

# **FORTSCHRITTE BEI DER ANALYSE VON BEWEGTEN FARBILDERN UNTER DEM EINFLUSS VON NICHT- KOOPERATIVEN MESSSITUATIONEN**

## **Habilitationsschrift**

von J.-Prof. Dr.-Ing. Ayoub Kassim Al-Hamadi

geboren am 26. September 1970

in Beni-Hammad (Jemen)

zur Verleihung des akademischen Grades

**Doktor-Ingenieur habitatus  
(Dr.-Ing. habil.)**

genehmigt von der Fakultät für Elektrotechnik und Informationstechnik der  
Otto-von-Guericke-Universität Magdeburg am 05. Mai 2010

Gutachter:

Prof. Dr.-Ing. habil. Bernd Michaelis, Universität Magdeburg

Prof. Dr.-Ing. Klaus Tönnies, Universität Magdeburg

Prof. Dr.rer.nat. Heiko Neumann, Universität Ulm



## Vorwort

Die hier vorliegende Arbeit wurde während meiner Tätigkeit als wissenschaftlicher Mitarbeiter am Institut für Elektronik, Signalverarbeitung und Kommunikationstechnik der Otto-von-Guericke-Universität Magdeburg unter der Betreuung von Herrn Prof. Dr.-Ing. habil. Bernd Michaelis angefertigt. An dieser Stelle möchte ich mich bei Prof. Michaelis für seine Anregung zu dieser Arbeit, seine wertvollen Hinweise und Diskussionsbereitschaft sowie für die freundliche Unterstützung bedanken.

Mein besonderer Dank gilt Herrn Prof. Dr.-Ing. Klaus Tönnies und Prof. Dr. rer.nat. Heiko Neumann für ihr Interesse an dieser Arbeit und die wertvollen Bemerkungen bei der Strukturierung und Darstellung der unterschiedlichen Schwerpunkte der Arbeit sowie die Übernahme der Begutachtung der Arbeit.

Mein Dank gebührt allen Mitarbeitern des IESK für ihre Hilfsbereitschaft insbesondere der Arbeitsgruppe von Prof. Michaelis für das gute Arbeitsklima. An dieser Stelle möchte ich meinen Dank den Herren Dr.-Ing. Gerald Krell und Dipl.-Ing. Robert Niese, Dr. Michael Tornow für die jahrelange Zusammenarbeit und freundliche Hilfsbereitschaft aussprechen. Mein besonderer Dank gilt auch meinen Diplomanden und Doktoranden, die ich für die Realisierung meiner Ideen und für die Bearbeitung von Teilaufgaben eingesetzt habe. Hier möchte ich Herrn Dipl.-Ing. Tobias Senst und Frau Dipl.-Ing. Ulrike Homberg danken.

Besonderer Dank gilt meinen persönlichen Korrekturlesern Robert Niese, Gerald Krell und Axel Panning, die mir gezeigt haben, dass es außer Hilfsverben auch noch andere Wörter gibt und viel von einem Informatiker in einem Informationstechnologen steckt. Des Weiteren möchte ich mich bei allen bedanken, die mich während der Habilitation unterstützt haben. Hier gebührt meinen Dank vor allem meiner Frau Asma Al-Shehabi, die mir durch ihre Hilfe, Geduld und ihr Verständnis für die Arbeitszeit am Wochenende diese Habilitation ermöglicht hat. Mein ausdrücklicher Dank gebührt darüber hinaus Herrn Prof. Dr.-Ing. habil. Franz Güttler, da er mich von der Bedeutung der Informationstechnologie überzeugt hat.

Schließlich danke ich allen meinen Freunden, die mir über die Jahre liebevoll zur Seite standen und mich während meines Aufenthalts in Magdeburg bei allen möglichen Problemen mit Rat unterstützt haben. Mein Aufenthalt in Magdeburg wie auch meine berufliche Entwicklung an der Otto-von-Guericke-Universität werden mir in guter Erinnerung bleiben und hinterlassen tiefe Spuren in meiner Persönlichkeit.

Magdeburg, 05. Mai 2010.

Ayoub Kassim Al-Hamadi

## Kurzfassung

Die Analyse von bewegten Bildern hat neben den aktuellen Forschungen im Bereich der intelligenten Überwachungstechnik, Umfelderkennung und Fahrerassistenzsysteme einen wesentlichen Einfluss auf Mensch-Maschine-Interaktionssysteme. Um sich bewegende Objekte in einer monokularen Farbbildsequenz automatisch erfassen und verfolgen zu können, wurden vier wesentliche Forschungsaspekte im Rahmen der vorliegenden Habilitationsarbeit thematisiert. Diese haben einen Beitrag zur Reduzierung der Komplexität und somit zur Lösung des Korrespondenzproblems in komplexen Situationen geleistet, insbesondere durch folgende Innovationen:

- Zum einen wurde ein systematischer Rahmen zur Entwicklung genauer intensitätsbasierter Ansätze auf der Grundlage der optischen Flussverfahren und Korrelationsverfahren vorgestellt. Dieser erlaubt eine Verbesserung der zurzeit bekannten Schätzverfahren, die in der Literatur zu finden sind. Gegenstand dieser Arbeiten waren die intensitätsbasierten optischen Flussverfahren und ihre Verwendung zur Selektion und Verfolgung mehrerer bewegter Objekte. Die hier vorgeschlagenen Lösungsansätze wurden für die Zustandsbeschreibung eines Knotens im Zustandsmodell bei der Datenfusion optimal eingesetzt. In einem weiteren Schritt wurde eine adaptive Systemstruktur auf Grundlage der Korrelationen zwischen Bildausschnitten entwickelt, die die Farbinformation optimal zur Verbesserung der Störsicherheit ausnutzt.
- Zum zweiten wurde ein neues Paradigma zur merkmalsbasierten Korrespondenzermittlung in Farbbildsequenzen vorgeschlagen, das sich vom klassischen Matching durch seine hierarchische Struktur und Adaptivität unterscheidet. Dieses strukturelle Konzept segmentiert, verfolgt und erkennt nicht nur fest gewählte Regionen, sondern auch beliebige *veränderliche* Objekte in Bildfolgen automatisch. Zusätzlich ermöglicht es die schnelle Berechnung und macht erstmals dieses äußerst robuste Matching für die Bewegungsbestimmung und Verhaltensbeschreibung praktisch einsetzbar. Die Echtzeitfähigkeit der Berechnung durch dieses Matching-Paradigma kann sich aus der inhärenten Parallelität des angewandten Teilverfahrens und seiner Schnelligkeit ergeben.
- Zum dritten wurde ein Partikelfiltersystem in Verbindung mit einer Datenassoziation für die Multi-Objektverfolgung entwickelt, dessen Verarbeitungsstufen die ursprünglichen Schritte eines Partikelfilters beinhalten, die in dieser Arbeit aber modifiziert werden. Die Modifikationen beziehen sich zuerst auf die Prädiktion, die durch ein erweitertes *adaptives Bewegungsmodell* erreicht wird. Zusätzlich wurde der ursprüngliche Ablauf des Partikelfilters um einen 3-stufigen Assoziationsschritt mit Datenauschluss, Datenzuweisung und Datenverwaltung ergänzt, um die Problematik der Fluktuationserkennung und den Umgang mit Objektinteraktionen zu lösen. Dadurch wird eine zuverlässige Bestimmung von Bewegungstrajektorien in Konfliktsituationen ermöglicht.
- Zuletzt wurde ein neuer Lösungsweg auf Grundlage der Datenfusion entwickelt, der eine Zustandschätzung und robuste Datenzuweisung von Objekten in Konfliktsituationen ermöglicht. Die Struktur für die Datenfusion nutzt die Stärken der intensitäts- und merkmalsbasierten Verfahren so aus, dass die Schwächen in beiden Verfahren kompensiert werden können. Die Grundlagen dafür sind die Bayes-Schätzung, die Fuzzy-Logik und die vorgestellten Bewegungsanalyseverfahren.

Ausgehend von diesen Aspekten wurde die vorliegende Arbeit in sechs Kapitel gegliedert, die die Basis für weitere Forschungsarbeiten im Bereich der Analyse von bewegten Bildern unter dem Einfluss von nicht kooperativen Messsituationen bilden.

## Abstract

The analysis of image sequences has an essential influence on the current research in the area of intelligent surveillance, environment perception, driver assistance systems and human computer interaction. In the context of detecting and tracking moving objects in color image sequences, four essential research aspects were treated in the postdoctoral lecture qualification thesis, which describe the contributions and innovations for solving the correspondence problem in complex image measuring situations. Moreover, the reduction of complexity is also an important aspect in this thesis. These contributions are listed as:

- The first contribution, a systematic framework of intensity-based approaches is introduced and modified. The focus in this part of research is to exploit optical flow approaches and its application on the automatic segmentation and tracking of several moving objects. The modified approaches are used as basis for the state description of nodes in the state model of data fusion. In the next step, an adaptive system is developed based on correlation of the image regions in conjunction with principle components analysis and neural networks. This exploits the color information optimally to improve the results under the influence of non-cooperative measuring situations.
- Secondly, a new paradigm is proposed for the feature-based correspondence analysis that differs from the classic matching through its hierarchical structure. Feature-based matching frequently suffers from accuracy and performance due to the higher number of image features which needs to be matched. This suggested paradigm is based on corresponding color segments which provide the basis for the hierarchical feature extraction and correlation. Unlike other approaches, the generation of color segments is accompanied by segmentation of moving objects. Using segmentation of moving objects, it is possible to substantially reduce the probable set of image features which may potentially match. Thus, a high robustness is reached under the influence of disturbed real image situations (e.g. brightness variations and shadows). In addition, it enables the fast determination of motion trajectories of objects and makes this developed feature-based matching practically useable for the crowd behavior analysis.
- In the third contribution, a particle filter system is developed in connection with the data association for the multi-object tracking, in which the processing steps include the original steps of the particle filter, which are modified in this work. In this context, a system of independent particle filters and an adaptive motion model is used which allows to separately handle the moving objects in conflict situations. To address the problems of the fluctuation detection and object interactions a data association step is suggested which includes data exclusion, data allocation and data administration. Furthermore, the proposed approach enables us to recognize conflicting image situations and to adapt the particle filters for these situations.
- Finally, a new solution is developed on the basis of the data fusion which enables a state estimation and a robust data assignment of real object in conflict situations. The concept for the data fusion exploits the strengths of the intensity- and feature-based procedure so that the weaknesses can be compensated in both procedures. The basis for this data fusion is the Baye's estimation, Fuzzy logic and the suggested intensity- and feature-based tracking analysis procedures.

Based on these aspects and contributions, the work is organized in six chapters establishing a fundament for the analysis of image sequences under the influence of non-cooperative measuring situations and opens the new possibilities for the future research.



