1, \dots, Q\} \quad (4.3)$$

- Berechnung der Referenzvektoren

$$\mathbf{r}_q(k) = \frac{1}{S_q} \sum_{\mathbf{s}_i \in \mathcal{C}_q(k)} \mathbf{s}_i \quad (4.4)$$

wobei S_q die Anzahl der Merkmalsvektoren im Cluster \mathcal{C}_q ist.

Um eine zeitliche Konsistenz der Partitionierungen in einer Bildfolge zu erreichen, werden als initiale Referenzvektoren $\mathbf{r}_1(0), \dots, \mathbf{r}_Q(0)$ die Referenzvektoren aus dem vorhergehenden Bild verwendet. Durch diese Initialisierung wird zugleich eine zeitliche Verknüpfung der Cluster in zwei aufeinanderfolgenden Bildern hergestellt. Im Unterschied zum Farbregionenfluß ist somit keine explizite Zuordnungsprozedur erforderlich. Nach erfolgter Initialisierung werden Referenzvektoren und Partitionen iterativ an die Bilddaten adaptiert. Wenn die Referenzvektoren in zwei aufeinanderfolgenden Iterationen gleich bleiben, hat man ein lokales Minimum des Quantisierungsfehlers erreicht. Da in der vorliegenden Anwendung nur wenig Rechenzeit zur Verfügung steht, muß schon früher abgebrochen werden. Bei einer Bildwiederholrate von 25 Hz sind die Änderungen zwischen zwei Folgebildern in der Regel so gering, daß bereits eine Iteration für eine brauchbare Adaption der Referenzvektoren an die aktuellen Bilddaten ausreicht. Einen Eindruck über das Konvergenzverhalten des Verfahrens vermittelt Bild 4.2. Dort ist die mittlere Längenänderung $\bar{l}(k)$ der Referenzvektoren in Abhängigkeit der Iterationen aufgetragen:

$$\bar{l}(k) = \frac{1}{Q} \sum_{i=1}^Q \|\mathbf{r}_q(k) - \mathbf{r}_q(k-1)\| \quad (4.5)$$

Die größte Änderung tritt, wie erwartet, in der ersten Iteration auf; schon in der zweiten Iteration ist sie um etwa die Hälfte kleiner. Ab der dritten Iteration sind die Änderungen für die gegebene Aufgabenstellung bereits ohne Belang. Auch wenn durch die geschickte

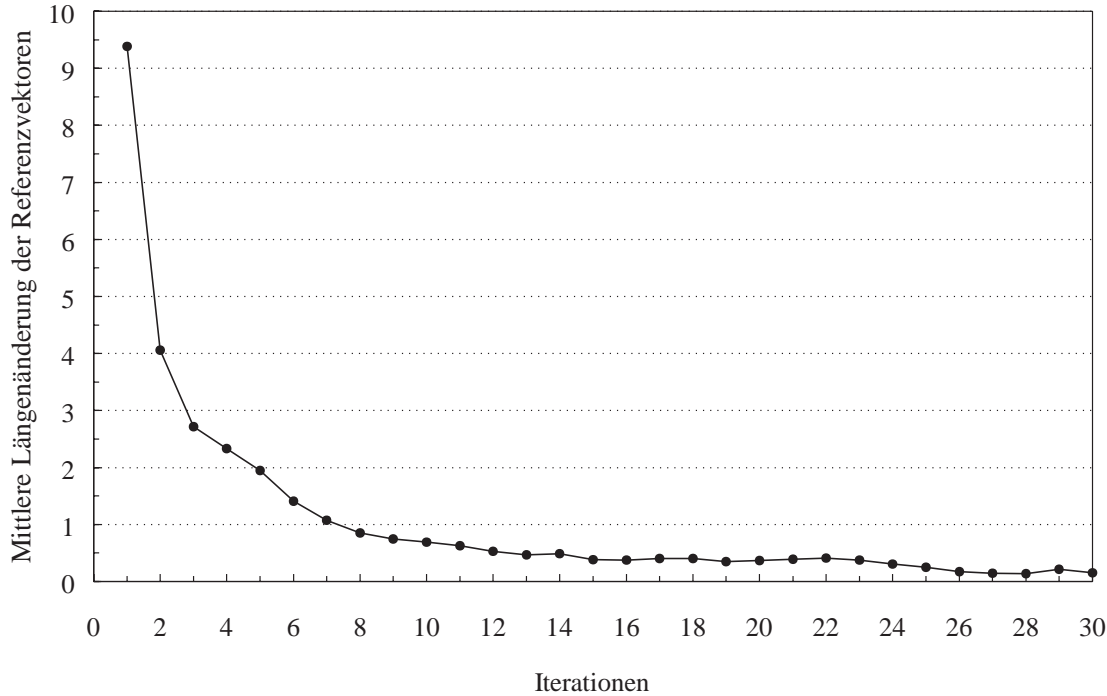


Bild 4.2: Beispiel für die mittlere Längenänderung der Referenzvektoren pro Iteration beim parallelen k-means Verfahren. Zur Initialisierung wurden die Referenzvektoren aus dem vorhergehenden Bild verwendet.

Initialisierung bereits nach einer Iteration abgebrochen werden kann, ist der Rechenaufwand erheblich. Wenn bei der Neuberechnung der Partitionierungen nach Gl. (4.3) für jeden der N Merkmalsvektoren jeweils die Abstände zu allen Q Referenzvektoren berechnet werden, ergibt sich auf einer Sun SPARC-20 eine Rechendauer von etwa 6,7 s pro Iteration bei 64 Partitionen und 12,5 s bei 128 Partitionen. Deshalb wird zur Minderung des Rechenaufwands ein Beschleunigungsverfahren [Judd et al. 96] eingesetzt. Das Prinzip des Verfahrens wird aus Bild 4.3 deutlich; der exakte Verfahrensablauf ist dem Anhang A.3 zu entnehmen. Aus Darstellungsgründen wird in Bild 4.3 ein 2D-Merkmalraum (x_1, x_2) betrachtet. In ihm befinden sich drei Referenzvektoren $\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3$ und drei Merkmalsvektoren $\mathbf{s}_1, \mathbf{s}_2, \mathbf{s}_3$. Um alle Referenzvektoren werden Kreise¹ gelegt, deren Durchmesser den Abständen zu den anderen Referenzvektoren entsprechen. Für \mathbf{r}_1 ergeben sich somit die Kreise K_1 und K_2 . Jedem Merkmalsvektor wird einer der Referenzvektoren als Bezugsvektor zugewiesen. Zu Beginn ist dies jener Referenzvektor, dem der Merkmalsvektor in einer

¹Im 5D-Merkmalraum werden keine Kreise, sondern sog. Hyperkugeln um die Referenzvektoren gelegt.

vorhergehenden Clusteranalyse bzw. Iteration zugeordnet wurde. Im Beispiel sei \mathbf{r}_1 der Bezugsvektor aller drei Merkmalsvektoren. Nun wird geprüft, ob die Merkmalsvektoren innerhalb des kleinsten Kreises um ihren Bezugsvektor liegen. Bei \mathbf{s}_1 ist dies der Fall, er kann ohne weitere Abstandsberechnungen dem ersten Referenzvektor zugeordnet werden. Bei \mathbf{s}_2 und \mathbf{s}_3 muß indes der Abstand zu \mathbf{r}_2 berechnet werden. Da \mathbf{s}_2 näher an \mathbf{r}_2 als an \mathbf{r}_1 liegt, wird der Referenzvektor \mathbf{r}_2 sein neuer Bezugsvektor. Der Zuordnungsprozeß für \mathbf{s}_2 beginnt mit geändertem Bezugsvektor von neuem. Für \mathbf{s}_3 wird geprüft, ob er innerhalb des nächst größeren Kreises um \mathbf{r}_1 liegt. Da dies erfüllt ist, kann er \mathbf{r}_1 zugewiesen werden. Im ungünstigsten Fall liegt ein Merkmalsvektor außerhalb aller Kreise, dann müssen seine Abstände zu allen Referenzvektoren berechnet werden. In Bild 4.4 sind die Rechenzeiten für eine k-means Iteration in Abhängigkeit der Clusteranzahl aufgetragen. Bei Berechnung aller Abstände führt eine Verdopplung der Clusteranzahl etwa zur Verdopplung der Rechenzeit. Durch das Beschleunigungsverfahren erreicht man einen deutlich flacheren Anstieg der Rechenzeitkurve.

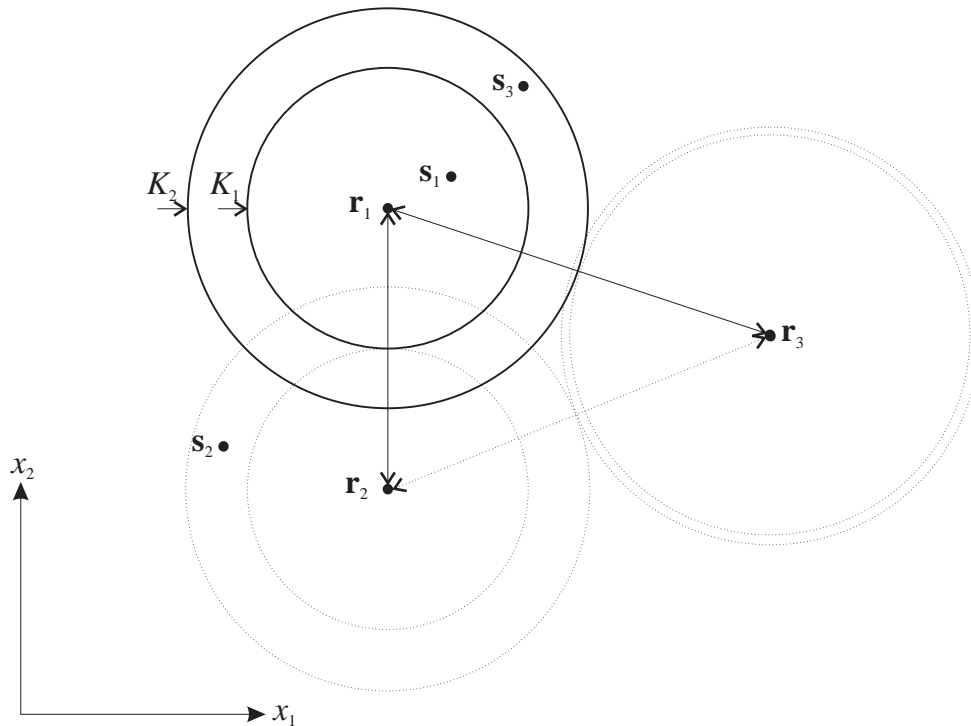


Bild 4.3: Beispiel für die Beschleunigung des parallelen k-means Verfahrens nach [Judd et al. 96] in einem 2D-Merkmalraum. Um jeden Referenzvektor $\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3$ werden konzentrische Kreise gebildet, deren Durchmesser durch die Abstände des jeweiligen Referenzvektors zu den anderen Referenzvektoren gegeben sind. Die Anzahl der notwendigen Abstandsberechnungen hängt davon ab, in welchem Kreis die Merkmalsvektoren $\mathbf{s}_1, \mathbf{s}_2, \mathbf{s}_3$ liegen.

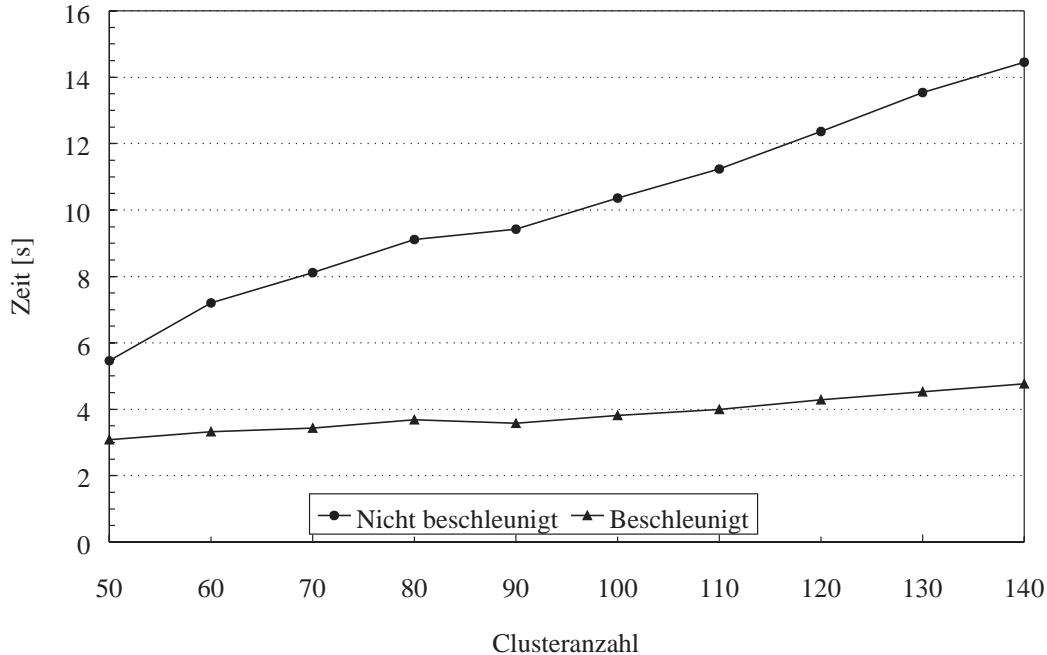


Bild 4.4: CPU-Zeit auf einer Sun SPARC-20 für einen Iterationsschritt des parallelen k-means Verfahrens in Abhängigkeit der Clusteranzahl. Gemessen wurde die Rechenzeit für ein 360×288 großes Bild. Zur Initialisierung des k-means Verfahrens wurden die Referenzvektoren aus dem vorhergehenden Bild verwendet.