# Inhaltsverzeichnis

VORWORT.....	I
INHALTSVERZEICHNIS .....	V
<b>1.          EINLEITUNG.....</b>	<b>1</b>
<b>1.1    Einordnung der eignen Arbeiten und Zielstellung .....</b>	<b>2</b>
<b>1.2    Gliederung der Arbeit .....</b>	<b>8</b>
<b>2.          INTENSITÄTSBASIERTE ANSÄTZE.....</b>	<b>11</b>
<b>2.1    Verfahren nach dem Optischen Fluss .....</b>	<b>12</b>
2.1.1    Horn-Schunck-Verfahren .....	14
2.1.2    Lucas-Kanade-Verfahren .....	17
2.1.3    Pyramidales LK-Verfahren .....	19
2.1.4    Bewertung der vorgestellten Verfahren .....	20
2.1.5    Verbesserung des PLK-Verfahrens .....	25
2.1.6    Kurzfassung und Diskussion.....	34
<b>2.2    Korrelationsverfahren.....</b>	<b>35</b>
2.2.1    Vorgeschlagene Systemstruktur zur Bewegungsschätzung .....	36
2.2.1.1    Aufbereitung der Farbinformation.....	36
2.2.1.2    Bewegungsanalyse für die selektierten Objekte .....	39
2.2.2    Experimentelle Ergebnisse mit der Systemstruktur .....	42
2.2.2.1    Einfluss bildspezifischer Störungen .....	44
2.2.3    Schlussfolgerung /Diskussion .....	49
2.2.4    Zusammenfassung.....	51
<b>3.          MATCHING-PARADIGMA .....</b>	<b>53</b>
<b>3.1    Klassisches Multi-Matching.....</b>	<b>54</b>
3.1.1    Farbsegmentierungsalgorithmen .....	54
3.1.1.1    Color-Structure-Code-Algorithmus.....	55
3.1.1.2    Forderungen und Vergleiche an die Segmentierungsverfahren .....	57
3.1.2    Korrespondenzanalyse mittels Multimatchs .....	59
<b>3.2    Ziele und Einordnung der eigenen Arbeiten .....</b>	<b>60</b>
<b>3.3    Bewegungsbasierter Segmentierungsansatz.....</b>	<b>62</b>
3.3.1    MD-Ansatz mit Nachbearbeitung .....	63
3.3.2    Schattenfilterung.....	66
<b>3.4    Hierarchische Merkmalsextraktion .....</b>	<b>70</b>
3.4.1    Merkmalsebene I - Motion-Blobs .....	71
3.4.2    Merkmalsebene II - CSC-Patches .....	73
<b>3.5    Hierarchisches merkmalsbasiertes Multi-Matching .....</b>	<b>76</b>
3.5.1    m:n Motion-Blob-Matching.....	77
3.5.2    1:1 CSC- Multi-Matching .....	80
3.5.3    Gruppierung der ermittelten Verschiebungsvektoren .....	88
3.5.4    Erweiterungsmöglichkeiten auf eine Meta-Korrelation .....	90
<b>3.6    Verbesserte Störsicherheit durch Optimierungsschritte .....</b>	<b>93</b>
3.6.1    Gütefunktion zur Gewichtsoptimierung.....	95
3.6.1.1    Threshold-Accepting-Algorithmus.....	97
3.6.1.2    Getrennte qualitative Bewertung der geschätzten Gewichtsvektoren.....	103
3.6.2    Erweiterung des Ähnlichkeitskriteriums.....	105
<b>3.7    Validierung des Gesamtmatching-Prozesses.....</b>	<b>110</b>
<b>3.8    Experimentelle Ergebnisse.....</b>	<b>120</b>
3.8.1    Einfluss von nicht-kooperativen Messsituationen.....	120
3.8.2    Sequenz II – Geringer Straßenverkehr, Schattenwurf.....	127

# INHALTSVERZEICHNIS

---

3.8.3	Sequenz III – Dichter Straßenverkehr .....	129
<b>3.9</b>	<b>Auswirkung verschiedener Segmentierungsgrade .....</b>	<b>132</b>
<b>3.10</b>	<b>Diskussion, Schlussfolgerung und Grenzen.....</b>	<b>133</b>
3.10.1	Besonderheit des vorgeschlagenen Matching-Paradigmas.....	135
3.10.2	Grenzen des Verfahrens .....	135
<b>3.11</b>	<b>Zusammenfassung und Ausblick.....</b>	<b>137</b>
<b>4.</b>	<b>OBJEKTVERFOLGUNG AUF DER BASIS VON DATENASSOZIATIONEN .....</b>	<b>139</b>
<b>4.1</b>	<b>Intuitive Einführung.....</b>	<b>139</b>
4.1.1	Kalman-Filter .....	141
4.1.2	Partikelfilter .....	144
<b>4.2</b>	<b>Einordnung und Ziele der eigenen Arbeiten.....</b>	<b>147</b>
<b>4.3</b>	<b>Partikelfiltersystem mit Datenassoziation .....</b>	<b>149</b>
4.3.1	Automatische Erfassung von Bildregionen.....	150
4.3.2	Zustandsbeschreibung.....	151
4.3.3	Korrespondenzanalyse und Datenassoziation .....	153
4.3.3.1	Adaptives Bewegungsmodell .....	155
4.3.3.2	Bestimmung des Beobachtungsmodells .....	157
4.3.4	Datenassoziation .....	159
4.3.4.1	Datenausschluss.....	162
4.3.4.2	Datenzuweisungen.....	164
4.3.4.3	Datenverwaltung .....	169
4.3.5	Verlauf einer Iteration des Partikelfiltersystems .....	173
<b>4.4</b>	<b>Analyse.....</b>	<b>175</b>
4.4.1	Bildsequenzen ohne Konfliktsituationen .....	175
4.4.2	Bildsequenzen mit Konfliktsituationen.....	177
<b>4.5</b>	<b>Schlussfolgerung .....</b>	<b>180</b>
<b>4.6</b>	<b>Zusammenfassung .....</b>	<b>182</b>
<b>5.</b>	<b>DATENFUSION AUF DER BASIS VON PARTIKELFILTERN .....</b>	<b>184</b>
<b>5.1</b>	<b>Intuitive Einführung und Einordnung der eigenen Arbeiten .....</b>	<b>184</b>
<b>5.2</b>	<b>Hybride Systemstruktur zur Multi-Objektverfolgung .....</b>	<b>185</b>
5.2.1	MB-Segmentierung.....	186
5.2.2	Datenaufbereitung.....	186
<b>5.3</b>	<b>Korrespondenzanalyse .....</b>	<b>189</b>
5.3.1	Zustandsmodell.....	191
5.3.2	Fuzzy-Bewegungsmodell.....	192
5.3.2.1	Fuzzifizierung.....	194
5.3.2.2	Fuzzyregelung .....	197
5.3.2.3	Defuzzifizierung und stochastische Entscheidung .....	201
5.3.3	Das Beobachtungsmodell.....	203
5.3.4	Sampling .....	209
<b>5.4</b>	<b>Untersuchungen und Ergebnisse.....</b>	<b>210</b>
5.4.1	Szene ohne Konfliktsituation .....	210
5.4.2	Szenen mit einfachen Konfliktsituationen .....	211
5.4.3	Szenen mit gehäuften Konfliktsituationen .....	215
<b>5.5</b>	<b>Schlussfolgerung und Zusammenfassung.....</b>	<b>218</b>
5.5.1	Diskussion der vorgeschlagenen Datenfusionsstruktur.....	219
5.5.2	Besonderheit .....	222
5.5.3	Grenzen.....	223
5.5.4	Zusammenfassung und Ausblick .....	223
<b>6</b>	<b>ZUSAMMENFASSUNG DER GESAMTARBEIT .....</b>	<b>225</b>
<b>7.</b>	<b>ABKÜRZUNGSVERZEICHNIS .....</b>	<b>233</b>

## INHALTSVERZEICHNIS

---

<b>8.</b>	<b>ANHANG.....</b>	<b>237</b>
<b>8.1</b>	<b>Der Einfluss von nicht-kooperativen Messsituationen .....</b>	<b>237</b>
<b>8.2</b>	<b>Normierungskoeffizienten für den Ellipsenabstandsfehler .....</b>	<b>243</b>
<b>8.3</b>	<b>Untersuchungen zum Vergleich von Farbhistogrammen .....</b>	<b>244</b>
<b>8.4</b>	<b>Beispielsequenzen: Optimierung der Gewichte .....</b>	<b>247</b>
<b>8.5</b>	<b>Beispielsequenzen: Einfache Konfliktsituationen.....</b>	<b>250</b>
8.5.1	Szene 1 .....	250
8.5.2	Szene 2 .....	253
<b>8.6</b>	<b>Beispielsequenzen: Gehäufte Konfliktsituationen .....</b>	<b>256</b>
8.6.1	Szene 3 .....	256
8.6.2	Szene 4 .....	258
8.6.3	Szene 5 .....	261
<b>9.</b>	<b>LITERATURVERZEICHNIS.....</b>	<b>265</b>



### 1. Einleitung

Das visuelle System des Menschen besitzt sehr hoch entwickelte Fähigkeiten zur Verfolgung und Erkennung von Objekten. So kann man mühelos ein Objekt im dichten Straßenverkehr verfolgen oder ein Auto eine bestimmte Fahrbahn entlang steuern. Diese Beispiele veranschaulichen, wie mit Hilfe der Bewegung zwischen zu beobachtendem Objekt und Hintergrund unterschieden wird. Die Ergebnisse der Objektverfolgung müssen interpretiert werden, um gegebenenfalls eine Aktion, wie z.B. eine Kopfdrehung oder die Lenkbewegung, auszulösen. Diese Aktion des Menschen steht am Ende einer Reihe physikalischer, physiologischer und psychologischer Prozesse im visuellen System des Menschen.

Der Wunsch, Objekte mit Hilfe technischer Geräte ähnlich wie beim biologischen System zu verfolgen, hat mit der zunehmenden Verfügbarkeit preiswerter Farbkameras und der Leistungssteigerung der Bildaufnahme- und Rechentechnik in den letzten Jahren stetig an Bedeutung gewonnen. Somit hat sich die Objektverfolgung zu einem wichtigen Forschungsschwerpunkt der Bildverarbeitung entwickelt und als Teilgebiet der Analyse von Bildsequenzen etabliert. Sie ist für viele Anwendungen des maschinellen Sehens aufgrund der Entwicklung einer Vielfalt leistungsfähiger Algorithmen sowie fallender Hardwarekosten von großer Bedeutung. Aktuelle Anwendungen reichen von der Bilddatenkompression [Ohm 95, Torres 00, Ahmad 07a, 07b, 08] über Einsatzgebiete in der Überwachungstechnik [Huwer 98] bis hin zur Steuerung autonomer Fahrzeuge [Dickmanns 92] und Roboter. Somit erweisen sich die bekannten Methoden bei der industriellen Anwendung (z.B. Fahrerassistenzsysteme [Ulmer 94, Saneyoshi 96, Tornow 03], Mensch-Maschine-Interaktionen [Corradini 00, Yang 02, Hofmann 03, Elmezian 08a, 08b, 09, Niese 08] usw.) bzw. für medizinische Zwecke (z.B. Verhaltensanalyse, Ganganalyse [Calow 03], Gesichts- und Schmerzanalyse [Craig 85, Handcock 05, Niese 07b, Al-Hamadi 08b, Niese 09] usw.) aber auch für Wachstumsvorgänge u.ä. als gut geeignete Hilfsmittel.