4.3.2 Prädiktion der Referenzvektoren

Um eine korrekte Bewegungsbestimmung zu ermöglichen, muß ein Cluster über den betrachteten Zeitraum stets dasselbe Bildobjekt repräsentieren. Ein wichtiger Schritt zur Erfüllung dieser Forderung wurde durch die Einführung einer dynamischen Segmentierung in Form der k-means Clusteranalyse gemacht. Bei der Initialisierung des k-means Verfahrens mit den Referenzvektoren des vorhergehenden Bildes geht man davon aus, daß Farbe und Position eines Bildobjekts in zwei aufeinanderfolgenden Bildern in etwa gleich bleiben. Die durch Szenenbewegung verursachten Abweichungen von dieser initialen Annahme werden durch Adaption der Referenzvektoren in den Iterationen des k-means Verfahrens berücksichtigt. Bei schnell bewegten Bildobjekten sind die Bild-zu-Bild Verschiebungen manchmal jedoch so groß, daß eine korrekte Adaption der Referenzvektoren

nicht mehr möglich ist. In Bild 4.5 a) ist ein solcher Fall dargestellt. Die Trajektorie des Clusters, das zu Anfang der Sequenz den unteren Teil des Kombis darstellt, weicht nach einigen Bildern signifikant von der Trajektorie des Kombis ab. Die Abweichung signalisiert, daß das Cluster nicht mehr dasselbe Bildobjekt wie am Anfang der Bildfolge repräsentiert – eine korrekte Bewegungsbestimmung des Kombis über die gesamte Sequenz ist somit nicht möglich.

Durch Prädiktionsverfahren lassen sich in Bildsequenzen mit schnell bewegten Objekten Verbesserungen erzielen. Man läßt die Annahme über die zeitliche Konstanz der Position eines Bildobjekts fallen und geht statt dessen von einer weitgehend gleichmäßigen Bewegung der Bildobjekte aus. Anhand der Trajektorien der Referenzvektoren in der Bildebene sagt man die Bildpositionen der Referenzvektoren im nächsten Bild vorher. Die Farbwerte der Referenzvektoren bleiben unverändert, da die Farbe eines Objekts von dessen Bewegung prinzipiell unabhängig ist. Die Initialisierung der Clusteranalyse für das nächste Bild erfolgt dann mit den prädizierten Referenzvektoren. In Bild 4.5 b) sind die Trajektorien bei Verwendung prädizierter Referenzvektoren zur Initialisierung der *k*-means Clusteranalyse dargestellt. Offensichtlich bleibt nun ein Cluster während der gesamten Sequenz dem Fahrzeug zugeordnet. Als Prädiktionsverfahren kommt der auf den Seiten 51 ff. beschriebene Kalmanfilter zum Einsatz. Zur Modellierung der Bildbewegung wird das auf Seite 53 beschriebene Systemmodell verwendet, das sich bereits im Farbregionenfluß bewährt hat. In ihm wird die Beschleunigung einer eindimensionalen Bewegung als weißes Rauschen modelliert. Die Initialisierung des Kalmanfilters für die Bewegung in *x*-Richtung wird unverändert vom Farbregionenfluß übernommen (siehe Anhang B.1). Bei der Initialisierung des Kalmanfilters für die *y*-Richtung wird die Anfangsgeschwindigkeit zu Null gewählt, da die Bewegungen der zu detektierenden Objekte hauptsächlich in *x*-Richtung verlaufen. In Untersuchungen hat sich gezeigt, daß durch diese Maßnahme störende Verschiebungen der Referenzvektoren in *y*-Richtung, die keinen Bezug zu realen Bildbewegungen haben, erfolgreich unterdrückt werden können.

Bleibt anzumerken, daß die Prädiktion im Farbclusterfluß einen stärkeren Einfluß auf die Bewegungsbestimmung hat als beim Farbregionenfluß. Dort erfolgte die Segmentierung der Einzelbilder unabhängig von den prädizierten Regionenschwerpunkten. Erst in der Regionenzuordnung wurde die Prädiktion berücksichtigt. Im Farbclusterfluß hingegen wird durch die Initialisierung der Clusteranalyse sowohl die Segmentierung, als auch die Clusterzuordnung beeinflusst. Dieser Umstand birgt die Gefahr, daß die berechneten Trajektorien von den realen Bildbewegungen abweichen, ohne daß sich dies in der Differenz zwischen Prädiktion und Messung bemerkbar macht.

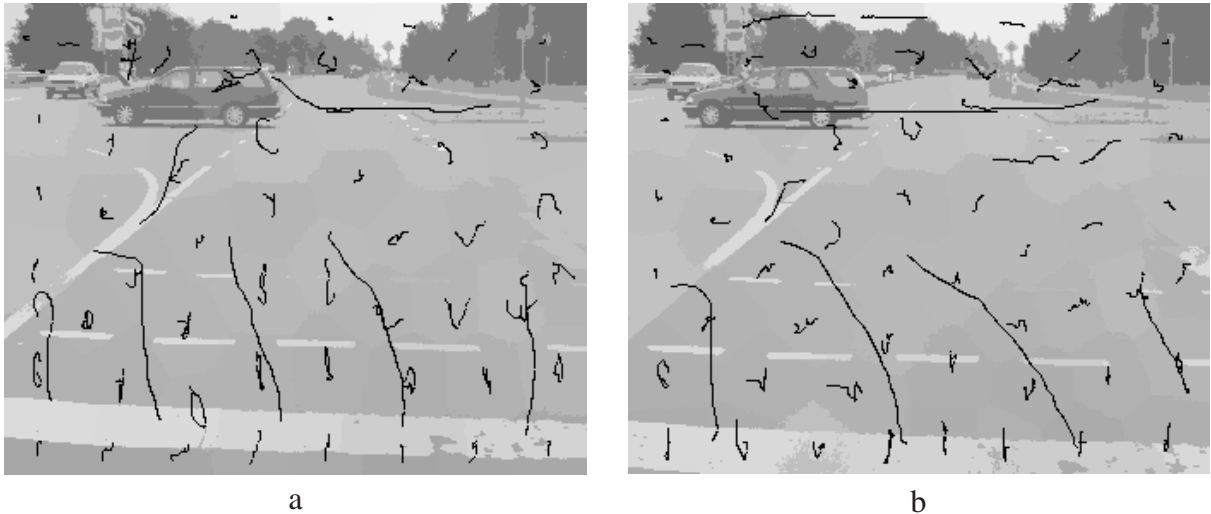


Bild 4.5: Trajektorien der Referenzvektoren (Clusterschwerpunkte) in der Bildebene beim Farbclusterfluß. In Bild a) sind die Ergebnisse ohne Prädiktion der Referenzvektoren dargestellt. Die Trajektorie des Clusters, das zu Beginn der Sequenz die untere Hälfte des Kombis darstellt, weicht nach einigen Bildern deutlich von der Trajektorie des Fahrzeugs ab. In b) erfolgte die Prädiktion der Bildpositionen der Referenzvektoren durch Kalmanfilterung. Bei dem Fahrzeug stimmt jetzt die berechnete Trajektorie deutlich besser mit der realen Bildbewegung des Fahrzeugs überein.

4.4 Bewegungssegmentierung

Die aus einer Bildfolge extrahierte Bewegungsinformation liegt in Form der Trajektorien der Referenzvektoren vor. Die Bewegungssegmentierung faßt nun Cluster, die in der Bildebene benachbart liegen und ähnliche Trajektorien besitzen, zu Objekthypothesen zusammen. Die Vorgehensweise orientiert sich an der in Kapitel 3.3 beschriebenen Bewegungssegmentierung. Die Verfahren unterscheiden sich im wesentlichen in der Festlegung der Nachbarschaft von Bildelementen. Beim Farbregionenfluß werden zwei Regionen als benachbart betrachtet, wenn sich ihre umschreibenden Rechtecke überlappen. Im Unterschied zur Farbregion müssen die Bildpunkte eines Clusters jedoch keine zusammenhängende Fläche bilden. Ein Cluster kann Bildpunkte enthalten, die von der Mehrheit der restlichen Bildpunkte des Clusters isoliert sind und eine große Distanz zum zugehörigen Clusterschwerpunkt in der Bildebene besitzen. Das umschreibende Rechteck eines Clusters würde in solchen Fällen durch diese exponierten Bildpunkte festgelegt und damit

eine zu große Ausdehnung des Clusters vorspiegeln. Eine von Ausreißern unabhängige Information über die Clusterausdehnung kann aus der Kovarianzmatrix des Clusters im x - y -Merkmalsraum gewonnen werden. Die beiden Eigenvektoren der Kovarianzmatrix entsprechen den Richtungen maximaler und minimaler Streuung der Clusterpunkte in der Bildebene. Die Streuungen selbst sind die Quadratwurzeln der Eigenwerte der Kovarianzmatrix. In Bild 4.6 sind die Eigenvektoren einiger Cluster abgebildet. Die Längen der Eigenvektoren entsprechen bis auf einen konstanten Faktor den zugehörigen Streuungen. Statt des umschreibenden Rechtecks wird in der Bewegungssegmentierung des Farbclusterflusses nun jenes achsenparallele Rechteck verwendet, das die im Clusterschwerpunkt zentrierten Eigenvektoren umfaßt (siehe Bild 4.6). Als praktikable Wahl für die Länge eines Eigenvektors hat sich das Vierfache der zugehörigen Streuung erwiesen. Eine Ergänzung im Vergleich zur Bewegungssegmentierung im Farbregionenfluß stellt die Einführung eines Konfidenzmaßes dar, das zur Bewertung des Informationsgehalts der Trajektorien

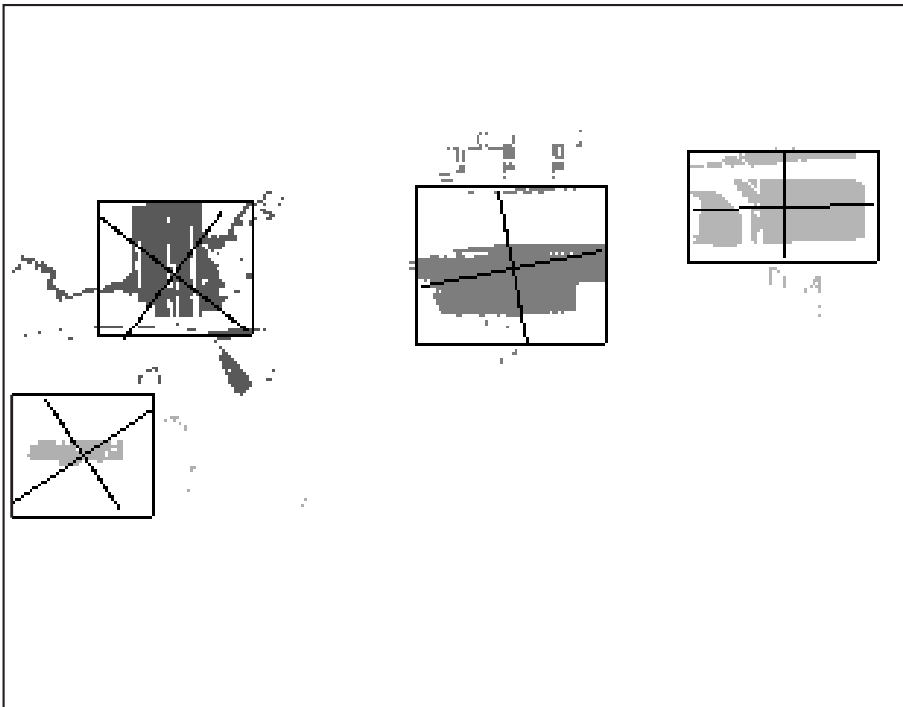


Bild 4.6: Beispiele für die Beschreibung der Clusterausdehnung durch Eigenvektoren und Eigenwerte der Kovarianzmatrix des Clusters im x - y -Merkmalsraum. Die eingezeichneten orthogonalen Achsenpaare verlaufen in Richtung der Eigenvektoren; ihre Längen entsprechen bis auf einen Faktor den Eigenwerten. Die Rechtecke umschließen die Achsenpaare und liefern eine Beschreibung der Clusterausdehnung.

dient. In der Clusteranalyse werden große Bildbereiche mit einer homogenen Farbverteilung (z.B. Straße, Himmel) in mehrere Cluster aufgeteilt, man spricht dann von einer Übersegmentierung. Die Cluster eines solchen Bildbereichs liegen im Farb-Positionsraum eng beieinander und weisen daher eine starke Kopplung auf. Lokale Bildänderungen in einem homogen gefärbten Bereich, z.B. durch Teilverdeckungen, können zu einer globalen Verschiebung der zugehörigen Cluster führen. Es ist somit sinnvoll, ein auf der Clusterverteilung basierendes Konfidenzmaß zu berechnen, das aussagt, wie gut die Trajektorien mit der Szenenbewegung übereinstimmen. Naheliegend ist die Ableitung des Konfidenzmaßes aus den Abständen der Clusterzentren im Farb-Positionsraum. Je größer der Abstand eines Clusterzentrums zu seinen nächsten Nachbarn ist, desto „glaubwürdiger“ ist seine Trajektorie. In der Arbeit wurde als Konfidenzmaß der mittlere Abstand jedes Clusterzentrums zu seinen drei nächsten Nachbarn bestimmt. In Bild 4.7 b) ist die Konfidenz der Cluster in Form des Grauwerts dargestellt. Hell bedeutet eine hohe und dunkel eine niedrige Konfidenz. Fahrzeugschatten, Fahrbahnmarkierungen und Verkehrszeichen besitzen eine hohe Konfidenz, da sie einen hohen Farbkontrast gegenüber ihrer Umgebung besitzen. Cluster im Bereich der Straße und des Himmels weisen dagegen eine niedrige Konfidenz auf. In der Bewegungssegmentierung wird von jedem Cluster eine minimale Konfidenz gefordert. Cluster, deren Konfidenz unterhalb dieser Schwelle liegt, werden bei der Objekthypothesenbildung nicht berücksichtigt.