Die in dieser Arbeit verwendete Definition der Objektverfolgung und -erkennung beschränkt sich auf die Beobachtung von Veränderungen in der Bildszene. Hierbei handelt es sich um eine Informationsverarbeitungsaufgabe sowohl aus Sicht der Mustererkennung als auch aus Sicht der künstlichen Intelligenz. Man versteht darunter das Beobachten von bestimmten zeitlichen Veränderungen in einer Szene. Häufig wird dabei die Aufmerksamkeit auf ein bestimmtes Objekt gerichtet und die Verfolgung dieses Objektes resultiert aus der zeitlichen Veränderung der Bildszene oder der Lage des Betrachters bezüglich der Szene.

Es darf allerdings nicht verallgemeinert werden, dass jede Veränderung im Verlauf einer Bildsequenz durch eine Bewegung verursacht wurde. Es können z.B. Beleuchtungsänderungen oder Reflexionen von sich bewegenden Objekten zu Bildinhaltsänderungen führen [Klette 90, Jahne 89, Al-Hamadi 01a, 02b]. Somit besteht das Problem, diese Bildinhaltsänderungen aufgrund direkter Objektbewegung von indirekt durch sie hervorgerufenen Bildinhaltsveränderungen zu unterscheiden. Prinzipiell werden die zeitlichen Veränderungen z.B. durch die Analyse der Bilddifferenz aufeinanderfolgender, möglichst (zeitlich) äquidistanter Bilder erfasst. Bildfolgen müssen nicht nur auf den für den Menschen sichtbaren Spektralbereich beschränkt sein, sondern können auch mit einer Infrarotkamera aufgenommen werden. Durch den Einsatz von leistungsfähigen

gen Speichermedien und hoher Rechenleistung sind die Analyse von längeren Bildsequenzen sowie die Berücksichtigung der Vorgeschichte der Bildinhaltsänderungen möglich [Mecke 99, Al-Hamadi 03g, 03f]. Allerdings ist dadurch zu erwarten, dass sich die Anforderungen hinsichtlich der Komplexität und der Genauigkeit bei der Bestimmung von Bewegungsparametern erhöhen.

Die Interpretation der Beobachtung der Bewegung, die sehr stark von dem Kontext und der Vorgeschichte sowie den jeweiligen Objekten abhängt, wird in dieser Arbeit untersucht. Es soll eine automatische Bewegungsbestimmung von Objekten mit Hilfe von *stationären* Farbkameras und leistungsfähiger Rechentechnik betrachtet werden, wobei zahlreiche Einzelprobleme auftreten, die beim visuellen System des Menschen zunächst nicht auffallen. Somit konzentriert sich die vorliegende Arbeit auf den Bereich von Beobachtungsszenarien, in denen sich unbekannte Objekte aufhalten und bewegen. Dies kann in Bereichen der Situationsanalyse, in Verhaltensstudien der Interaktionen sich bewegender Objekte sowie der Mensch-Maschine-Interaktionen vorkommen. Dafür werden hier diverse Methoden entwickelt und eingesetzt, die auf der Idee beruhen, dass ein Ähnlichkeitsmaß minimiert wird, um Korrespondenzen in zeitlichen Bildfolgen zu finden. Dies setzt allerdings eine Spezifikation und Optimierung der dafür benötigten Funktionsräume und Ähnlichkeitsmaße für die Beobachtungen und gewünschten Ergebnisse voraus.

### 1.1 Einordnung der eigenen Arbeiten und Zielstellung

Das prinzipielle Anliegen der Verfahren zur Bewegungsanalyse oder Objektverfolgung besteht in der Bestimmung objektspezifischer Bewegungsgrößen aus zeitlichen Bildfolgen. Eine typische Verarbeitungskette hierfür ist in Abbildung 1-1-1a dargestellt. Zu Beginn erfolgt die Extraktion von bestimmten Bildmerkmalen (*Primitiven*), wie bspw. Pixel, Ecken, Konturlinien oder Texturmerkmalen von Objekten. Bei der direkten Verwendung extrahierter Bildregionen kann der Grauwertverlauf sinngemäß als eine Primitive betrachtet werden.



**Abbildung 1-1-1a:** Verarbeitungskette bei der Bewegungsanalyse [Al-Hamadi 01]

Durch Zuordnung (*Matching*) dieser Primitiven aus aufeinander folgenden Bildern der Sequenz entstehen Korrespondenzen, aus denen dann entsprechende *Bewegungsgrößen* ermittelt werden können. Die geschätzten Bewegungsgrößen<sup>1</sup> geben die Richtung, die Position und die Geschwindigkeit der Primitiven an. Die Extraktion der Primitiven sowie insbesondere die Ermittlung der Korrespondenzen werden in der Praxis von vielfältigsten Störungen beeinflusst. Dies hat einfache Gründe:

- Die beobachtete Bewegung eines Objektes in einer Bildszene wird durch verschiedene Einflüsse (Anfangen vom Kamerarauschen, über Schattenwurf und Helligkeitsschwankungen, Verdeckungen bis hin zu Reflexionen z.B. an Fahrzeugen, Deforma-

---

<sup>1</sup> Die Bewegungsgrößen können Verschiebungsvektoren mit Richtung und Geschwindigkeit oder Bewegungstrajektorien eines Objektes beinhalten.

## 1. EINLEITUNG

---

tionen u.ä.) überlagert und dadurch die Korrespondenzermittlung zum Teil stark verfälscht. Diese Einflüsse werden im Rahmen der vorliegenden Arbeit als *nicht-kooperative Messsituationen* bezeichnet [Mysliwetz 86, Schnelting 98, Mecke 99, Al-Hamadi 01a, 02a, 05a, 06a].

- Die automatische Erfassung von einzelnen Objekten, Objektteilen oder Objektprimitiven in realen Bildsequenzen gilt als schwierig, wenn die Objekte interagieren oder mit gestörten Bildsituationen verbunden sind. Dazu stellt die genaue Trennung von lokalen und globalen Bewegungen eines einzigen Objektes eine weitere Herausforderung (z.B. in der Mimik [Niese 07c, Al-Hamadi 08b] oder Gestik [Elmezain 07, 08a, 08b, 09a]) dar. Besonders lassen sich oft Bewegungen verdeckter Objekte bzw. Körperteile nicht oder nur mit erheblichen Einschränkungen statistisch rekonstruieren.

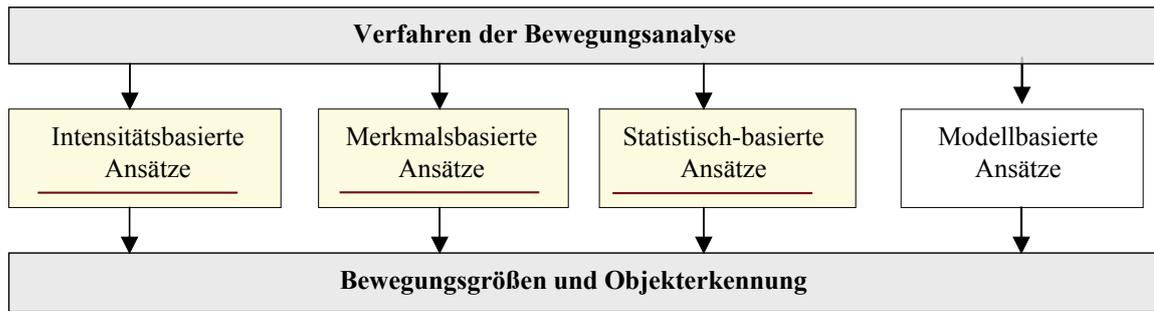
Infolgedessen stehen bereits leistungsfähige Verfahren zur Berechnung von Bewegungsparametern bzw. für das Tracking von Bildregionen zur Verfügung [Mecke 99, Schnelting 98, Daniilidis 92, Deriche 92; Borgefors 88, Rehrmann 94, Horn 81, Nagel 85, Jähne 97, Frank 96, Balke 93, Möller 01, Brandner 02 and Nummiaro 02, Al-Hamadi 02c, Al-Hamadi 0d, Bruhn 02, Brox 04], die meist von einfachen (*idealen*) Bedingungen ausgehen oder nur eine Art der Störungen berücksichtigen. Eine wesentliche Problemstellung stellt derzeit u.a. die Ermittlung und Untersuchung geeigneter Verfahren zur Bewegungsanalyse von relativ dichtem Straßenverkehr in realen Bildsequenzen dar [Coifman 98, Al-Hamadi 04b, 06a, Pathan 09a, 09b].

Die klassischen Verfahren zur Bestimmung von Bewegungsgrößen (z.B. Verschiebungsvektoren, Optischer Fluss oder Bewegungstrajektorien) von sich bewegenden Objekten in Bildsequenzen lassen sich wie folgt einordnen (Abb. 1-1-1b):

- intensitätsbasierte Ansätze, die sich wiederum in zwei Gruppen einteilen lassen, die als Verfahren nach dem ‘optischen Fluss’ (auch differentielle Verfahren genannt) [Horn 81, Nagel 85, Jähne 97, Weickert 00, Bruhn 02, Brox 04] und Matchingverfahren [Burt 82, Klette 96, Musmann 85, Mecke 99, Al-Hamadi 01a] bekannt sind,
- merkmalsbasierte Verfahren [Bandopadhyay 86, Kories 89, Jähne 89, Borgefors 88, Rehrmann 98a, Coifman 98, Al-Hamadi 03e, Al-Hamadi 04b, Al-Hamadi 05a],
- statistisch-basierte Verfahren [Isard 96, Isard 98, Schuster 95, Al-Hamadi 07a, Al-Hamadi 08a], sowie
- modellbasierte Verfahren [Musmann 85, Blake 98, Mecke 94, Calow 03, Al-Hamadi 06f, Al-Hamadi 06e].

Die Übergänge zwischen den Methoden sind fließend. Manche der genannten Arbeiten sind deswegen nicht unbedingt eindeutig einer bestimmten Gruppe zuzuordnen. Dennoch lassen sich diese Kategorien verallgemeinern und auf Besonderheiten analysieren. In diesem Zusammenhang spielen beispielsweise die intensitätsbasierten Ansätze eine große Rolle, die eine vollständige und genaue Bestimmung des optischen Flussfelds in idealen Situationen zwischen aufeinander folgenden Bildern einer Bildfolge ermöglichen. Daher bilden diese Ansätze eine gute Grundlage für verschiedene Verfahren, die in dieser Arbeit untersucht werden.

## 1. EINLEITUNG



**Abbildung 1-1-1b:** Die Einteilung der Verfahren der Bewegungsbestimmung (Unterstrichene Ansätze werden in der vorliegende Arbeit behandelt).

Während die Ansätze der ersten Gruppe in der Regel örtliche bzw. zeitliche Intensitätsänderungen in den Bildfolgen auswerten, bieten die *merkmalsbasierten* Ansätze eine weitere Möglichkeit zur Bewegungsbestimmung von Objekten in Bildfolgen (Abb. 1-1-1b). Die Idee der merkmalsbasierten Ansätze liegt nicht in der Beschreibung des Objekts als Ganzes, sondern durch mehrere Teilmerkmale. Im Vergleich zu den intensitätsbasierten Ansätzen, bei denen Pixel für Pixel untersucht werden, erfolgt hier die Extraktion von Bildelementen, die zu Merkmalen auf höherer Ebene zusammengefasst werden. Diese sind dann der Gegenstand des zeitlichen Matchings.

Die Wahl der Merkmale hängt von den Objekteigenschaften und der Anwendung ab [Kanade 80, Kories 89, Jähne 89, Borgefors 88, Guse 92, Rehrmann 96, Al-Hamadi 05e, Al-Hamadi 06d, Al-Hamadi 06e]. Meist ist eine vollständige Suche nach Bildprimitiven im Bild uneffizient, da sich beispielsweise bei einer durchschnittlichen Anzahl von nur  $N$ -Bildprimitiven (*Ecken und Kanten*) in aufeinander folgenden Bildern schon ( $2^{N+N}$ ) verschiedene Möglichkeiten der Zuordnung ergeben. Davon ausgehend, arbeiten die merkmalsbasierten Verfahren so allgemein, dass sie generelle Informationen liefern, die von anderen, spezielleren Ansätzen nicht abgedeckt werden können. Die Informationen müssen dafür in zusätzlichen Prozessen weiterverarbeitet werden. Hierbei liegt die Herausforderung in der korrekten Zuweisung der Merkmale zu den Objekten.

Die dritte Gruppe beinhaltet die statistisch-basierten Ansätze, die auf der Bayes-Schätzung [Daniilidis 92, Fox 03, Isard 98, Stüker 04, Martnez 04] basieren. Die rekursive Bayes-Schätzung ist ein Filter, der für die Schätzung von Parameter in dynamischen Systemen eingesetzt wird. Die Bayes-Schätzung beruht auf der Theorie, dass vorhandenes Wissen für die Bewertung von beobachteten Signal- oder geschätzten Parameterwerten genutzt werden kann. Beispielsweise werden mit Kalman-Filtern [Kalman 60, Daniilidis 92] lineare Problemstellungen, wie die Positionsschätzung von Objekten, gelöst [Köhler 05, Mecke 98, Pathan 09a, Pathan 09b, Al-Hamadi 03g].

Anwendung findet dieses effiziente Verfahren bei der autonomen Fahrzeugsteuerung [Mysliwetz 95, Dickmanns 92, Coifman 98, Pathan 09a, Pathan 09b]. In vielen realen Situationen ist es jedoch nicht der Fall, dass sich die Bewegungen und Messungen durch lineare Modelle abbilden lassen. Der Einsatz von Partikelfiltern zur Multi-Objektverfolgung wurde in verschiedenen Arbeiten untersucht [Sidenbladh 03, Martnez 04, Al-Hamadi 07a, Al-Hamadi 08a]. Dazu wurden Ansätze entwickelt, in denen die Modellierung der Szene durch bloße Partikelfilteranwendung erfolgt, die Identifizierung der einzelnen Objekte und Ermittlung deren Trajektorien jedoch ausbleibt [Sidenbladh 03, Vermaak 03, Homberg 07].

In anderen Ansätzen wird durch die Verwendung der Datenassoziation<sup>2</sup> die Unterscheidung der Objekte auf simulierten Daten oder auf Radardaten ohne komplizierte Situationen untersucht [Kreucher 04, Karlsson 01].

Im Gegensatz zu den kurz vorgestellten Verfahren detektieren und verfolgen die modellbasierten Ansätze (vierte Gruppe) die Zielobjekte mit einem anhand von Vorwissen konstruierten Modell. Es wird ein Matching zwischen Vorwissen und Bilddaten durchgeführt. Vor allem sind die Modelle von der Art der Zielobjekte abhängig, die starr oder nicht starr sein können. Dazu gehören u.a. aktive Konturen oder Formmodelle (z.B. modifiziertes Activ-Shape-Modell [Al-Hamadi 05b, Al-Hamadi 06e]). Während aktive Konturen und Formmodelle durch ihre Definitionen auf den Bilddaten arbeiten und somit ihre Zielstrukturen auch bei vorhandenen Störungen, wie Rauschen, finden können, zeigen die Merkmalsansätze durch ihre Abhängigkeit von den Bilddaten eine größere Störanfälligkeit. Des Weiteren profitiert die Genauigkeit bei der Ermittlung der Bewegungsparameter von eingebrachtem Modellwissen. Hingegen beschränkt das eingebrachte Vorwissen der Modelle den Verfolgungsprozess auf Objekte der modellierten Art. Je detaillierter die Modelle formuliert werden, desto spezieller und enger ist der Einsatzbereich. Darin liegt die große Schwäche der modellbasierten Ansätze. Sind sie einmal initialisiert, sind sie durchaus in der Lage, ihr Zielobjekt zu verfolgen – auch bei partiellen Verdeckungen. Des Weiteren erfordert die automatische Analyse von Szenen mit Objektfluktuationen<sup>3</sup> eine automatische Initialisierung der aktiven Konturen. Bezüglich unkooperativer Situationen liefern die modell- und konturbasierten Ansätze – eine erfolgreiche Initialisierung vorausgesetzt – ein weitaus stabileres Ergebnis. Problematisch kann auch die Differenzierung mehrerer Objekte werden, wenn sich beispielsweise der Schattenwurf eines Objekts mit der Region eines anderen Objekts verbindet oder wenn sich Objekte teilweise verdecken und dadurch als eine Region erfasst werden. Eine Trennung ist dann ohne weitere Analyse nicht möglich.

Über die Einsatzmöglichkeiten entscheidet letztendlich das Ausgangswissen über die zu analysierenden Szenen. Für Szenarien, in denen ein bestimmtes Objekt oder bestimmte Arten von Objekten untersucht werden sollen, eignen sich die kontur- und modellbasierten Ansätze. Wichtig ist dabei ein geeignetes Initialisierungsverfahren oder eine konstante Anzahl von Objekten [Al-Hamadi 06f].

Die regionen- und merkmalsbasierten Ansätze sind für Szenen mit variierenden Objekten geeignet, da sie vergleichsweise wenig Vorwissen benötigen und die Szenen allgemein verarbeiten [Bandopadhyay 86, Kories 84, Dubuisson 93, Rehrmann 96, Rehrmann 98a, Rehrmann 98b, Coifman 98, Al-Hamadi 03e, Al-Hamadi 04b, Al-Hamadi 05a]. Werden Szenarien mit Objektfluktuation untersucht, liefert ein intensitätsbasierter Ansatz durch seine Bewegungsanalyse eine gute Grundlage für die automatische Initialisierung und Verarbeitung der Objekte [Al-Hamadi 99a, 01a]. Die einzelnen Ansätze und Beispielarbeiten verdeutlichen, dass bei der Erkennung eine Kombination der Methoden nicht ausgeschlossen werden kann. Im Gegenteil, einerseits bauen sie aufeinander auf, andererseits können durch die Kombination die Defizite der einen Methode durch die Vorteile einer anderen abgedeckt werden, wie beispielsweise die

---

<sup>2</sup> Wahrscheinlichkeitsbasierte Datenerfassung und –zuweisung auf der Grundlage von Hypothesengenerierung.

<sup>3</sup> Der Begriff *Fluktuation* bezeichnet eine wiederkehrende Veränderung von Gegebenheiten und Zuständen. Je nach Bedeutungszusammenhang und Fachgebiet kann Fluktuation unterschiedliche Sachverhalte wiedergeben und diese anderweitig darstellen [WIKIPEDIA].

## 1. EINLEITUNG

---

Kombination von Active-Shape-Modells mit intensitätsbasiertem Matchingverfahren bei [Al-Hamadi 06f].

Neben diesen prinzipiellen Methoden gibt es eine Vielzahl kombinierter und hierarchischer Ansätze, die jeweils die günstigsten Eigenschaften einzelner Ansätze ausnutzen. Hier kann, ausgehend von den aus Korrespondenzen ermittelten Bewegungsgrößen, die Bewegungsbestimmung entweder direkt erfolgen oder es werden Schätzverfahren wie z.B. Kalman-Filter [Al-Hamadi 03g, 03f, Pathan 09a, 09b] oder Partikelfilter [Al-Hamadi 07a, 08a] eingesetzt, die eine Filterung der ermittelten Bewegungsgrößen realisieren. Um die Leistungsfähigkeit eines Verfahrens zur Bewegungsanalyse bewerten zu können, soll das Verfahren die Anforderung hinsichtlich der Robustheit und Flexibilität erfüllen, die wie folgt zusammenzufassen sind:

- 1). *Objekt- und Umgebungsunabhängigkeit*: Die Vielzahl von Objektarten bzw. Objektvariationen in Szenen mit Fußgängern und Fahrzeugen erfordert eine allgemeine Objektbeschreibung, die für jedes Objekt auf gleiche Weise und automatisch erstellt und verarbeitet werden kann. Um eine hohe Einsatzflexibilität zu erreichen, verbietet es sich, Wissen über die Umgebung, z.B. in Form eines Umgebungsmodells, einzubringen. Auf diese Weise wäre zwar eine Einschränkung im Bewegungsraum der Objekte, wie Straßen- oder Fußwegbereiche möglich. Sie würde jedoch die Analyse verschiedener Szenen erschweren.
- 2). *Automatische Initialisierung*: Für ein automatisches Verfolgungssystem muss jeder Zeitschritt automatisch ablaufen und wiederum den nächsten einleiten. Das bedeutet für Szenen mit Objektfluktuationen, dass neue Objekte, die die Szene betreten, automatisch, ohne interaktive Eingriffe erkannt und vom System verarbeitet werden müssen. Im Zusammenhang mit der Fluktuationserkennung ist demnach eine selbstständig verlaufende Merkmalsextraktion zur Objektbeschreibung erforderlich, die der Initialisierung der neuen Objekte dient. Dazu ist durch die unbekannte und veränderliche Anzahl der Objekte eine Organisation der hinzukommenden und verschwindenden Objekte nötig. So müssen hier alle neuen, aber auch verschwundenen Objekte als solche identifiziert und entsprechend in das System eingeführt oder aus dem System gelöscht werden.
- 3). *Erkennung von Konfliktsituationen*: In realen Szenen ist nicht nur die Anzahl der Objekte unbekannt und veränderlich. Hinzu kommen Szenen, in denen Objekte Konfliktsituationen auslösen. Das ist beispielsweise der Fall, wenn Objekte sich in ihrem Weg kreuzen oder sich überholen. Ebenso können Objekte nicht nur von anderen Objekten, sondern auch von Hindernissen verdeckt werden. Hier ist in den Augenblicken der Verdeckung keine eindeutige Zuordnung möglich und erfordert eine spezielle Handhabung dieser Situationen. Ebenso liegt keine Datengrundlage für die Bewegungsbeschreibung und -schätzung vor. Auslöser solcher Situationen können verschiedene Szenarien sein, beispielsweise Kreuzen, Überholen, Stehen bleiben oder Störungen.