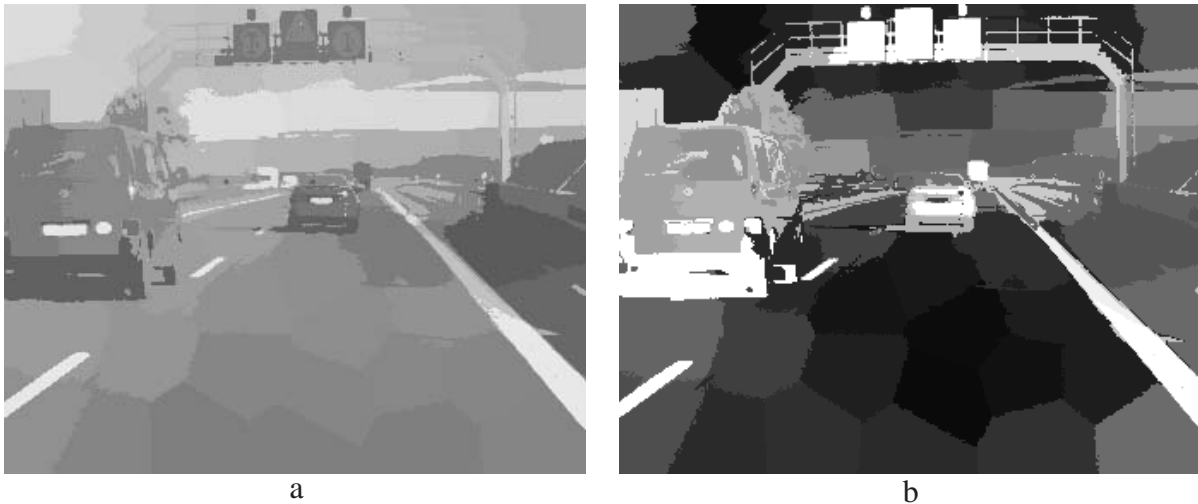


Bild 4.7: Beispiel für die Konfidenz der Cluster, mit der die Übereinstimmung zwischen Cluster- und Szenenbewegungen bewertet wird. In Bild a) sind die Ergebnisse der Clusteranalyse dargestellt. In Bild b) sind Konfidenzen der Cluster als Grauwerte kodiert. Hell kennzeichnet hohe, dunkel niedrige Konfidenz.

4.5 Ergebnisse der Objektdetektion anhand des Farbclusterflusses

Die Bewegungssegmentierung wurde an den in Kapitel 3.3 genannten vier Bildsequenzen quantitativ evaluiert. Zusätzlich wurden noch zwei weitere Szenen mit Fußgängern, im folgenden als Szene 5 und 6 bezeichnet, ausgewertet.

- In Szene 5 bewegt sich ein Fußgänger quer zur Kamera aus einem abgeschatteten in einen sonnenbestrahlten Bereich. Gleichzeitig führt die Kamera eine leichte Rückwärtsbewegung mit ca. 2 km/h aus; die Sequenz erstreckt sich über 85 Bilder.
- In Szene 6 ist ein partiell verdeckter Fußgänger hinter einem Gittertor. Die Kamerabewegung verläuft mit einer Geschwindigkeit von ca. 5 km/h auf den Fußgänger zu; die Sequenz umfaßt 55 Bilder.

Tabelle 4.1: Ergebnisse der Objektdetektion durch den Farbclusterfluß

Sequenz	Richtig	Nicht	Falsch	U_P	U_B	Verschm.
1	62	0	0	40	61	0
2	45	0	2	23	33	0
3	118	75	45	39	42	1
4	229	198	116	33	30	10
5	82	0	14	28	60	0
6	52	2	1	27	55	0

Zur Bewertung der Detektionsleistung wurden die in Kapitel 3.3 genannten Größen berechnet. Bei der Berechnung des Überlappungsgrads U_B wurden, statt der umschreibenden Rechtecke der Cluster, die aus den Kovarianzmatrizen der Cluster berechneten und im vorigen Abschnitt beschriebenen Rechtecke verwendet. In den Sequenzen 1, 2 und 4 wurden 64, in den restlichen Sequenzen 128, Cluster verwendet. Die Ergebnisse sind in Tabelle 4.1 angegeben. In Bild 4.8 sind einige detektierte Objekte aus den ersten 4 Sequenzen dargestellt. In der Motorradsequenz traten keine Probleme auf. In der zweiten Sequenz konnte zwar das Fahrzeug in jedem Bild detektiert werden, die Überdeckung zwischen Objekt und Objekthypothese war jedoch gering. Wie aus dem linken oberen Bild aus Bild 4.9 hervorgeht, ist dies darauf zurückzuführen, daß nur die Bewegung des unteren Fahrzeugteils korrekt bestimmt wurde. Die obere Fahrzeughälfte konnte, aufgrund ihres geringen Farbkontrasts zur Umgebung und der schnellen Bewegung des Fahrzeugs, nicht korrekt verfolgt werden. Die hohe Anzahl der nicht detektierten Objekte in den Szenen

3 und 4 hat zwei Ursachen: Erstens können, wie beim Farbregionenfluß auch, Fahrzeuge ohne signifikante Bildbewegung prinzipiell nicht detektiert werden; zweitens werden Fahrzeuge nicht unmittelbar bei ihrem Eintreten in das Blickfeld der Kamera detektiert. Wie aus Bild 4.9 b) deutlich wird, „springen“ Cluster vom Hintergrund auf das ins Bild eintretende Objekt. Die Trajektorien der Cluster des Objekts stimmen somit nicht mit der Objektbewegung überein. Entsprechend der Länge des Zeitfensters für die Korrelation der Trajektorien in der Bewegungssegmentierung, kann das neue Objekt erst nach etwa 5 Bildern detektiert werden. Trotz der Verwendung eines Konfidenzmaßes, ist ein Teil der Falschdetektionen in Szene 3 und 4 auf Clusterbewegungen in den übersegmentierten Bildbereichen der Straße und des Himmels zurückzuführen (siehe Bild 4.9 d)). In Bild 4.9 c) sind die Bewegungen der zur Straße gehörigen Cluster deutlich zu erkennen. Besonders in der Umgebung des Fahrzeugs führt die Verdeckung bzw. Freilegung von Straßenteilen zu Clusterverschiebungen, die nicht mit der Szenenbewegung korrespondieren.

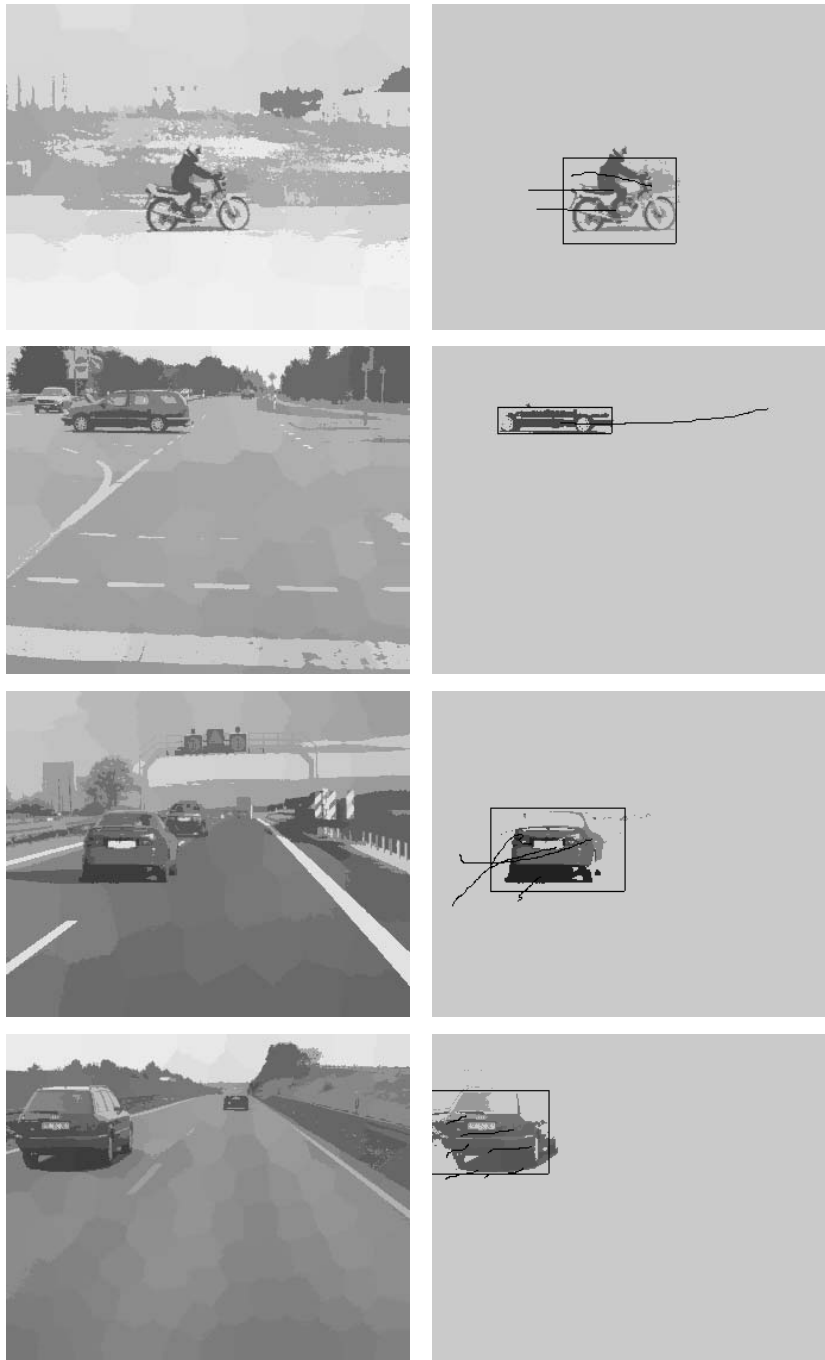


Bild 4.8: Beispiele der Objektdetektion durch den Farbclusterfluß. In der linken Spalte sind die Ergebnisse der Clusteranalyse im Farb-Positionsraum dargestellt. Die Originalfarben der Bildpunkte sind durch die Farben der Referenzvektoren ersetzt. In der rechten Spalte sind die Objekthypothesen dargestellt. Die kalmangeglätteten Trajektorien der Referenzvektoren und die objektbeschreibenden Rechtecke sind schwarz eingezeichnet.

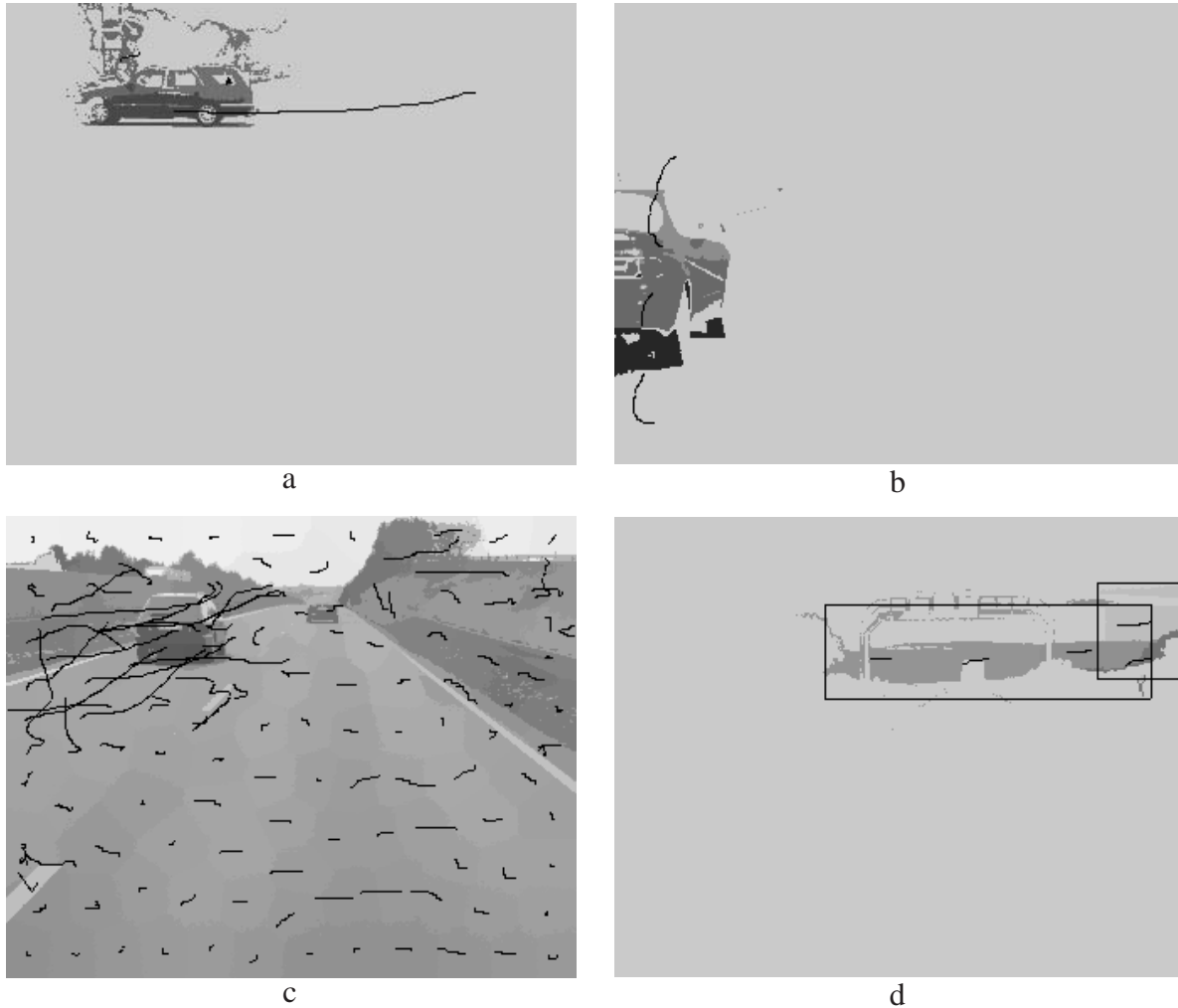


Bild 4.9: Problemfälle bei der Objektdetektion durch den Farbclusterfluß. In Bild a) kann aufgrund der großen Bildverschiebung des Fahrzeugs, in Verbindung mit dessen geringem Farbkontrast gegenüber dem Hintergrund, nur die Bewegung des unteren Fahrzeugteils korrekt bestimmt werden. In b) ist ein ins Blickfeld der Kamera eintretendes Fahrzeug dargestellt. Die Cluster ändern ihre Objektzugehörigkeit, sie wandern von der Straße zum Fahrzeug ab. Folglich stimmen ihre Trajektorien nicht mit denen der Bildobjekte überein. In Bild c) sind die Effekte bei Verdeckung bzw. Freilegung von Straßenteilen durch ein fahrendes Auto dargestellt. Deutlich sind die Bewegungen der zur Straße gehörigen Cluster rund um das Fahrzeug zu erkennen. In d) sind schließlich Falschdetektionen dargestellt, die ihre Ursache in der Übersegmentierung des Himmels haben.