Die Verwendung von herkömmlichen Methoden zur Bestimmung von Bewegungsparametern (z.B. Verschiebungsvektoren) in Bildbereichen, in denen bisher keine eindeutige Messung möglich ist, stellt die Effizienz der erwähnten Methoden und die Verwertbarkeit der Ergebnisse in Frage. Erzwingt man nämlich in einem solchen Fall die Bestimmung der Bewegungsparameter eines Objektes mittels solcher Methode, so wird in der Regel das Ergebnis fehlerhaft sein. Die Auswertung solcher fehlerhaften

Messergebnisse führt entweder zu einer Verfälschung nachfolgender Klassifikationsergebnisse oder bedingt weiteren Rechenaufwand, wenn mit zusätzlichen Kriterien (z.B. Kalmann-Filter, Datenassoziationen) versucht wird, solche Fehlmessungen nachträglich zu eliminieren.

Um die vorgestellten Anforderungen erfüllen zu können, werden in dieser *vorliegenden Arbeit* mehrere Beiträge auf dem Gebiet der Analyse von bewegten Bildern entwickelt, die sich wie folgt charakterisieren lassen:

- Zum einen wird ein systematischer Rahmen zur Entwicklung genauer intensitätsbasierter Ansätze auf der Grundlage der optischen Flussverfahren und Korrelationsverfahren vorgestellt. Dieser erlaubt eine Verbesserung der zurzeit bekannten Schätzverfahren, die in der Literatur zu finden sind. Gegenstand dieser vorliegenden Forschungsarbeiten sind die intensitätsbasierten optischen Flussverfahren und ihre Verwendung zur Selektion und Verfolgung mehrerer bewegter Objekte. Da diese Ansätze einen hinreichenden linearen Verlauf von Änderungen der Bildintensität von Bild  $I_k$  zu  $I_{k+1}$  voraussetzen, wird im weiteren Schritt eine adaptive Systemstruktur auf der Grundlage der Korrelationen von Bildregionen bzw. –ausschnitten entwickelt. Diese nutzt die Farbinformation zur Verbesserung der Störsicherheit aus. Die Adaptivität dieser Systemstruktur bezieht sich auf die optimale Farbraumtransformation und das kanalspezifische Gesamtkriterium, die zur Verbesserung der Matchingqualität führt. Dabei wird die Bewegungsbestimmung von Objekten nicht wie bisher überwiegend in der Schwarz-Weißdomäne sondern in der Farbdomäne durchgeführt.
- Zum zweiten wird ein neues Matching-Paradigma zur merkmalsbasierten Korrespondenzermittlung in Farbbildsequenzen vorgeschlagen, das sich vom klassischen Matching durch seine hierarchische Struktur und Adaptivität unterscheidet. Dieses Konzept segmentiert, verfolgt und erkennt nicht nur fest gewählte Regionen, sondern auch beliebige *veränderliche* Objekte in Bildfolgen automatisch. Zusätzlich ermöglicht es die Berechnung in relativer Echtzeitnähe und macht dieses äußerst robuste merkmalsbasierte Matching praktisch einsetzbar.
- Zum dritten wird ein Partikelfiltersystem in Verbindung mit einer Datenassoziation für die Multi-Objektverfolgung entwickelt, dessen Verarbeitungsstufen die ursprünglichen Schritte eines Partikelfilters beinhalten. Diese Schritte werden in dieser Arbeit aber so modifiziert, dass eine zuverlässige Bestimmung von Bewegungstrajektorien in Konfliktsituationen möglich ist. Diese Modifikationen beziehen sich zuerst auf die Prädiktion, die durch ein erweitertes *adaptives Bewegungsmodell* erreicht wird. Zusätzlich wird der ursprüngliche Ablauf bei der Anwendung eines Partikelfilters um einen 3-stufigen Assoziationsschritt mit Datenausschluss, Datenzuweisung und Datenverwaltung ergänzt, um die Problematik der Fluktuationserkennung und den Umgang mit Objektinteraktionen lösen zu können.
- Zuletzt wird ein neuer Lösungsweg auf der Grundlage der Datenfusion entwickelt, der eine Zustandsschätzung und robuste Datenzuweisung von Objekten in Konfliktsituationen ermöglicht. Die Struktur für die Datenfusion nutzt die Stärken der intensitäts- und merkmalsbasierten Verfahren so aus, dass die Schwächen in beiden Verfahren weitgehend kompensiert werden können. Die Grundlagen dafür sind die bekannte Bayes-Schätzung, die Fuzzy-Logik und die vorgestellten Bewegungsanalyseverfahren.

### 1.2 Gliederung der Arbeit

In der vorliegenden Arbeit werden Verfahren entwickelt, die die genaue Erfassung und Verfolgung von sich bewegenden Objekten unter dem Einfluss von gestörten Messsituationen ermöglichen. Die Möglichkeiten der Analyse hinsichtlich der Genauigkeit und der analysierbaren Bewegungsform hängen unmittelbar von der Korrespondenzermittlung ab, die durch mehrere Methoden realisiert wurden. Für die Analyse wurde von einem statischen Aufnahmesystem, das Farbbildsequenzen liefert, ausgegangen. Davon ausgehend wird die vorliegende Arbeit in sechs Kapitel gegliedert, die die Fortschritte bei der Analyse von bewegten Farbbildern unter dem Einfluss von nicht-kooperativen Messsituationen erörtern<sup>4</sup>.

Im Anschluss an die Einleitung im *ersten Kapitel* werden in dem *zweiten Kapitel* die bekannten intensitätsbasierten Ansätze hinsichtlich ihrer Leistungsfähigkeit und der Robustheit untersucht und bewertet. Experimente zur Evaluation der Qualität belegen die Eignung der Verfahren zur Bestimmung der Verschiebungsvektoren eines Objektes. Die notwendigen Optimalitätsbedingungen werden für zwei einfache Variationsmethoden präsentiert, das so genannten Horn-Schunck-Verfahren (HS), und das Lucas-Kanade-Verfahren (LK). Sie bilden eine mathematische Grundlage und Rechtfertigung für den zweiten Abschnitt der Arbeit, indem die praktische und numerische Behandlung im Vordergrund steht. Um meine Ergebnisse in einer Reihe von Experimenten geeignet analysieren und auswerten zu können, werden Darstellungsformen und Gütekriterien für den berechneten optischen Fluss eingeführt. Zur praktischen Behandlung und Auswertung werden in diesem Kapitel grundsätzlich zwei unterschiedliche Ansätze verfolgt:

- Zum einen wird die weit verbreitete Lösungsstrategie „*Optimierung und anschließende Diskretisierung*“ verwendet. Dabei werden die Lösungen zunächst ausgehend von Euler-Lagrange-Gleichungen optimiert und dann in diskretisierter Form iterativ gelöst<sup>5</sup>.
- Zum anderen wird ein adaptives Verfahren nach dem (direkten) Prinzip der „*Diskretisierung und anschließenden Verfeinerung der resultierenden Ergebnisse*“ behandelt. Die Idee beruht darauf, alle vorkommenden lokalen Informationen und Differentialgleichungen erst mit dem Lucas-Kanade-Verfahren zu lösen und dann das resultierende nichtlineare Optimierungsproblem, das eine große Anzahl an Optimierungsvariablen enthält, hierarchisch zu lösen. Eine Verfeinerung erfolgt durch zwei Schritte, die u.a. eine adaptive Anpassung des Abtastfensters und eine Anti-Ghost-Adaption beinhalten.

Da die qualitative Behandlung solcher optischen Flussverfahren mit realen Bildsequenzen schwierig ist, greifen wir auf selbst generierte Bilder mit bekannten Grundwahrheiten zurück, die sich auf reale Situationen beziehen. Modifizierte Variationsmethoden sowie die oben angesprochene Variationsmethode zur simultanen optischen Flussberechnung werden zum Vergleich mit vorgeschlagenen Fehlerkriterien behandelt. Dazu wird in einer Reihe von Experimenten der Einfluss von Parameterrestriktionen auf die Ergebnisse bei der Berechnung mit dem optischen Fluss untersucht. Dies beinhaltet eine simultane Bewegungserkennung über Parameterrestriktionen sowie Auswirkungen einer Regularisierung auf den optischen Fluss und die Bewegungssegmentierung.

---

<sup>4</sup> Durch den Aufbau der Arbeit ist es durchaus möglich nur einzelne Kapitel zu lesen. Leser, die mit dem Thema des optischen Flusses noch nicht vertraut sind, sollten sich allerdings zum Einstieg Kapitel 2 widmen.

<sup>5</sup> Diese Verfahren wurden in C++ implementiert und mittels realer und synthetischer Bildsequenzen erprobt.

Im zweiten Teil des zweiten Kapitels wird eine adaptive Systemstruktur auf der Grundlage von Korrelationen von erfassten Bildausschnitten bzw. –regionen entwickelt, die die Farbinformation optimal zur Verbesserung der Matchingqualität ausnutzt. Dabei legt konkret die funktionelle Beschreibung dieser Systemstruktur zur Bewegungsanalyse zunächst weder ein Bewegungsmodell noch den verwendeten Farbraum fest. Es wird die Bewegungsbestimmung von Objekten nicht wie bisher überwiegend in der Schwarz-Weißdomäne sondern in der Farbdomäne durchgeführt. Diese beiden Teile lassen sich den intensitätsbasierten Ansätzen (erste Gruppe in Abbildung 1-1-1b) zuordnen.

In *Kapitel 3* wird ein neues Matching-Paradigma zur Verfolgung multipler Objekte auf der Grundlage der hierarchischen Merkmalsextraktion vorgeschlagen. Die Verarbeitungsschritte, die zur Umsetzung dieses Matching-Paradigmas führen, werden begründet und detailliert beschrieben. Der vorgeschlagene Segmentierungsansatz, der vorgeschlagene Schattenfilter und der Nachverarbeitungsalgorithmus werden dargestellt und die auf der Grundlage der Bewegungssegmentierung generierten hierarchischen Merkmalskomponenten vorgestellt. Im Anschluss wird das spezielle merkmalsorientierte Prinzip der Zuordnung von Bildregionen spezifiziert. Erweiterungen und Optimierungsansätze zur Verbesserung der Matchinggüte werden entwickelt und getestet. Die durch das vorgeschlagene Multi-Matching ermittelten Korrespondenzen führen zu Bewegungstrajektorien, welche Auskunft über die Bewegung in längeren Bildsequenzen geben können und somit das wichtigste Ergebnis des Zuordnungsprozesses darstellen. Dazu werden Ergebnisse vorgestellt, die in umfangreichen experimentellen Untersuchungen anhand realer Bildsequenzen gewonnen wurden.

Zum Nachweis der Funktions- und Leistungsfähigkeit des Matching-Paradigmas wurden Analysen sowohl in einfachen, künstlichen als auch in realen, gestörten Bildfolgen durchgeführt. Die erzielten Ergebnisse werden erläutert und mit denen herkömmlicher Verfahren verglichen und bewertet. Es werden die erreichten Ziele beschrieben, sowie Stärken und Schwächen des entwickelten Verfahrens aufgezeigt. Zum Ende dieses Kapitels werden die Ergebnisse unserer Untersuchungen, Erweiterungen und die Grenzen des vorgeschlagenen Paradigmas diskutiert und zusammengefasst.

In *Kapitel 4* wird ein statistisch-basiertes Verfahren entwickelt, das auf der Basis von Partikelfiltern und Datenassoziationen die Verfolgung multipler Objekte sowie die stabile Ermittlung ihrer Trajektorien auch in Konfliktszenen ermöglicht. Für die Darstellung wurde von einem statischen Aufnahmesystem ausgegangen. Die zu analysierenden Szenen enthalten Objekte unterschiedlicher Art und unterschiedlichen Verhaltens, so genannte Multi-Objekte, die sich größtenteils unabhängig von einander bewegen. Die Multi-Objektverfolgung wurde mit einem Partikelfiltersystem realisiert und um weitere Techniken ergänzt und modifiziert. Dazu wird basierend auf Erkenntnissen aus Kapitel 1 und 2 die Ausgangssituation analysiert und daraus die Anforderungen an ein robustes System zur Objektverfolgung definiert. Es wird ein System entworfen, das die notwendigen Schritte von automatischer Erfassung über Merkmalsextrahierung bis zur Multi-Objektverfolgung realisiert. Diese Schritte werden im weiteren Verlauf des vierten Kapitels einzeln betrachtet und diskutiert. Dies beinhaltet die Bewegungssegmentierung, die Merkmalsextraktion, das Aufstellen des Partikelfilters und die multi-objekt-fähige Systematisierung des Partikelfilters mit einer Datenassoziation.

## 1. EINLEITUNG

---

Die Feinstruktur der Methode wird mit Untersuchungen innerhalb des Kapitels untermauert. Dazu werden in diesem Kapitel Untersuchungen mittels realer Bildsequenzen beschrieben. Zum einen wird damit die Verwendung bestimmter Techniken der in Kapitel 3 entworfenen Schritte begründet. Zum anderen wird die Funktionstüchtigkeit der entwickelten Methode zur Multi-Objektverfolgung untersucht. Mit speziellen Bildsequenzen werden Stärken und Schwächen der vorgeschlagenen Methode aufgezeigt. Anhand der Ergebnisse werden die erreichten Ziele aber auch die Grenzen der vorgeschlagenen Methode erläutert. Abschließend erfolgt eine Zusammenfassung der Erkenntnisse dieser Arbeit.

Im *Kapitel 5* wird basierend auf den Erkenntnissen aus Kapitel 2, 3 und 4 ein neuer Lösungsweg vorgeschlagen und als hybrides System dargestellt, das durch Fusion der intensitätsbasierten und merkmalsbasierten Ansätze mittels Partikelfilter (Kapitel 4), die Vorteile beider Verfahren nutzt und dadurch ihre Nachteile kompensiert. Dabei wird auf statistische Verfahren und Fuzzy-Logik zurückgegriffen sowie das Partikelfiltersystem vorgestellt. Anschließend werden Untersuchungen von realen Bildsequenzen beschrieben, die Konfliktsituationen beinhalten. Zum einen wird damit die Verwendung bestimmter Techniken in den in Kapitel 5 entworfenen Schritten begründet. Zum anderen wird die Funktionstüchtigkeit der entwickelten Methode zur Datenfusion untersucht. Mit Bildsequenzen werden Stärken und Schwächen der Methode aufgezeigt. Anhand der Ergebnisse werden die erreichten Ziele, aber auch die Grenzen der Methode erläutert. Den Abschluss dieser Arbeit bildet das sechste Kapitel, welches zugleich eine Zusammenfassung der Erkenntnisse dieser vorgelegten Arbeit und den Ausblick darstellt.

### 2. Intensitätsbasierte Ansätze

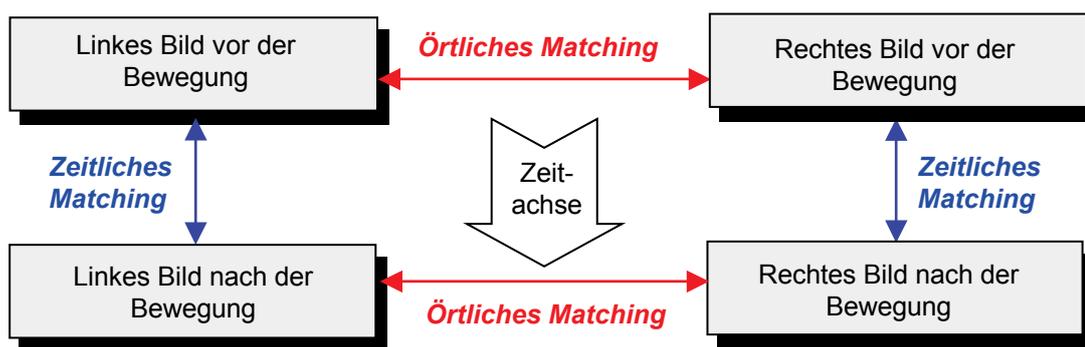
Zahlreiche Fragestellungen des Maschinensehens erfordern die Ermittlung von Bewegungsinformation aus zeitlichen Bildfolgen [Dickmanns 92, Ohm 95, Huwer 98, Al-Hamadi 08b, Elmezain 09a, Niese 07b]. Jedoch ist es ohne Vorwissen über statische Objekte kaum möglich, die absolute Bewegung von Objekten zu ermitteln. Die Verwendung einer einzelnen Kamera erlaubt die Berechnung der 3D Bewegungsprojektion der realen Welt nur in der 2D Bildebene. Daher spricht man in Analogie zur Strömungslehre bei der projizierten Bewegung auch vom optischen Fluss. Unmittelbare theoretische Zusammenhänge und Gemeinsamkeiten zwischen den genannten optischen Flussmethoden werden hier diskutiert, da diese eine wichtige Grundlage für die vorliegende Arbeit darstellen.

Zur besseren Einordnung wird in diesem Kapitel ein Überblick über die wesentlichsten intensitätsbasierten Ansätze zur Bewegungsanalyse in Bildsequenzen sowie übergreifende Methoden gegeben. Hierbei handelt es sich im ersten Teil um die Ausführungen von Horn-Schunck aus ihrem grundlegenden Artikel aus dem Jahr 1981 [Horn 81] und das Lucas-Kanade-Verfahren [Lucas 81]. Das von uns implementierte Verfahren arbeitet mit Graustufenbildern. Es gibt zwar inzwischen verschiedene Ansätze, um den optischen Fluss auch aus Farbbildern zu berechnen. Die größere Anzahl an Informationen, die Farbbilder durch ihre drei Farbkanäle liefern, ist jedoch selten von entscheidender Bedeutung bei der Berechnung des optischen Flusses und kann nur in wenigen Ausnahmefällen zu einer besseren Approximation der zugrunde liegenden Bewegung führen.

Im Unterschied zur Analyse von Grauwertbildern mittels des bekannten optischen Flussverfahrens wird im zweiten Teil die Farbinformation berücksichtigt. Die Betrachtung des Problems mittels Korrelationsverfahrens ist aus [Al-Hamadi 01a, 02a, 03f, 03g] entnommen. Daher wird eine adaptive Systemstruktur zur Ermittlung von Bewegungsgrößen für beliebige Blockgrößen und objektangepasste Größen entwickelt. Verglichen mit anderen Ansätzen, die die Bewegungsanalyse auf bestimmte Merkmale in Intensitätsbildern oder flexible Konturmodelle [Al-Hamadi 06f] beschränken, wird in dem vorgestellten System unmittelbar die optimale Erfassung der Farbintensitäten des Bildkanals verwendet. Da sich die Korrespondenzermittlung mittels eines Ähnlichkeitsmaßes auf flächenhafte Ausschnitte bezieht, ist in der vorgeschlagenen adaptiven Systemstruktur in [Al-Hamadi 01a, 02a, 03f] keine explizite Merkmalsextraktion erforderlich. Es werden deshalb keine konkreten Annahmen über bestimmte Objektmerkmale oder Verknüpfungsfunktionen bei der Korrespondenzermittlung getroffen. Konkret legt die funktionelle Beschreibung des Systems zur Bewegungsanalyse zunächst weder ein Bewegungsmodell noch den verwendeten Farbraum fest.

Eine aktuelle Tendenz auf dem Gebiet der Bewegungsanalyse in Forschung und Anwendung ist die Erweiterung der Verfahren auf den 3D-Raum [Stiller 97, Mecke 01, Knöppel 01, Tornow 03, Elmezian 09a]. Damit ergeben sich neue Analysemöglichkeiten wie bspw. die Erfassung von dreidimensionalen Bewegungsgrößen oder der Änderung der räumlichen Messobjektform. Eine mögliche Ausgangsbasis hierfür sind mehrkanalige Video-Bildfolgen, die von mehreren Kameras aufgenommen werden. Ausgehend von der Tatsache, dass Objektveränderungen (*Freiheitsgrade*) im 3D-Raum sowohl hinsichtlich ihrer Lage (*Ortskoordinate*) als auch bezüglich ihrer Dynamik

(Zeitkoordinate) zu berücksichtigen sind, kann man zur Problembeschreibung von einem vierdimensionalen Raum ausgehen, der durch die drei Ortskoordinaten und eine Zeitachse gebildet wird. Klassische Verfahren der Bildauswertung bilden meist die Ausgangsbasis und werden miteinander kombiniert. In diesem Kontext kann beispielhaft die Kombination von Verfahren der *monokularen Bewegungsanalyse* und der *photogrammetrischen 3D-Vermessung* gesehen werden, die u.a. von [Liu 93] vorgeschlagen wurde. In Abbildung 2-1-0 ist dieser Zusammenhang schematisch dargestellt. Bei der photogrammetrischen Vermessung werden durch die Auswertung von Stereobildpaaren die 3D-Koordinaten einer Objektoberfläche für jeweils einen Zeitpunkt berechnet. Die Basis bildet hierbei das Finden von Korrespondenzen in den Bildpaaren (*örtliches Matching*).



**Abbildung 2-1-0:** Zusammenhang zwischen monokularer Bewegungsanalyse und photogrammetrischer 3D-Vermessung.

Durch eine monokulare Bewegungsanalyse können je Kanal (Bildsequenz einer Kamera) Korrespondenzen entlang der Zeitachse ermittelt werden (*zeitliches Matching*). Dadurch ist es prinzipiell möglich, eine zeitliche Verknüpfung der 3D-Koordinaten zu realisieren. Auf diese Weise können Bewegungstrajektorien im Raum ermittelt werden. Für Beispiele, an denen das Korrespondenzproblem auch im dreidimensionalen Raum betrachtet wird, sei auf [Liu 93, Saneyoshi 96, Tornow 03, Mecke 01] verwiesen.