Überraschend gute Ergebnisse konnten in den Fußgängersequenzen erzielt werden. Bei Betrachtung von Bild 4.10 a) und f) würde man vermuten, daß ein bewegungs- und gebietsbasierter Ansatz für die Detektion von Fußgängern prinzipiell ungeeignet ist. Folgende Probleme werden deutlich:

1. Die Abbildungen der Extremitäten haben keinen ausgeprägten Flächencharakter. Die farbliche Homogenität eines Fußgängers wird außerdem durch Faltenwurf und mögliche Texturierung der Kleidung gestört.
2. Durch Gehbewegungen werden Formänderungen der in die Bildebene projizierten Arme und Beine verursacht.
3. Die Bewegungen des linken und rechten Armes bzw. Beines unterscheiden sich stark voneinander. Die modellfreie Bewegungssegmentierung beruht aber auf der Ähnlichkeit der Bewegungen von Objektteilen.

Trotz dieser Probleme und einer abrupten Intensitätsänderung, beim Übergang vom abgeschatteten in den sonnenbeschienenen Bereich, konnte der Fußgänger aus Szene 5 mit dem Farbclusterfluß erfolgreich detektiert werden (siehe Bild 4.10 b)–e)). Zu oben genannten Problemen lassen sich folgende Erklärungen geben:

1. Cluster eignen sich auch zur Repräsentation von Bildobjekten, die keinen ausgeprägten Flächencharakter besitzen, da von den Bildpunkten eines Clusters nicht gefordert wird, daß sie eine zusammenhängende Fläche bilden.
2. Im Farbclusterfluß wird keine Formstabilität der Cluster in der Bildebene vorausgesetzt.
3. In der Regel werden in einem Cluster Teile beider Arme bzw. Beine repräsentiert. Daraus folgt, daß die Clusterbewegung nicht die Bewegung *eines* Arms bzw. Beins, sondern die Bewegung des Schwerpunkts beider Arme bzw. Beine widerspiegelt. Diese Schwerpunktbewegungen verlaufen gleichmäßig und ähneln der Bewegung des Oberkörpers, so daß eine modellfreie Bewegungssegmentierung möglich ist.

Die Unempfindlichkeit des Verfahrens gegenüber partiellen Verdeckungen der zu detektierenden Objekte wird in den Ergebnissen aus Szene 6 (siehe Bild 4.11) deutlich. Aufgrund des Farbkontrastes zwischen Fußgänger und Gitter ist eine Separation dieser Objekte in der Clusteranalyse möglich. Sie führt dazu, daß die Cluster des Fußgängers jeweils in mehrere Flächen in der Bildebene zerfallen, die durch Gitterstäbe getrennt sind.

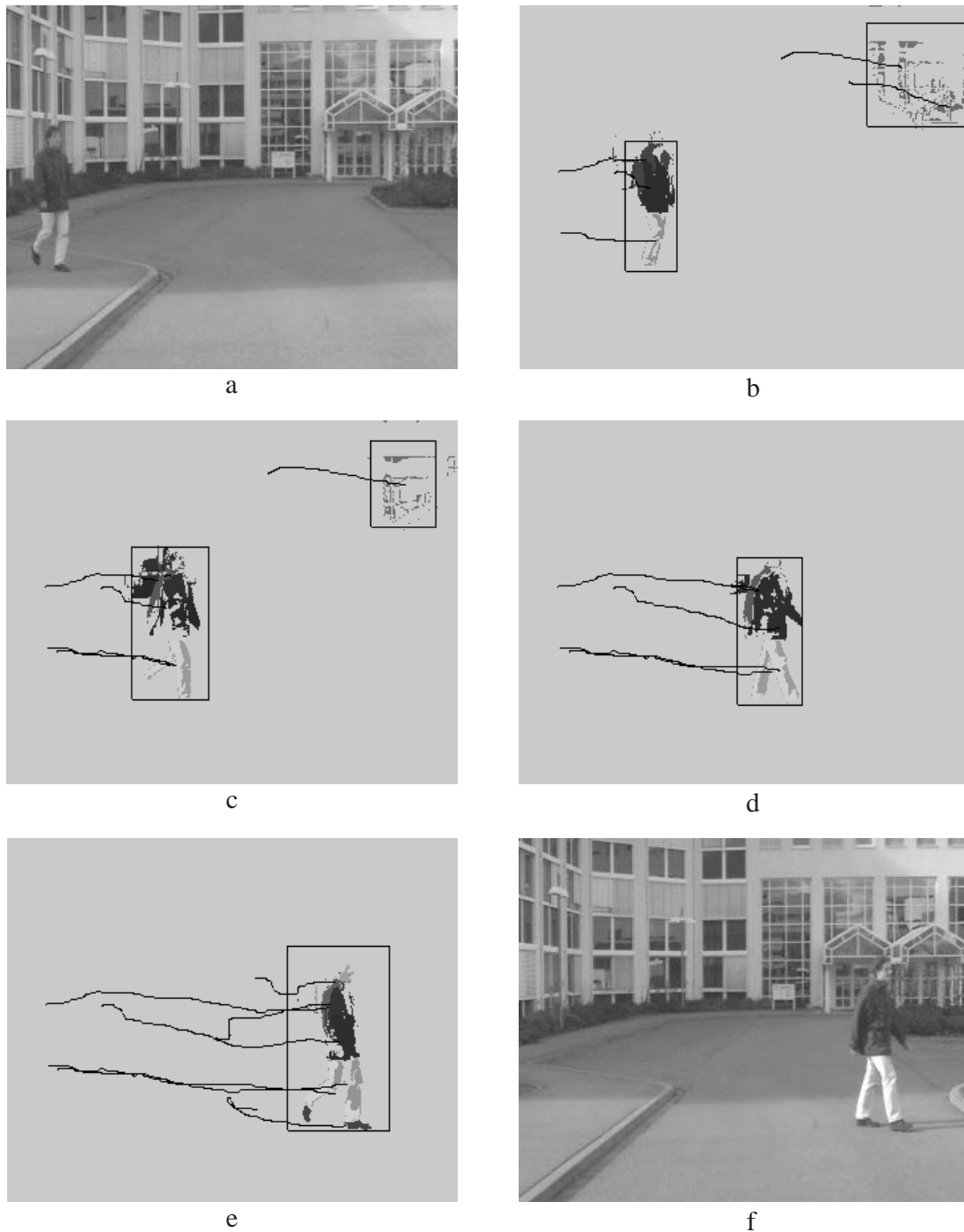


Bild 4.10: Detektion eines Fußgängers bei starken Beleuchtungsänderungen anhand des Farbclusterflusses. Die Bilder a) und f) sind dem Anfang bzw. Ende der Originalsequenz entnommen. In b)–e) sind einige Ergebnisse der Objektdetektion dargestellt. Die kalmangeglätteten Trajektorien der Referenzvektoren und die objektbeschreibenden Rechtecke sind schwarz eingezeichnet.

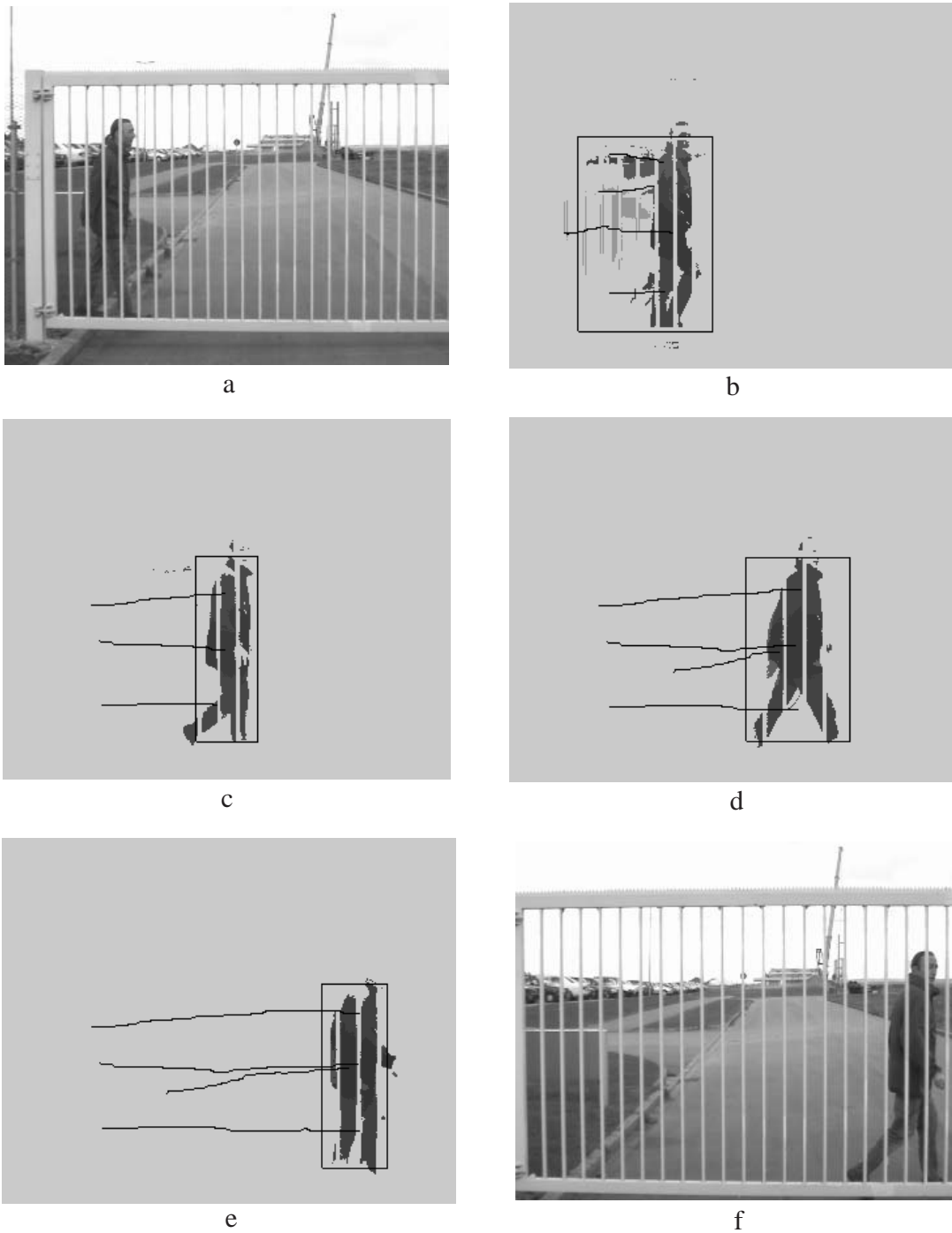


Bild 4.11: Detektion eines partiell verdeckten Fußgängers. Die Bilder a) und f) sind dem Anfang bzw. Ende der Originalsequenz entnommen. In b)–e) sind einige Ergebnisse der Objektdetektion dargestellt. Die kalmangeglätteten Trajektorien der Referenzvektoren und die objektbeschreibenden Rechtecke sind schwarz eingezeichnet.

Im folgenden werden Stärken und Schwächen des Verfahrens zusammengefaßt:

- + Das Verfahren benötigt keine kalibrierte Kamera, kein Straßendmodell, keine Messung der Eigengeschwindigkeit und keine Kenntnisse über das Aussehen der zu detektierenden Objekte.
- + In der Bewegungsbestimmung sind nur zwei Parameter einzustellen, die Anzahl der Cluster und die Gewichtung zwischen Farb- und Positionsmerkmalen.
- + Auch bei Bildobjekten, die keinen ausgesprochenen Flächencharakter besitzen, ist eine Bewegungsbestimmung möglich.
- + Objekte mit signifikanter Relativbewegung zur Kamera lassen sich zuverlässig detektieren.
- + Das Verfahren ist robust gegenüber partiellen Verdeckungen.
- + Das Verfahren eignet sich zur Detektion von formveränderlichen Objekten.
 - Es lassen sich nur Objekte detektieren, die sich relativ zur Kamera bewegen.
 - Objekte, die in der Bildebene aneinandergrenzen und dieselbe Bildbewegung besitzen, können nicht voneinander getrennt werden. Genausowenig kann zwischen dem Schatten und dem Objekt selbst unterschieden werden.
 - Wenn Objekte in das Blickfeld der Kamera eintreten bzw. deren Blickfeld verlassen, ändern die betroffenen Cluster ihre Objektzugehörigkeit. Ihre Trajektorien stimmen nicht mit der Bildbewegung überein.
 - In großen Bildbereichen mit einer homogenen Farbverteilung (z.B. Straße, Himmel) treten häufig Falschdetektionen auf.
 - Sehr große Bildbewegungen können nicht erfaßt werden.
 - Relativ hoher Rechenaufwand in der Clusteranalyse.