### 2.1 Verfahren nach dem Optischen Fluss

Um das Prinzip des optischen Flusses zu verdeutlichen, geht man im Allgemeinen davon aus, dass die Änderungen der Bildintensität von Bild  $I_k$  zu Bild  $I_{k+1}$  nur durch relative oder absolute Objektbewegungen verursacht werden. Allerdings darf nicht verallgemeinert werden, dass jede Intensitätswertänderung im Verlauf einer Bildsequenz nur durch Bewegung verursacht wurde. Es können z.B. Beleuchtungsänderungen oder Reflexionen von sich bewegenden Objekten auftreten. Dabei besteht die Schwierigkeit, Grauwertänderungen aufgrund direkter Objektbewegung von indirekt durch sie hervorgerufene Änderungen zu unterscheiden. Die ortsabhängige Intensitätswertänderung lässt sich durch den Verschiebungsvektor darstellen, der für jeden Pixel des Ursprungsbildes Richtung und Geschwindigkeit der Verschiebung beschreibt.

Für die Bestimmung der Verschiebungsvektorfelder wird das differentielle Verfahren angewendet [Nagel 85, Singh 91, Klette 96, Black 93, Jähne 89, Horn 81], das den Übergang von Bild  $I_k$  zu  $I_{k+1}$ , also den Verlauf von Änderungen der Bildintensität von Bild  $I_k$  zu  $I_{k+1}$  auswertet. Diese Änderungen werden durch ein Vektorfeld dargestellt, das die Verschiebung aller Pixelgrauwerte von Bild  $I_k$  zu Bild  $I_{k+1}$  beinhaltet. Dieses

## 2. INTENSITÄTSBASIERTE ANSÄTZE

Vektorfeld wird auch als *optisches Flussfeld* und der Verlauf der Änderung der Bildintensitäten als *optischer Fluss* bezeichnet. Deshalb wird das Verfahren als eine Approximation des lokalen Verschiebungsvektorfeldes betrachtet.

Um den Verlauf der Grauwerte in einer Bildfolge vom Zeitpunkt  $k$  zum Zeitpunkt  $k+1$  repräsentieren zu können, wird an den optischen Fluss die Bedingung der Bildwerttreue  $I(x+u, y+v, t_k + \delta t_{const}) = I(x, y, t_k)$  gestellt. Dies bedeutet, dass ein Pixel aus dem Bild zum Zeitpunkt  $t=k$  im Bild zum Zeitpunkt  $t=k+1$  zwar seine Position, nicht aber den Wert der Bildirradianz ändert. Somit lässt sich von der Bildirradianzänderung auf eine Pixelverschiebung schließen. Der Grauwert des Pixels  $I(x, y, t)$  zum Zeitpunkt  $k$  ist zum Zeitpunkt  $k+1$  um den Vektor  $\mathbf{v}$  verschoben. Ziel ist es nun, aus den Grauwerten einer Bildfolge das dazugehörige optische Flussfeld zu errechnen. Bei geringen Objektbewegungen und konstanten Beleuchtungsverhältnissen besitzt die Grauwertfunktion  $I_k(x, y, t)$  eine hohe Linearität und kann durch eine Taylorreihe erster Ordnung angenähert werden und somit lässt sich die Bildfunktion durch die Taylorreihenentwicklung lokal für kleine Schritte ( $dx, dy, dt$ ) wie folgt darstellen:

$$I(x + dx, y + dy, t + dt) = I(x, y, t) + dx \cdot \frac{\partial I(x, y, t)}{\partial x} + dy \cdot \frac{\partial I(x, y, t)}{\partial y} + dt \cdot \frac{\partial I(x, y, t)}{\partial t} + e \quad (2-1-1a)$$

Bei der Konstanz des Grauwertes in aufeinander folgenden Bildern einer Bildfolge können die nichtlinearen Terme zweiter und höherer Ordnung<sup>6</sup>  $e$  vernachlässigt werden. Aus der Gleichung (2-1-1a/b) und nach Division durch  $dt$  ergibt sich die aus der Thermodynamik bekannte Kontinuitätsgleichung, die man zur Ermittlung der optischen Flussgeschwindigkeiten  $u$  und  $v$  verwendet, wobei  $u = dx/dt$  und  $v = dy/dt$  sind.

$$\nabla(x, y) \in \Omega: \quad u \cdot \frac{\partial I(x, y, t)}{\partial x} + v \cdot \frac{\partial I(x, y, t)}{\partial y} + I_t \cdot \frac{\partial I(x, y, t)}{\partial t} = u \cdot I_x + v \cdot I_y + I_t \cdot I_t = 0 \quad (2-1-1b)$$

Die Gleichung (2-1-1b) beschreibt eine Gerade im  $u$ - $v$ -Raum, in der Literatur auch als *velocity space* bezeichnet, die die Koordinatenachsen in den Punkten  $P(0, -I_t/I_y)$  und  $P(-I_t/I_x, 0)$  schneidet. Allerdings besteht für diese Gleichung keine eindeutige Lösung, da sie nur eine Bedingung für zwei gesuchte Parameter  $u$  und  $v$  formuliert und somit im Sinne von *Hadamard*, ein schlecht gestelltes Problem darstellt. Diese partielle Differentialgleichung hat eine Vielzahl von Lösungen, die durch die Vorgabe von Randwertbedingungen auf eine eindeutig bestimmte Lösung eingeschränkt werden können.

Zur Lösung dieses Problems wurden in der Literatur diverse Verfahren vorgeschlagen, die sich im Wesentlichen in *lokale* und *globale* Methoden unterteilen lassen. Während lokale Verfahren nur eine feste Pixelnachbarschaft verwenden und dadurch an Stellen ohne Texturinformation keine Bewegungsschätzung ermöglichen, propagieren globale Verfahren durch geeignete Glattheitsannahmen bezüglich der Lösung implizit Informationen durch das gesamte Bild und führen in jedem Fall zu einer vollständigen Schätzung des Verschiebungsvektorfeldes. Besonders bedeutende Vertreter der globalen und lokalen Methoden, sind die so genannten intensitätsbasierten Ansätze, die den optischen Fluss durch Minimierung eines Energiefunktionalen berechnen. Zur

<sup>6</sup> Im Gegensatz dazu würde bei der Verwendung eines nicht-linearen Grauwertmodells die Berücksichtigung des Fehlerterms  $e$  notwendig sein und würde in einer Differentialgleichung mit Koeffizienten höherer Ordnung resultieren.  $I_x, I_y$  und  $I_t$  wird als Kurzschreibweise für die Bildgradienten benutzt.

Verdeutlichung betrachten wir im folgenden das Horn-Schunck-Verfahren (globales Verfahren) und das Lucas-Kanada Verfahren (lokales Verfahren).

### 2.1.1 Horn-Schunck-Verfahren

Das *Horn-Schunck-Verfahren* (HS) bezieht neben der Bildwerttreue, als zweite Bedingung an den optischen Fluss, die Glattheit des Vektorfeldes  $(u, v)$  mit ein. Dies wird mit der Annahme begründet, dass benachbarte Oberflächenpunkte eines Objektes ähnliche Bewegungen ausführen. Diese Bedingung dient zur Regularisierung des schlecht gestellten Problems und benutzt *globale* Informationen für die Bewegungsschätzung. Formal entspricht die Bedingung nach Glattheit der Minimierung folgender Funktion:

$$F(\mathbf{v})_{\text{Glatt}} = \left( \frac{u(x, y)}{\partial x} \right)^2 + \left( \frac{u(x, y)}{\partial y} \right)^2 + \left( \frac{v(x, y)}{\partial x} \right)^2 + \left( \frac{v(x, y)}{\partial y} \right)^2 = |\nabla u|^2 + |\nabla v|^2 \rightarrow \min \quad (2-1-1c)$$

Die Berechnung des optischen Flusses wird nach Horn-Schunck durch die Minimierung der Gleichungen (2-1-1b) und (2-1-1c), die durch einen Wichtungsfaktor  $\alpha$  verknüpft sind (s. Gl. (2-1-1d)), erreicht. Somit ist  $\mathbf{v}$  dann der optische Fluss der Bildfolge  $I_k, I_{k+1}$ , wenn gilt:

$$F(\mathbf{v}) = \iint_{x, y} (F_{\text{HS}}(\mathbf{v}) + \alpha \cdot F_{\text{Glatt}}(\mathbf{v})) dx dy \rightarrow \min \quad (2-1-1d)$$

$$F(\mathbf{v}) = \iint_{x, y} ((u \cdot I_x + v \cdot I_y + I_t)^2 + \alpha \cdot (|\nabla u|^2 + |\nabla v|^2)) dx dy \rightarrow \min \quad (2-1-1e)$$

wobei  $\mathbf{v} = (u, v, \delta t = 1)^T$  das gesuchte Verschiebungsvektorfeld ist,  $F_{\text{HS}}(\mathbf{v})$  den Datenterm (Bildwerttreue) bezeichnet und  $F_{\text{Glatt}}(\mathbf{v})$  für den Glattheitsterm steht. Hierbei ist zu beachten, dass der erste Term aufgrund der Berechnung der partiellen Ableitungen der Helligkeitswerte ( $I_x, I_y$  und  $I_t$ ) sowohl positiv als auch negativ sein kann. Deshalb wird dieser Teil in Gleichung (2-1-1d) quadriert. Der zweite Teil, auch als Glattheitsbedingung bezeichnet, nimmt nur nichtnegative Werte an. Durch den Wichtungsparameter von  $\alpha > 0$  kann der Einfluss der beiden Funktionsterme festgelegt werden. Während bei einem kleinem  $\alpha$  in der späteren Optimierung mehr Wert auf die Glattheit des Vektorfeldes gelegt wird, führt ein höherer Wert des Parameters zu einer Ausrichtung des Vektorfeldes an die Bildwerttreue. Die Berechnung des Vektorfeldes entspricht nun einer Minimierungsaufgabe der Funktion  $F$  in der Gleichung (2-1-1d). Es wird das Minimum der Fehlerterme, d.h. der Fehler der Glattheit des Vektorfeldes und der Fehler der Bildwerttreuen, gesucht.

Während im Datenterm Annahmen über Konstanz von Objektmerkmalen formuliert werden, z.B. die Konstanz des Grauwertes oder der ersten räumlichen Ableitung des Grauwertes, verkörpert der Glattheitsterm Annahmen über die stückweise Glattheit des gesuchten Verschiebungsvektorfelds<sup>7</sup>. Aus der Variationsrechnung ist bekannt, dass die Lösung dieser Gleichung durch die so genannten Euler Lagrange-Gleichungen bestimmt ist. Diese im Detail zu erläutern, würde leider den Rahmen dieser Arbeit sprengen, so

---

<sup>7</sup> Unter dieser Annahme macht es Sinn, bewusst auf die übliche Linearisierung der Konstanzannahmen bei größeren Verschiebungen zu verzichten und eine wesentlich komplexere Optimierung in Kauf zu nehmen. Durch Minimierung dieses Funktionals wird dann dasjenige Flussfeld als Lösung bestimmt, das optimal bezüglich aller berücksichtigten Modellannahmen ist.

## 2. INTENSITÄTSBASIERTE ANSÄTZE

dass an dieser Stelle nur auf einige Aspekte eingegangen wird. Bei Interesse ist die Ableitung in [Klette 96, Meyberg 01] nachzulesen.