5 Zusammenfassung und Ausblick

5.1 Zusammenfassung

Sichtgestützte Fahrerassistenzsysteme benötigen zur Interpretation von Verkehrsszenen Kenntnisse über die Position und Bewegung anderer Verkehrsteilnehmer. Vor diesem Hintergrund wurde ein Ansatz zur Detektion von Verkehrsteilnehmern, anhand optischer Sensordaten, untersucht. Ein besonderes Augenmerk wurde auf Verkehrsteilnehmer mit einer großen Relativbewegung zum eigenen Fahrzeug gerichtet, da von ihnen ein erhöhtes Gefahrenpotential im Hinblick auf mögliche Kollisionen ausgeht. Das vorgestellte Verfahren trifft keine vereinfachenden Einschränkungen bezüglich des Verkehrsszenarios und der Art der Verkehrsteilnehmer, ferner wurden keine Kenntnisse über die Eigenbewegung des Fahrzeugs und den Verlauf der Fahrbahn vorausgesetzt. Für Aufnahmen stand eine unkalibrierte, in einem Kraftfahrzeug montierte, 3-Chip CCD-Farbkamera zur Verfügung.

Die Detektionsaufgabe wird durch ein Verfahren der Bildfolgenanalyse gelöst. Den Ausgangspunkt der Bildfolgenanalyse stellt die 3D-Bewegung zwischen der aufnehmenden Kamera und der Szene dar. Durch die Projektionsgleichung der Kamera läßt sich die 3D-Bewegung in eine Bildbewegung auf der Sensorfläche der Kamera umrechnen. Die Bildbewegung ist keine meßbare Größe im Ausgangssignal der Kamera und muß erst durch geeignete Verfahren aus den zeitlichen Bildänderungen geschätzt werden. In der Bewegungssegmentierung wird dann jedes Einzelbild in Segmente aufgeteilt, in denen die Bildbewegung einen homogenen Verlauf aufweist. Ein reales Objekt (z.B. Auto, Motorrad), das sich durch seine Bewegung von benachbarten Objekten der Szene unterscheidet, wird im segmentierten Bild als *ein* Bewegungssegment dargestellt. Die Bewegungssegmentierung ermöglicht somit die Extraktion von Objekten, unabhängig von deren Form und Farbe. Zur Lösung der Detektionsaufgabe fehlt noch die Selektion jener Bewegungssegmente, die den Abbildungen von Verkehrsteilnehmern entsprechen. Da sich relevante Verkehrsteilnehmer durch ihre große Relativbewegung zur Kamera auszeichnen, werden nur solche Bewegungssegmente selektiert, die eine signifikante Bildbewegung besitzen.

In dieser Arbeit wurden zwei als Farbreionen- und Farbcusterfluß bezeichnete Detekti-

onsverfahren vorgestellt. Die Besonderheit beider Verfahren ist die Schätzung der Bildbewegung durch Beobachtung des zeitlichen Verhaltens von Farbgebieten. In einem ersten Schritt werden die Einzelbilder durch Farbsegmentierung in Farbgebiete aufgeteilt. Anschließend wird die Bewegung der Farbgebiete in Form der Trajektorien ihrer Schwerpunkte bestimmt. Die Bewegungssegmentierung faßt schließlich benachbarte Farbgebiete mit ähnlichen Trajektorien zu Objekthypothesen zusammen.

Im Farbreionenfluß wird jedes Einzelbild einer Clusteranalyse im Farbraum unterzogen. Sie reduziert die im Bild auftretenden Farben auf eine kleine Anzahl signifikanter Farben. Durch die Farbreduktion wird das Bild in Regionen aufgeteilt, die jeweils aus zusammenhängenden Bildpunkten gleicher Farbe bestehen. Wegen des Meßrauschens der Kamera werden die farbquantisierten Bilder nachträglich durch Medianfilterung bzw. Relaxation geglättet. Die anschließende Farbzusammenhangsanalyse wandelt die segmentierten Bilder in eine symbolische Darstellungsform um, die den folgenden Verarbeitungsschritten einen effizienten Zugriff auf Regionen und deren Merkmale ermöglicht. Danach erfolgt die Bestimmung der Bild-zu-Bild Verschiebungen der Regionen durch paarweise Zuordnung einander entsprechender Regionen in zwei aufeinanderfolgenden Bildern. Die zeitliche Verkettung der Bild-zu-Bild Verschiebungen liefert die Trajektorien der Farbreionen über eine Bildsequenz. Kalmanfilter werden zur Prädiktion der Regionenverschiebungen und zur Glättung der Trajektorien eingesetzt. Den Abschluß bildet die Bewegungssegmentierung, die benachbarte Farbreionen mit ähnlichen Trajektorien zu Objekthypothesen zusammenfaßt. Qualitative Experimente wurden an einigen hundert Bildsequenzen durchgeführt, die hauptsächlich von Autobahnfahrten stammen. Quantitativ wurde die Detektionsleistung des Farbreionenflusses an vier verschiedenen Sequenzen mit insgesamt 466 Einzelbildern gemessen. In allen Untersuchungen hat sich gezeigt, daß Fahrzeuge, die eine signifikante Bildbewegung besitzen, zuverlässig detektiert werden. Hingegen werden Verkehrsteilnehmer, die sich nicht oder nur gering gegenüber der Kamera bewegen, bei der Bildung von Objekthypothesen nicht berücksichtigt. Als Problem hat sich das Meßrauschen der Kamera erwiesen; es kann trotz zeitlicher Glättung der Bilddaten zu erheblichen Änderungen in Folgebildern führen. Betroffen sind hauptsächlich Regionen, die einen schwachen Farbkontrast gegenüber ihren Nachbarregionen besitzen. Auch schmale Regionen, wie sie z.B. bei Armen und Beinen von Personen entstehen, erweisen sich häufig als instabil. Dort genügt bereits die Farbänderung weniger Bildpunkte, um eine Region auseinanderzubrechen bzw. zwei Regionen zu verschmelzen.

Die Suche nach Verfahren, die eine zeitliche Konsistenz der Segmentierungsergebnisse gewährleisten, hat schließlich zur Entwicklung des Farbclusterflusses geführt. Im Farbclusterfluß wird die Kenntnis, daß sich der Bildinhalt in aufeinanderfolgenden Bildern in der Regel nur geringfügig ändert, unmittelbar in das Segmentierungsverfahren integriert. Anstatt die Einzelbilder unabhängig voneinander zu segmentieren, wird auf dem

Segmentierungsergebnis aus dem vorhergehenden Bild aufgebaut – man spricht von einer dynamischen Segmentierung. Die Basis des Farbclusterflusses bildet die Segmentierung der einzelnen Bilder durch Clusteranalyse im Farb-Positionsraum. Jeder Bildpunkt wird dabei durch seine drei Farbwerte sowie seine horizontale und vertikale Position in der Bildebene charakterisiert. Die Clusteranalyse faßt Bildpunkte, die sich in ihrer Farbe und ihrer Position in der Bildebene ähnlich sind, in Cluster zusammen. Jedes Cluster ist durch seinen Referenzvektor (Schwerpunktvektor) im Farb-Positionsraum beschrieben. Den Kern der dynamischen Segmentierung bildet das k-means Clusterverfahren. Ausgehend von den Referenzvektoren des vorhergehenden Bildes führt es eine iterative Anpassung der Referenzvektoren an die aktuellen Bilddaten durch. Das k-means Verfahren bewirkt nicht nur eine zeitlich konsistente Segmentierung, es legt außerdem die Zuordnung einander entsprechender Referenzvektoren in Folgebildern fest. Ohne eine explizite Zuordnungsprozedur erhält man die Bewegung der Referenzvektoren im Farb-Positionsraum und damit die Trajektorien der Clusterschwerpunkte in der Bildebene. Eine Verbesserung der Bewegungsbestimmung wird durch Prädiktion der Clusterverschiebungen mittels Kalmanfilter erzielt. Die Initialisierung des k-means Verfahrens erfolgt dann mit den prädizierten Referenzvektoren, wobei die Prädiktion nur auf die Positionskoordinaten der Referenzvektoren angewandt wird. In der abschließenden Bewegungssegmentierung werden Cluster, die in der Bildebene benachbart sind und ähnliche Trajektorien besitzen, zu Objekthypothesen zusammengefaßt. Neben zahlreichen qualitativen Untersuchungen, wurde die Detektionsleistung des Farbclusterflusses an sechs verschiedenen Sequenzen mit insgesamt 606 Einzelbildern gemessen. Hervorzuheben sind die Ergebnisse bei der Detektion von Fußgängern. Obwohl Arme und Beine eines Fußgängers keinen ausgeprägten Flächencharakter besitzen und Gehbewegungen zu starken Formänderungen führen, können Fußgänger durch den Farbclusterfluß detektiert und vom Hintergrund getrennt werden. Selbst unter erschwerten Bedingungen, wie z.B. abrupten Helligkeitsänderungen oder Teilverdeckungen der Fußgänger, erweist sich die Detektion als robust. Eine Schwäche des Verfahrens ist die ungenaue Bewegungsbestimmung in großen, homogen gefärbten Bildbereichen. Probleme bereiten außerdem sehr schnell bewegte und in das Bild eintretende Verkehrsteilnehmer; sie führen dazu, daß Cluster ihre Objektzugehörigkeit ändern. Die Folge ist, daß die Trajektorien der betroffenen Cluster von der tatsächlichen Bildbewegung abweichen.

5.2 Ausblick

Die vorliegende Arbeit bildet die Grundlage für eine bewegungsbasierte Detektion von Verkehrsteilnehmern. Für eine praktische Anwendung der Verfahren in Fahrerassistenzsystemen sind allerdings noch ergänzende Arbeiten erforderlich:

- Portierung des Farbclusterflusses auf eine fahrzeugtaugliche Hardware. Nur so sind umfangreiche Untersuchungen in Echtzeit möglich.
- Umfangreiche quantitative Auswertungen von Bildsequenzen. Insbesondere sind noch Untersuchungen bei widrigen Witterungs- und Beleuchtungsbedingungen durchzuführen (z.B. Nebel, Regen, Dämmerung).
- Fusion der bewegungsbasierten Ansätze mit anderen optischen Detektionsverfahren (z.B. Stereo). Durch Untersuchung von Bilddaten derselben Verkehrsszenen ist zu klären, wo die Stärken und Schwächen der verschiedenen Verfahren liegen und wie sich die Verfahren sinnvoll ergänzen lassen.
- Fusion mit Sensordaten im nicht-sichtbaren Spektralbereich. Von besonderem Interesse sind Sensordaten, die unmittelbaren Aufschluß über Entfernung und Geschwindigkeit der Verkehrsteilnehmer geben. In [Heisele et al. 96] sind erste Ansätze über die Kombination des Farbregionenflusses mit Radardaten zu finden.

A Struktogramme

A.1 Zuordnung von Regionen im Farbregionenfluß

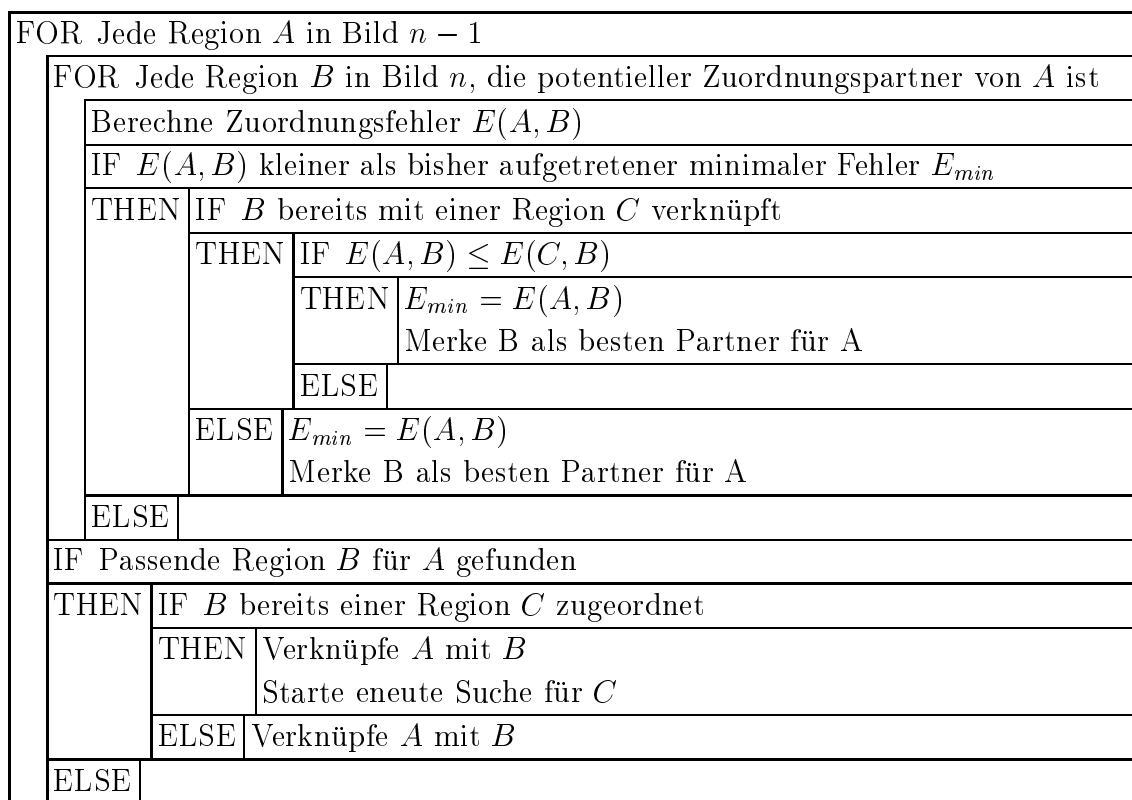


Bild A.1: Struktogramm für die Regionenzuordnung im Farbregionenfluß

A.2 Bewegungssegmentierung im Farbregionenfluß

FOR Jede Objekthypothese O	
FOR Jede Objekthypothese P , die nicht O ist	
IF O und P benachbart und Korrelationen der Trajektorien der Regionen von O mit den Trajektorien der Regionen von P ausreichend groß	
THEN	Verschmelze O und P zu einer Objekthypothese
ELSE	
FOR Jede Region A	
IF A gehört zu einer Objekthypothese O	
THEN	IF Korrelation der Trajektorie von A mit der Trajektorie einer anderen Region B von O ausreichend groß
THEN	Behalte A in Objekthypothese O
ELSE	Entferne A aus Objekthypothese O Fahre fort, wie wenn A zu keiner Objekthypothese gehört hätte
ELSE	FOR Jede Objekthypothese O
IF Korrelation der Trajektorie von A mit der Trajektorie einer Region B von O ausreichend groß	
THEN	Füge A zur Objekthypothese O hinzu Beende Schleife für Objekthypothesen O
ELSE	
IF A gehört zu keiner Objekthypothese	
THEN	FOR Jede Region B , die zu keiner Objekthypothese gehört und nicht A ist
IF Korrelation der Trajektorien von A und B ausreichend groß	
THEN	Bilde neue Objekthypothese aus A und B Beende Schleife für Regionen B
ELSE	
ELSE	

Bild A.2: Struktogramm der Bewegungssegmentierung im Farbregionenfluß

A.3 Beschleunigung des parallelen k-means Clusterverfahrens

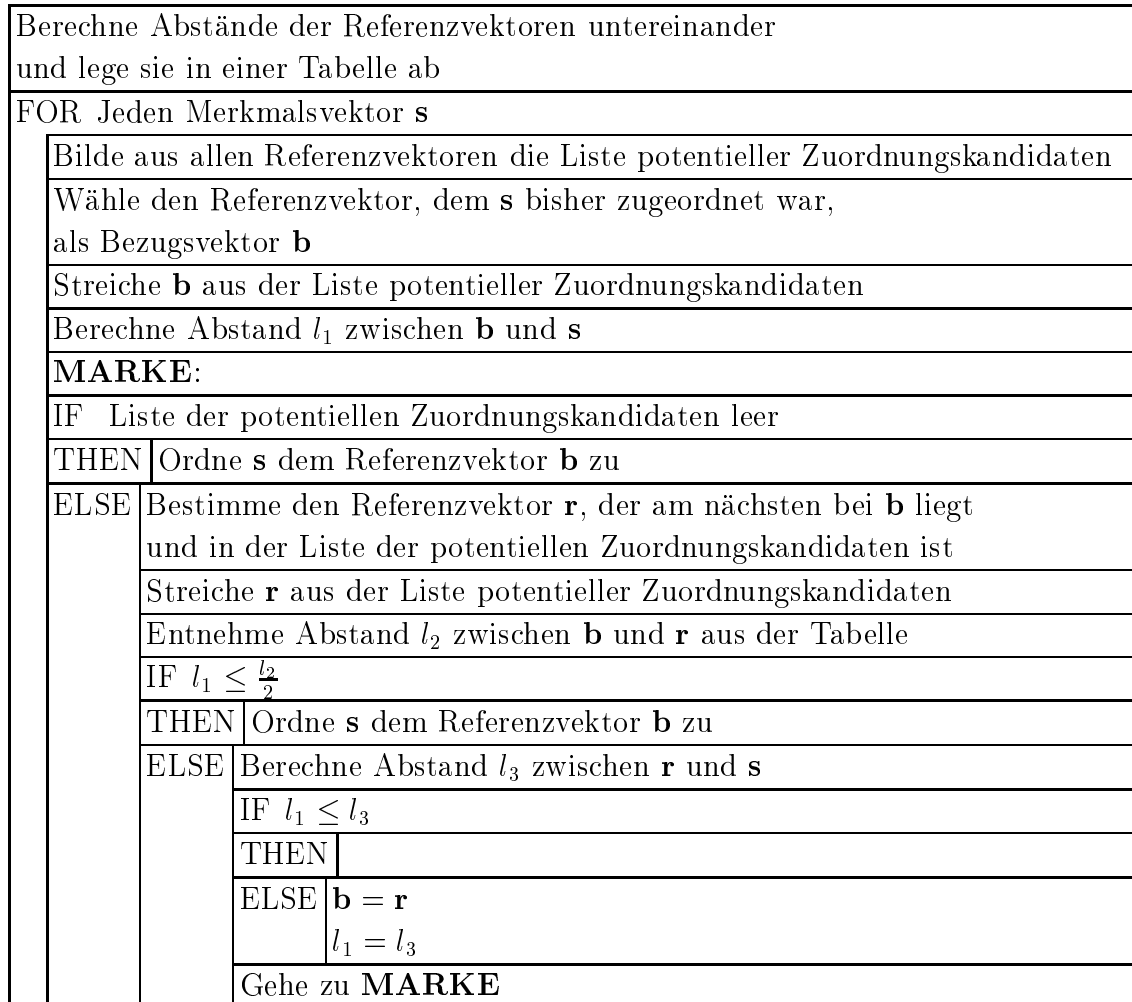


Bild A.3: Struktogramm für die schnelle Berechnung der Partitionierung im parallelen k-means Clusterverfahren

B Kalmanfilterung

B.1 Modellierung der Beschleunigung als weißes Rauschen

In diesem Modell wird die Beschleunigung einer zeitkontinuierlichen, eindimensionalen Bewegung als weißes Rauschen modelliert.

$$\frac{d^2 \xi(t)}{dt^2} = w(t) \quad (\text{B.1})$$

Für das Rauschen gilt:

$$E[w(t)] = 0 \quad (\text{B.2})$$

$$E[w(t)w(\tau)] = \sigma_w^2 \delta(t - \tau) \quad (\text{B.3})$$

Dabei symbolisiert $\delta(\cdot)$ die Diracfunktion, σ_w^2 ist die spektrale Leistungsdichte des Rauschens. Das dynamische Verhalten des Systems ist durch folgende Gleichung beschrieben:

$$\frac{d\mathbf{x}(t)}{dt} = \mathbf{F}\mathbf{x}(t) + \mathbf{w}(t) \quad (\text{B.4})$$

mit

$$\mathbf{x}(t) = \left(\xi, \frac{d\xi}{dt} \right)^T \quad (\text{B.5})$$

$$\mathbf{F} = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \quad (\text{B.6})$$

$$\mathbf{w}(t) = (0, w(t))^T \quad (\text{B.7})$$

Geht man zum zeitdiskreten Fall mit einer Abtastperiode T über, so erhält man [Bar-Shalom & Li 93]:

$$\mathbf{x}(k+1) = \mathbf{A}\mathbf{x}(k) + \mathbf{w}(k) \quad (\text{B.8})$$

wobei

$$\begin{aligned}
 \mathbf{A} &= e^{\mathbf{F}T} \\
 &= \mathbf{I} + \mathbf{F}T + \frac{\mathbf{F}^2 T^2}{2!} + \dots \\
 &= \begin{pmatrix} 1 & T \\ 0 & 1 \end{pmatrix}
 \end{aligned} \tag{B.9}$$

Für das Systemrauschen gilt:

$$\mathbf{w}(k) = \int_0^T e^{\mathbf{A}(T-\tau)} (0, w(kT + \tau))^T d\tau \tag{B.10}$$

Mit den Gleichungen (B.3) und (B.10) läßt sich die Kovarianzmatrix des zeitdiskreten Systemrauschens angeben:

$$\begin{aligned}
 \mathbf{Q} &= E[\mathbf{w}(k)\mathbf{w}(k)^T] \\
 &= \int_0^T \begin{pmatrix} T - \tau \\ 1 \end{pmatrix} \begin{pmatrix} T - \tau & 1 \end{pmatrix} \sigma_w^2 d\tau \\
 &= \begin{pmatrix} \frac{1}{3}T^3 & \frac{1}{2}T^2 \\ \frac{1}{2}T^2 & T \end{pmatrix} \sigma_w^2
 \end{aligned} \tag{B.11}$$

Bei Messung der Position lautet die Meßgleichung:

$$\mathbf{y}(k) = \mathbf{C}\mathbf{x}(k) + \mathbf{v}(k) \tag{B.12}$$

mit

$$\mathbf{C} = \begin{pmatrix} 1 & 0 \end{pmatrix} \tag{B.13}$$

$$\begin{aligned}
 \mathbf{R} &= E[\mathbf{v}(k)\mathbf{v}(k)^T] \\
 &= \sigma_v^2
 \end{aligned} \tag{B.14}$$

Die Initialisierung des Filters wird zum Zeitpunkt T mit folgenden Werten durchgeführt:

$$\hat{\mathbf{x}}^+(1) = \left(y(1), \frac{y(1) - y(0)}{T} \right)^T \tag{B.15}$$

$$\hat{\mathbf{P}}^+(1) = \begin{pmatrix} \sigma_v^2 & \frac{\sigma_v^2}{T} \\ \frac{\sigma_v^2}{T} & \frac{T\sigma_w^2}{3} + \frac{2\sigma_v^2}{T^2} \end{pmatrix} \tag{B.16}$$

Die Initialisierung ist so gewählt, daß die Schätzfehler mittelwertfrei sind. Für die folgende Herleitung der Initialisierungsgrößen werden die Vektorkomponenten benötigt:

$$\mathbf{x}(k) = (x_1(k), x_2(k))^T \tag{B.17}$$

$$\mathbf{y}(k) = y(k) \tag{B.18}$$

$$\mathbf{w}(k) = (w_1(k), w_2(k))^T \tag{B.19}$$

$$\mathbf{v}(k) = v(k) \tag{B.20}$$

Die Messungen für $k = 0$ und $k = 1$ lauten:

$$y(0) = x_1(0) + v(0) \tag{B.21}$$

$$y(1) = x_1(0) + Tx_2(0) + w_1(0) + v(1) \tag{B.22}$$

Durch Einsetzen der Meßwerte in Gl. (B.15) erhält man:

$$\hat{x}_1^+(1) = x_1(1) + v(1) \tag{B.23}$$

$$\hat{x}_2^+(1) = x_2(0) + \frac{1}{T}(w_1(0) + w(1) - v(0)) \tag{B.24}$$

Für die zweite Systemgröße zum Zeitpunkt $kT = 1$ gilt:

$$x_2(1) = x_2(0) + w_2(0) \tag{B.25}$$

Für die Schätzfehler ergibt sich:

$$\hat{x}_1^+(1) - x_1(1) = v(1) \tag{B.26}$$

$$\hat{x}_2^+(1) - x_2(1) = \frac{1}{T}w_1(0) - w_2(0) + \frac{1}{T}[v(1) - v(0)] \tag{B.27}$$

Damit ist bewiesen, daß die Schätzfehler mittelwertfrei sind. Aus den Gleichungen (B.11), (B.14), (B.26) und (B.27) folgt für die Kovarianzmatrix von $\hat{\mathbf{x}}^+(1)$:

$$\hat{\mathbf{P}}^+(1) = E[\hat{\mathbf{x}}^+(1)\hat{\mathbf{x}}^+(1)^T] = \begin{pmatrix} \sigma_v^2 & \frac{\sigma_v^2}{T} \\ \frac{\sigma_v^2}{T} & \frac{T\sigma_w^2}{3} + \frac{2\sigma_v^2}{T^2} \end{pmatrix} \tag{B.28}$$

Nun soll die Einstellung der Filterparameter im Fall der Verfolgung von Regionen in Bildsequenzen betrachtet werden. Die einstellbaren Parameter des Kalmanfilters sind die Größen σ_v^2 und σ_w^2 . Im Meßrauschen berücksichtigt man, daß Verschiebungen der Regionenschwerpunkte nicht nur durch die Bewegung der Region, sondern auch durch Störungen in den segmentierten Bildern verursacht werden. Nach Begutachtung der Trajektorien wurde $\sigma_v^2 = 2$ gewählt. Es wird also ein mittlerer Fehler von $\sqrt{2}$ Bildpunkten pro Koordinate angenommen. Im Prozeßrauschen wird berücksichtigt, daß das Systemmodell die reale Bewegung nur ungenügend modelliert. Laut [Bar-Shalom & Li 93] ist σ_w^2 so zu wählen, daß die Änderung der Geschwindigkeit über eine Periode T in der Größenordnung von $\sqrt{T\sigma_w^2}$ liegt. Bei der Initialisierung wurde angenommen, daß die Geschwindigkeitsänderung im Mittel zu einer Schwerpunktverschiebung von 0,5 Bildpunkten pro Abtastperiode führt; daraus folgt $\sigma_w^2 = 3906$.

B.2 Modellierung der Ableitung der Beschleunigung als weißes Rauschen

In diesem Modell wird die zeitliche Ableitung der Beschleunigung als weißes Rauschen modelliert:

$$\frac{d^3\xi(t)}{dt^3} = w(t) \quad (\text{B.29})$$

Für das Systemrauschen gilt:

$$E[w(t)] = 0 \quad (\text{B.30})$$

$$E[w(t)w(\tau)] = \sigma_w^2 \delta(t - \tau) \quad (\text{B.31})$$

Das zeitkontinuierliche System wird durch folgende Gleichungen beschrieben:

$$\frac{d\mathbf{x}(t)}{dt} = \mathbf{F}\mathbf{x}(t) + \mathbf{w}(t) \quad (\text{B.32})$$

wobei

$$\mathbf{x}(t) = \left(\xi, \frac{d\xi}{dt}, \frac{d^2\xi}{dt^2} \right)^T \quad (\text{B.33})$$

$$\mathbf{F} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix} \quad (\text{B.34})$$

$$\mathbf{w}(t) = (0, 0, w(t))^T \quad (\text{B.35})$$

Die Herleitung der Systemgleichungen für den zeitdiskreten Fall erfolgt wie in Anhang B.1. Man erhält [Bar-Shalom & Li 93]:

$$\mathbf{x}(k+1) = \mathbf{A}\mathbf{x}(k) + \mathbf{w}(k) \quad (\text{B.36})$$

wobei

$$\mathbf{A} = \begin{pmatrix} 1 & T & \frac{1}{2}T^2 \\ 0 & 1 & T \\ 0 & 0 & 1 \end{pmatrix} \quad (\text{B.37})$$

$$\begin{aligned} \mathbf{Q} &= E[\mathbf{w}(k)\mathbf{w}(k)^T] \\ &= \begin{pmatrix} \frac{1}{20}T^5 & \frac{1}{8}T^4 & \frac{1}{6}T^3 \\ \frac{1}{8}T^4 & \frac{1}{3}T^3 & \frac{1}{2}T^2 \\ \frac{1}{6}T^3 & \frac{1}{2}T^2 & T \end{pmatrix} \sigma_w^2 \end{aligned} \quad (\text{B.38})$$

Die Meßgleichung lautet:

$$\mathbf{y}(k) = \mathbf{C}\mathbf{x}(k) + \mathbf{v}(k) \quad (\text{B.39})$$

mit

$$\mathbf{C} = \begin{pmatrix} 1 & 0 & 0 \end{pmatrix} \quad (\text{B.40})$$

$$\begin{aligned} \mathbf{R} &= E[\mathbf{v}(k)\mathbf{v}(k)^T] \\ &= \sigma_v^2 \end{aligned} \quad (\text{B.41})$$

Die Initialisierung des Filters wird wie in [Meyer & Bouthemy 92] zum Zeitpunkt $2T$ durchgeführt:

$$\hat{\mathbf{x}}^+(2) = \left(y(2), \frac{\frac{3}{2}y(2) - 2y(1) + \frac{1}{2}y(0)}{T}, \frac{y(2) - 2y(1) + y(0)}{T^2} \right)^T \quad (\text{B.42})$$

$$\hat{\mathbf{P}}^+(2) = \begin{pmatrix} \sigma_v^2 & \frac{3\sigma_w^2}{2T} & \frac{\sigma_w^2}{T^2} \\ \frac{3\sigma_v^2}{2T} & \frac{T^3\sigma_w^2}{2T} + \frac{13\sigma_v^2}{2T^2} & \frac{9T^2\sigma_w^2}{40} + \frac{6\sigma_v^2}{T^3} \\ \frac{\sigma_v^2}{T^2} & \frac{9T^2\sigma_w^2}{40} + \frac{6\sigma_v^2}{T^3} & \frac{23T\sigma_w^2}{30} + \frac{6\sigma_v^2}{T^4} \end{pmatrix} \quad (\text{B.43})$$

Die Initialisierung ist so gewählt, daß die Schätzfehler mittelwertfrei sind. Dies soll im folgenden bewiesen werden. Die ersten drei Messungen lauten:

$$y(0) = x_1(0) + v(0) \quad (\text{B.44})$$

$$y(1) = x_1(0) + Tx_2(0) + \frac{T^2}{2}x_3(0) + w_1(0) + v(1) \quad (\text{B.45})$$

$$\begin{aligned} y(2) &= x_1(0) + 2Tx_2(0) + 2T^2x_3(0) + w_1(0) + Tw_2(0) + \frac{T^2}{2}w_3(0) \\ &\quad + w_1(1) + v(2) \end{aligned} \quad (\text{B.46})$$

Durch Einsetzen der Meßwerte in Gl. (B.42) erhält man:

$$\hat{x}_1^+(2) = x_1(2) + v(2) \quad (\text{B.47})$$

$$\begin{aligned} \hat{x}_2^+(2) &= x_2(0) + 2Tx_3(0) - \frac{1}{2T}w_1(0) + \frac{3}{2}w_2(0) + \frac{3T}{4}w_3(0) + \frac{3}{2T}w_1(1) \\ &\quad + \frac{1}{T} \left[\frac{1}{2}v(0) - 2v(1) + \frac{3}{2}v(2) \right] \end{aligned} \quad (\text{B.48})$$

$$\begin{aligned} \hat{x}_3^+(2) &= x_3(0) + \frac{1}{T^2}w_1(0) + \frac{1}{T}w_2(0) + \frac{1}{2}w_3(0) + \frac{1}{T^2}w_1(1) \\ &\quad + \frac{1}{T^2} [v(0) - 2v(1) + v(2)] \end{aligned} \quad (\text{B.49})$$

Die Systemgrößen bei $k = 2$ lauten:

$$x_2(2) = x_2(0) + 2Tx_3(0) + w_2(0) + Tw_3(0) + w_2(1) \quad (\text{B.50})$$

$$x_3(2) = x_3(0) + w_3(0) + w_3(1) \quad (\text{B.51})$$

Damit ergeben sich die Schätzfehler:

$$\hat{x}_1^+(2) - x_1(2) = w(1) \quad (\text{B.52})$$

$$\begin{aligned} \hat{x}_2^+(2) - x_2(2) &= -\frac{1}{2T}w_1(0) + \frac{1}{2}w_2(0) - \frac{T}{4}w_3(0) + \frac{3}{2T}w_1(1) - w_2(1) \\ &\quad + \frac{1}{T} \left[\frac{1}{2}v(0) - 2v(1) + \frac{3}{2}v(2) \right] \end{aligned} \quad (\text{B.53})$$

$$\begin{aligned} \hat{x}_3^+(2) - x_3(2) &= -\frac{1}{T^2}w_1(0) + \frac{1}{T}w_2(0) - \frac{1}{2}w_3(0) + \frac{1}{T^2}w_1(1) - w_3(1) \\ &\quad + \frac{1}{T^2} [v(0) - 2v(1) + v(2)] \end{aligned} \quad (\text{B.54})$$

Somit ist bewiesen, daß die Schätzfehler mittelwertfrei sind. Aus den Gleichungen (B.38), (B.41) und (B.52)–(B.54) folgt für die Kovarianzmatrix von $\hat{\mathbf{x}}^+(2)$:

$$\begin{aligned} \hat{\mathbf{P}}^+(2) &= E[\hat{\mathbf{x}}^+(2)\hat{\mathbf{x}}^+(2)^T] \\ &= \begin{pmatrix} \sigma_v^2 & \frac{3\sigma_v^2}{2T} & \frac{\sigma_v^2}{T^2} \\ \frac{3\sigma_v^2}{2T} & \frac{T^3\sigma_w^2}{12} + \frac{13\sigma_v^2}{2T^2} & \frac{9T^2\sigma_w^2}{40} + \frac{6\sigma_v^2}{T^3} \\ \frac{\sigma_v^2}{T^2} & \frac{9T^2\sigma_w^2}{40} + \frac{6\sigma_v^2}{T^3} & \frac{23T\sigma_w^2}{30} + \frac{6\sigma_v^2}{T^4} \end{pmatrix} \end{aligned} \quad (\text{B.55})$$

Für das Meßrauschen wird wie in Anhang B.1 $\sigma_v^2 = 2$ gewählt. In [Bar-Shalom & Li 93] wird empfohlen σ_w^2 so einzustellen, daß die Änderung der Beschleunigung über eine Periode T in der Größenordnung von $\sqrt{T\sigma_w^2}$ liegt. Bei der Initialisierung wurde angenommen, daß die Beschleunigungsänderung im Mittel zu einer Änderung der Geschwindigkeit um 1 [Bildpunkte/s] in einer Abtastperiode führt. Daraus folgt $\sigma_w^2 = 15625$.

Literaturverzeichnis

- [Adelson & Bergen 86] E. H. Adelson, J. R. Bergen. *The extraction of spatiotemporal energy in human and machine vision*. Proc. IEEE Workshop on Visual Motion, Charleston, 1986, 151–156.
- [Adiv 85] G. Adiv. *Determining three-dimensional motion and structure from optical flow generated by several moving objects*. IEEE Transactions on Pattern Analysis and Machine Intelligence 7 (1985) 384–401.
- [Aggarwal & Nandhakumar 88] J. K. Aggarwal, N. Nandhakumar. *On the computation of motion from sequences of images—a review*. Proceedings of the IEEE 76 (1988) 917–935.
- [Anandan 89] P. Anandan. *A computational framework and an algorithm for the measurement of visual motion*. International Journal of Computer Vision 2 (1989) 283–310.
- [Ayer et al. 94] S. Ayer, P. Schroeter, J. Bigün. *Segmentation of moving objects by robust motion parameter estimation*. Computer Vision: 3rd European Conference on Computer Vision 1994, J.-O. Eklund (ed.), Stockholm: Springer-Verlag, 1994, 316–327.
- [Bar-Shalom & Li 93] Y. Bar-Shalom, X.-R. Li. *Estimation and tracking: Principles, techniques, and software*. Boston: Artech House, 1993.
- [Barnard & Thompson 80] S. T. Barnard, W. B. Thompson. *Disparity analysis of images*. IEEE Transactions on Pattern Analysis and Machine Intelligence 2 (1980) 333–340.
- [Barron et al. 94] J. L. Barron, D. J. Fleet, S. S. Beauchemin. *Performance of optical flow techniques*. International Journal of Computer Vision 19 (1994) 43–77.
- [Bartneck 87] N. Bartneck. *Ein Verfahren zur Umwandlung der ikonischen Bildinformation digitalisierter Bilder in Datenstrukturen zur Bildauswertung*. Ph.D. thesis, Technische Universität Carolo-Wilhelmina, Braunschweig, 1987.
- [Bennamoun et al. 94] M. Bennamoun, B. Boashash, A. Bower. *Object contour extraction from motion parameters*. Second Australian and New Zealand Conference on Intelligent Information Systems, Brisbane, 1994, 337–341.

- [Bhanu & Burger 88] B. Bhanu, W. Burger. *Approximation of displacement fields using wavefront region growing*. Computer Vision Graphics Image Processing 41 (1988) 306–322.
- [Black & Jepson 96] M. J. Black, A. D. Jepson. *Estimating optical flow in segmented images using variable-ordered parametric models with local deformations*. IEEE Transactions on Pattern Analysis and Machine Intelligence 18 (1996) 972–986.
- [Bober & Kittler 94] M. Bober, J. Kittler. *A Hough transform based hierarchical algorithm for motion segmentation and estimation*. TVIPMOR 4th International Workshop 1994, V. Capellini (ed.), Florenz, Italien: Elsevier Science Amsterdam, 1994, 335–342.
- [Bouthemy & François 93] P. Bouthemy, E. François. *Motion segmentation and qualitative dynamic scene analysis from an image sequence*. International Journal of Computer Vision 10 (1993) 157–182.
- [Brauckmann et al. 94] M. E. Brauckmann, C. Goerick, J. Groß, T. Zielke. *Towards all around automatic visual obstacle sensing for cars*. Proc. Intelligent Vehicles Symposium, Paris, 1994, 79–84.
- [Broida & Chellappa 91] T. J. Broida, R. Chellappa. *Estimating the kinematics and structure of a rigid object from a sequence of monocular images*. IEEE Transactions on Pattern Analysis and Machine Intelligence 13 (1991) 497–513.
- [Burt et al. 83] P. J. Burt, C. Yen, X. Xu. *Multi resolution flow through motion analysis*. Proc. Conference on Computer Vision and Pattern Recognition, Washington, 1983, 246–252.
- [Chen & Chang 92] L.-H. Chen, S. Chang. *A video tracking system with adaptive predictors*. Pattern Recognition 25 (1992) 1171–1180.
- [Chow & Aggarwal 77] W. K. Chow, J. K. Aggarwal. *Computer analysis of planar curvilinear moving images*. IEEE Transactions on Computers 26 (1977) 179–185.
- [Deo 74] N. Deo. *Graph theory with applications to engineering and computer science*. Prentice-Hall, 1974.
- [Dickmanns 93] E. D. Dickmanns. *Bifocal vision for mobile robot control*. IEEE Symposium on Industrial Electronics, Budapest, 1993, 1–6.
- [Dickmanns 94] E. D. Dickmanns. *The 4-D approach to visual control of autonomous systems*. AIAA/NASA Conference on Intelligent Robots in Field, Factory, Service and Space, Houston, Texas, 1994, 1–11.

- [Dubuisson & Jain 95] M.-P. Dubuisson, A. K. Jain. *Contour extraction of moving objects in complex outdoor scenes*. International Journal of Computer Vision 14 (1995) 83–105.
- [Duda & Hart 73] R. O. Duda, P. E. Hart. *Pattern classification and scene analysis*. Wiley-Interscience, 1973.
- [Enkelmann et al. 94] W. Enkelmann, V. Gengenbach, W. Krüger, S. Rössle, W. Tölle. *Obstacle detection by real-time optical flow evaluation*. Proc. Intelligent Vehicles Symposium, Paris, 1994, 97–102.
- [Etoh & Shirai 93] M. Etoh, Y. Shirai. *Segmentation and 2D motion estimation by region fragments*. Proc. 4th International Conference on Computer Vision, Berlin, 1993, 192–199.
- [Fleet & Jepson 90] D. J. Fleet, A. D. Jepson. *Computation of component velocity from local phase information*. International Journal of Computer Vision 5 (1990) 77–104.
- [Fleet 92] D. J. Fleet. *Measurement of image velocity*. Norwell, Massachusetts: Kluwer Academic, 1992.
- [Foresti & Murino 94] G. L. Foresti, V. Murino. *Moving object recognition from an image sequence for autonomous driving*. TVIPMOR 4th International Workshop 1994, V. Capellini (ed.), Florenz, Italien: Elsevier Science Amsterdam, 1994, 383–390.
- [Förstner & Gülch 87] W. Förstner, E. Gülch. *A fast algorithm for detection and precise location of distinct points, corners, and circular features*. Proc. Intercommission Conference on Fast Processing of Photogrammetric Data, Interlaken, 1987, 281–305.
- [Franke & Kutzbach 96] U. Franke, I. Kutzbach. *Fast stereo based object detection for stop & go traffic*. Proc. Intelligent Vehicles Symposium, Tokyo, 1996, 339–344.
- [Geman & Geman 84] S. Geman, D. Geman. *Stochastic Relaxation, Gibbs distribution, and the Bayesian restoration of images*. IEEE Transactions on Pattern Analysis and Machine Intelligence 6 (1984) 721–741.
- [Gil et al. 96] S. Gil, R. Milanese, T. Pun. *Comparing features for target tracking in traffic scenes*. Pattern Recognition 29 (1996) 1285–1296.
- [Graefe 93] V. Graefe. *Vision for intelligent road vehicles*. Proc. Intelligent Vehicles Symposium, Tokyo, 1993, 135–140.
- [Healey 93] G. Healey. *Hierarchical segmentation based approach to motion analysis*. Image and Vision Computing 11 (1993) 570–576.

- [Heckbert 80] P. Heckbert. *Color image quantization for frame buffer display*. M.Phil. thesis, Massachusetts Institute of Technology, 1980.
- [Heeger 88] D. J. Heeger. *Optical flow using spatiotemporal filters*. International Journal of Computer Vision 1 (1988) 279–302.
- [Heisele & Ritter 95] B. Heisele, W. Ritter. *Obstacle detection based on color blob flow*. Proc. Intelligent Vehicles Symposium, Detroit, 1995, 282–286.
- [Heisele et al. 96] B. Heisele, H. Neef, W. Ritter, R. Schneider, G. Wanielik. *Object detection in traffic scenes by a colour video and radar data fusion approach*. Proc. First Australian Data Fusion Symposium, Adelaide, 1996, 48–52.
- [Heisele et al. 97] B. Heisele, U. Kreßel, W. Ritter. *Tracking non-rigid, moving objects based on color cluster flow*. Proc. Computer Vision and Pattern Recognition, San Juan, 1997, 253–257.
- [Heitz & Bouthemy 93] F. Heitz, P. Bouthemy. *Multimodal estimation of discontinuous optical flow using Markov random fields*. IEEE Transactions on Pattern Analysis and Machine Intelligence 15 (1993) 1217–1232.
- [Horn & Schunck 81] B. K. P. Horn, B. G. Schunck. *Determining optical flow*. Artificial Intelligence 17 (1981) 185–203.
- [Irani & Anandan 96] M. Irani, P. Anandan. *A unified approach to moving object detection in 2D and 3D scenes*. Proc. International Conference on Pattern Recognition, Vol. 1, Wien, 1996, 712–716.
- [Jain & Jain 81] J. R. Jain, A. K. Jain. *Displacement measurement and its application in interframe image coding*. IEEE Transactions on Communications 29 (1981) 1799–1808.
- [Jain & Nagel 79] R. Jain, H.-H. Nagel. *On the analysis of accumulative difference pictures from image sequences of real world scenes*. IEEE Transactions on Pattern Analysis and Machine Intelligence 1 (1979) 206–214.
- [Judd et al. 96] D. Judd, P. K. McKinley, A. K. Jain. *Large-scale parallel data clustering*. Proc. International Conference on Pattern Recognition, Vol. 1, Wien, 1996, 488–493.
- [Kahl et al. 78] D. J. Kahl, A. Rosenfeld, A. Danker. *Some experiments in point pattern matching*. Research report TR-690, University of Maryland, Maryland, 1978.
- [Kalivas & Sawchuk 91] D. S. Kalivas, A. A. Sawchuk. *A region matching motion estimation algorithm*. CVGIP: Image Understanding 54 (1991) 275–288.

- [Kearney et al. 87] J. K. Kearney, W. B. Thompson, D. L. Boley. *Optical flow estimation: An error analysis of gradient-based methods with local optimization*. IEEE Transactions on Pattern Analysis and Machine Intelligence 9 (1987) 229–244.
- [Kolb 95] T. Kolb. *Fahrzeugklassifikation mit neuronalen Netzen*. M.Phil. thesis, Universität Ulm, 1995.
- [Koller et al. 91] D. Koller, N. Heinze, H.-H. Nagel. *Algorithmic characterization of vehicle trajectories from image sequences by motion verbs*. Proc. Computer Vision and Pattern Recognition, Maui, 1991, 90–95.
- [Koller et al. 93] D. Koller, J. Weber, J. Malik. *Robust multiple car tracking with occlusion reasoning*. Research report UCB/CSD 93–780, Computer Science Division (EECS) University of California, Berkley, California, 1993.
- [Koller et al. 94] D. Koller, Q.-T. Luong, J. Malik. *Using binocular stereopsis for vision-based vehicle control*. Proc. Intelligent Vehicles Symposium, Paris, 1994, 237–242.
- [Krüger et al. 95] W. Krüger, W. Enkelmann, S. Rössle. *Real-time estimation and tracking of optical flow vectors for obstacle detection*. Proc. Intelligent Vehicles Symposium, Detroit, 1995, 304–309.
- [Legters & Young 82] G. R. Legters, T. Y. Young. *A mathematical model for computer image tracking*. IEEE Transactions on Pattern Analysis and Machine Intelligence 4 (1982) 583–594.
- [Levine & O’Handley 73] M. D. Levine, D. A. O’Handley. *Computer determination of depth maps*. Computer Graphics and Image Processing 2 (1973) 131–150.
- [Loffeld 90] O. Loffeld. *Estimationstheorie*. Vol. 2. R. Oldenbourg, 1990.
- [Longuet-Higgins & Prazdny 80] H. C. Longuet-Higgins, K. Prazdny. *The interpretation of a moving retinal image*. Proc. Roy. Soc. London, B. 208 (1980) 385–397.
- [Lucas 84] B. D. Lucas. *Generalized image matching by the method of differences*. Ph.D. thesis, Carnegie-Mellon University, 1984.
- [MacQueen 67] J. MacQueen. *Some methods for classification and analysis of multivariate observations*. Proceedings of the 5th Berkley Symposium on Mathematics Statistics and Probability, 1967, 281–297.
- [Mandler & Oberländer 90] E. Mandler, M. Oberländer. *One pass encoding of connected components in multi-valued images*. Proc. International Conference on Pattern Recognition, Atlantic City, 1990, 64–69.

- [Markandey & Flinchbaugh 90] V. Markandey, B. E. Flinchbaugh. *Multispectral constraints for optical flow computation*. Proc. 3rd International Conference on Computer Vision, Osaka, 1990, 38–41.
- [McIntosh & Mutch 88] J. H. McIntosh, K. M. Mutch. *Matching Straight Lines*. Computer Vision Graphics Image Processing 43 (1988) 386–408.
- [Meyer & Bouthemy 94] F. G. Meyer, P. Bouthemy. *Region-based tracking using affine motion models in long image sequences*. CVGIP: Image Understanding 60 (1994) 119–140.
- [Meyer & Bouthemy 92] F. Meyer, P. Bouthemy. *Region-based tracking in an image sequence*. Research report 1723, IRISA/INRIA, Rennes Cedex, France, 1992.
- [Mitiche & Habelrih 89] A. Mitiche, G. Habelrih. *Interpretation of straight line correspondences using angular relations*. Pattern Recognition 22 (1989) 299–308.
- [Mitiche et al. 89] A. Mitiche, O. Faugeras, J. K. Aggarwal. *Counting Straight Lines*. Computer Vision Graphics Image Processing 47 (1989) 353–360.
- [Mitiche & Bouthemy 96] A. Mitiche, P. Bouthemy. *Computation and analysis of image motion: a synopsis of current problems and methods*. International Journal of Computer Vision 19 (1996) 29–55.
- [Nagel & Enkelmann 86] H.-H. Nagel, W. Enkelmann. *An investigation of smoothness constraints for the estimation of displacement vector fields from image sequences*. IEEE Transactions on Pattern Analysis and Machine Intelligence 8 (1986) 565–593.
- [Nagel 78] H. Nagel. *Formation of an object concept by analysis of systematic time variations in the optically perceptible environment*. Computer Graphics and Image Processing 7 (1978) 149–194.
- [Nakayama & Loomis 74] K. Nakayama, J. M. Loomis. *Optical velocity patterns, velocity-sensitive neurons, and space perception: A hypothesis*. Perception 3 (1974) 63–80.
- [Odobez & Bouthemy 94] J. M. Odobez, P. Bouthemy. *Detection of multiple moving objects using multiscale MRF with camera motion compensation*. International Conference on Image Processing, 1994, 257–261.
- [Pal & Pal 93] N. R. Pal, S. K. Pal. *A review on image segmentation techniques*. Pattern Recognition 26 (1993) 1277–1294.
- [Price & Reddy 77] K. Price, D. R. Reddy. *Change detection and analysis in multispectral images*. Proc. Int. Joint Conf. Artificial Intell., 1977, 619–625.

- [Ranke et al. 97] M. Ranke, R. Ott, B. Heisele. *Farbsegmentierung mit Hopfield-Netzen*. Technischer Bericht F3-97-005, Daimler-Benz AG, Ulm, 1997.
- [Rehrmann & Rothaar 97] V. Rehrmann, M. Rothaar. *Detection and tracking of moving objects in colour outdoor scenes*. Proceedings of the 30th Symposium on Automotive Technology and Automation, Florenz, Italien, 1997, 85–91.
- [Richter 81] M. Richter. *Einführung in die Farbmatrik*. zweite edn. Berlin: Walter de Gruyter, 1981.
- [Saito & Komatsu 94] T. Saito, T. Komatsu. *Extended block-matching algorithms for estimating multiple image motions*. International Conference on Image Processing, 1994, 735–739.
- [Schalkoff & McVey 82] R. J. Schalkoff, E. S. McVey. *A model and tracking algorithm for a class of video targets*. IEEE Transactions on Pattern Analysis and Machine Intelligence 4 (1982) 2–10.
- [Schmid 94] M. Schmid. *3D Erkennung von Fahrzeugen in Echtzeit aus monokularen Bildfolgen*. Ph.D. thesis, Universität der Bundeswehr München, 1994.
- [Schnörr 91] C. Schnörr. *Determining optical flow from irregular domains by minimizing quadratic functionals of certain classes*. IEEE Transactions on Pattern Analysis and Machine Intelligence 6 (1991) 25–38.
- [Schwarzinger et al. 92] M. Schwarzinger, T. Zielke, D. Noll, M. Brauckmann, W. von Seelen. *Vision-based car following: detection, tracking, and identification*. Proc. Intelligent Vehicles Symposium, Detroit, 1992, 24–29.
- [Shirai 87] Y. Shirai. *Three-dimensional computer vision*. Springer Verlag, 1987.
- [Singh 90] A. Singh. *An estimation-theoretic framework for image-flow computation*. Proc. 3rd International Conference on Computer Vision, Osaka, 1990, 168–177.
- [Skarbek & Koschan 94] W. Skarbek, A. Koschan. *Colour image segmentation - a survey* -. Research report 94-32, Technische Universität Berlin, Berlin, 1994.
- [Smith & Brady 94] S. M. Smith, J. M. Brady. *A scene segmenter; visual tracking of moving vehicles*. Engineering Applications of Artificial Intelligence 7 (1994) 191–204.
- [Smith & Brady 95] S. M. Smith, J. M. Brady. *ASSET-2: real-time motion segmentation and shape tracking*. IEEE Transactions on Pattern Analysis and Machine Intelligence 17 (1995) 814–820.

- [Snyder 91] M. A. Snyder. *On the mathematical foundations of smoothness constraints for the determination of optical flow and for surface reconstruction*. IEEE Transactions on Pattern Analysis and Machine Intelligence 13 (1991) 1105–1114.
- [Tekalp 95] A. M. Tekalp. *Digital Video Processing*. Prentice Hall, 1995.
- [Thompson & Pong 90] W. B. Thompson, T.-C. Pong. *Detecting moving objects*. International Journal of Computer Vision 4 (1990) 39–57.
- [Tomasi & Kanade 92] C. Tomasi, T. Kanade. *Shape and motion from image streams under orthography: a factorization method*. International Journal of Computer Vision 9 (1992) 137–154.
- [Tretiak & Pastor 84] O. Tretiak, L. Pastor. *Velocity estimation from image sequences with second order differential operators*. Proc. International Conference of Pattern Recognition and Image Processing, Montreal, 1984, 16–19.
- [Ullman 79] S. Ullman. *The interpretation of visual motion*. Cambridge, Massachusetts: The MIT Press, 1979.
- [Verri & Poggio 89] A. Verri, T. Poggio. *Motion field and optical flow: qualitative properties*. IEEE Transactions on Pattern Analysis and Machine Intelligence 11 (1989) 490–498.
- [Verri et al. 90] A. Verri, F. Girosi, V. Torre. *Differential techniques for optical flow*. Journal of the Optical Society of America A 7 (1990) 912–922.
- [Weng et al. 92] J. Weng, J. Ahuja, N. Huang. *Matching two perspective views*. IEEE Transactions on Pattern Analysis and Machine Intelligence 14 (1992) 806–825.
- [Wu 91] X. Wu. *Efficient statistical computations for optimal color quantization*. Graphics Gems II, J. Arvo (ed.). Academic Press, 1991.
- [Zielke et al. 92] T. Zielke, M. Brauckmann, W. von Seelen. *Intensity and edge-based symmetry detection applied to car-following*. Computer Vision: European Conference on Computer Vision 1992, (Lecture Notes in Computer Science 588), G. Sandini (ed.), Springer-Verlag, 1992, 865–873.
- [Zucker et al. 78] S. W. Zucker, E. V. Krishnamurthy, R. L. Haar. *Relaxation processes for scene labeling: Convergence, speed, and stability*. IEEE Transactions on Systems, Man, and Cybernetics 8 (1978) 41–48.