Für den diskreten Fall werden die Bildpunkte bei ganzzahligen Koordinaten  $(i,j)$  für ein Bild der Größe  $W \times H$  betrachtet. Hierzu werden die ersten Ableitungen des Gesamtfehlers Gl. (2-1-1e) nach  $u_{ij}$  und  $v_{ij}$  gebildet, diese gleich Null gesetzt, und schließlich die gesuchten Funktionen  $u$  und  $v$ , für die F genügt, aus den resultierenden Gleichungen bestimmt.

Die Diskretisierung nimmt die Ableitung der Helligkeitswerte an einer Stelle zwischen den beiden Frames  $k$  und  $k+1$  und zwischen den Pixeln  $(i,j)$  und  $(i+1, j+1)$  vor. Für diese Stelle werden jeweils vier Vorwärtsdifferenzen ermittelt. Da die Werte  $u_{ij}$  und  $v_{ij}$  von ihrer Vierernachbarschaft abhängig sind, bietet sich ein iteratives Lösungsschema an [Klette 98]. Mit diesem Lösungsschema kann der Fehler Schritt für Schritt seinem Minimum angenähert werden. Da sich die jetzige Lösung immer auf die Lösung eines Zeitschrittes zuvor bezieht, ist es nicht möglich, das absolute Minimum mit einem Rechenschritt zu erreichen. Die Gleichungen müssen daher iterativ gelöst werden und somit wird das HS-Verfahren durch die Anzahl der Iterationen und der Wichtungsfaktor  $\alpha$  beeinflusst.

$$u_{ij}^{(n+1)} = \bar{u}_{ij}^n - I_x \left( \frac{I_x \bar{u}_{ij}^n + I_y \bar{v}_{ij}^n + I_t}{\alpha + I_x^2 + I_y^2} \right) \quad \text{und} \quad \bar{u}_{ij} = u_{i,j+1} + u_{i-1,j} + u_{i+1,j} + u_{i,j-1} \quad (2-1-2a)$$

$$v_{ij}^{(n+1)} = \bar{v}_{ij}^n - I_y \left( \frac{I_x \bar{u}_{ij}^n + I_y \bar{v}_{ij}^n + I_t}{\alpha + I_x^2 + I_y^2} \right) \quad \text{und} \quad \bar{v}_{ij} = v_{i,j+1} + v_{i-1,j} + v_{i+1,j} + v_{i,j-1} \quad (2-1-2b)$$

Dabei ist  $n$  die Iterationsanzahl.  $u^0$  und  $v^0$  deuten auf eine initiale Vektorschätzung hin, welche den Wert Null aufweist.  $\bar{u}^n$  und  $\bar{v}^n$  deuten in dieser Gleichung auf den Mittelwert der Nachbarschaft von  $u^n$  und  $v^n$  hin. Der Faktor  $\alpha$ , der das Gewicht des Regularisierers zur optischen Fluß-Bedingung angibt, kann den Grad der Unschärfe beeinflussen. Eine Auswahl an optischen Fluss-Vektorfeldern mit unterschiedlichem  $\alpha$  findet sich im Abschnitt 2.1.4. Die Annahmen, dass Bewegungen im Bild meist glatt sind, die letztlich zur Formulierung des Regularisierers geführt haben, begründen somit auch das „Verwaschen“ der Informationen. Diese Modelllösung zeichnet sich durch seine Eigenschaften als globales Verfahren aus, was eine 100%ige Vektorfelddichte, selbst bei stark homogenen Flächen bedeutet. Es wird also für jedes Pixel ein Verschiebungsvektor und somit eine Aussage über den optischen Fluss des jeweiligen Bildpunktes geliefert. Für die Randpunkte muss eine gesonderte Betrachtung erfolgen.

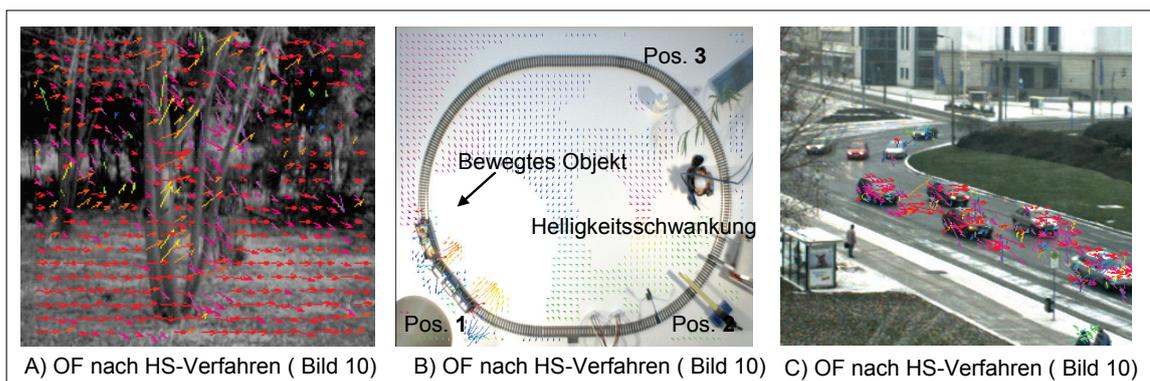
Die oben getroffenen Annahmen, die zu der Formulierung des HS-Verfahrens geführt haben, zeigen gleichzeitig auch dessen größtes Problem. An Objektgrenzen unterscheidet sich der optische Fluss benachbarter Bildpunkte der unterschiedlichen Objekte. Die HS-Bedingung glättet über diese Grenzen hinweg, was zu einem Verlust von Informationen führt.

Eine Vielzahl von Verfahren geht über diese HS-Annahme hinaus, bei der eine Glattheit der Bewegung eng mit dem Grauwertverlauf verbunden ist. Nagel schlägt in diesem Zusammenhang eine „gerichtete Glattheit“ vor, die nur entlang von geradlinigen Grauwertübergängen gefordert wird, um die Ermittlung der beiden Komponenten des Verschiebungsvektors zu gewährleisten [Nagel 85]. Das bestimmte, quasi kontinuierliche Verschiebungsvektorfeld ist aber gegen Rauschen und Beleuchtungsänderungen

## 2. INTENSITÄTSBASIERTE ANSÄTZE

empfindlich. Im Gegensatz dazu stellen Weickert und Schnörr [Weickert 00] einen isotropen, bildbasierten Regularisierer (*richtungsunabhängig*) vor, dem die Annahme zugrunde liegt, dass Kanten im Bild im Allgemeinen eine Objektgrenze und damit auch oft eine Bewegungsgrenze anzeigen. Dies gilt vor allem für nicht stark texturierte Bilder. An diesen Grenzen wird nun wenig geglättet, wohingegen innerhalb von Objekten, also an Stellen, an denen im Bild keine starken Grauwertänderungen vorliegen, stark geglättet wird. Der bildbasierte isotrope Regularisierer lässt sich zu einem anisotropen Regularisierer erweitern. Dieser Regularisierer hat den Vorteil, dass er entlang aller Bildkanten glättet, senkrecht zu ihnen jedoch nicht. Die Verwendung dieses Regularisierers innerhalb der optischen Fluss-Gleichungen ist in [Weickert 00] erläutert.

Ein mit dem vorgestellten HS-Verfahren berechneter optischer Fluss ist in Abbildung 2-1-1 dargestellt. In diesem Vektorfeldbild kann man die sich bewegenden Objekte leicht erkennen. Somit ist das Ergebnis schon mit dem einfachsten der hier vorgestellten HS-Regularisierer durchaus nicht nutzbar. Unschön an diesem Ergebnis sind lediglich die geglätteten Umrisse des Baumes und Ausreißer A) und Objekte in B und C), die ihre Identifizierung aufgrund Ausreißer erschweren. Solche "verschwommenen" Ergebnisse sind typisch für diesen Regularisierer bei den gestörten Bildsituationen.



**Abbildung 2-1-1:** Die Abbildungen zeigen die Vektorfelder für den optischen Fluss (OF) mit dem HS-Verfahren. Für die Regularisierungsparameter wurden jeweils die Werte  $\alpha=10$  und 50 Iterationen gewählt. Man erkennt den stark glättenden Effekt des HS-Verfahrens. Für die Bildfolgen A, B und C sind die Ergebnisse des HS-Verfahrens besonders enttäuschend, was an der instabilen Diskretisierung und größeren Pixelverschiebung liegen kann. An Position (1) ist eine Lampe, die eine starke Helligkeitsschwankung hervorruft. An Position (2) und (3) in B sind Teilverdeckung und Schattenwurfeffekte.

**Diskussion:** Durch die Berechnung des optischen Flusses ist das HS-Verfahren allerdings sehr rauschanfällig und die daraus folgende Annahme über die Konstanz des Grauwertes ist nicht mehr zu gewährleisten. Dazu enthält das durch das HS-Verfahren generierte Bewegungsvektorfeld Vektoren geringer Güte<sup>8</sup> (z.B. Ausreißer). Liegt beispielsweise das Fehlermaß zwischen zwei korrespondierenden Punkten oder Bildern über einem definierten mittleren Fehlermaß, dann wird der betreffende Vektor nicht mehr zur weiteren Analyse herangezogen. Eine Folge dieser Vorgehensweise ist, dass

<sup>8</sup> Deshalb empfiehlt es sich entweder eine Tiefpass- oder Medianfilterung in der Vorverarbeitung zu benutzen oder eine Bereinigung des Verschiebungsvektorfeldes anhand der Betrachtung der Fehlerenergie oder der Regionshomogenitäten durchzuführen. Die Fehlermaße können sich auf räumliche Informationen oder auch auf einen zeitlichen Kontext beziehen.

Bildbereiche, für die keine bzw. eine schlechte Punktkorrespondenz gefunden wurde, nicht mehr zur Analyse beiträgt. Dazu kommt, dass auch die lokale Bewegungsinformation nicht genau erfasst und abgeschwächt wird. Aufgrund von Aufdeckungen, Verdeckungen und Bereichen starker Homogenität der Intensität, können nicht immer Punktkorrespondenzen zwischen benachbarten Bildern gefunden werden. Bei Szenen mit sich unterschiedlich bewegenden Objekten, die somit der Glattheitsbedingung widerstreben, ist mit systematischen Fehlern, vor allem an den Objektändern oder in Bereichen mit einem harten Objektübergang (ohne Unschärfe), zu rechnen. Im folgenden Abschnitt wird dies genauer betrachtet. Eine weitere Schwachstelle des Verfahrens liegt in der Forderung der Bildwerttreue begründet und lässt sich auch nicht eliminieren. Eine weitere mögliche Alternative zur Ermittlung der Verschiebungsvektorfelder bildet bei hohen Stabilitätsforderungen das Lucas-Kanade-Verfahren, das sich *lokal* dem gesuchten Verschiebungsvektor annähert.

### 2.1.2 Lucas-Kanade-Verfahren

Die grundlegende Idee des Lucas-Kanade-Verfahrens (LK) ist es, einen Signalabschnitt in einem Fenster  $\Omega$  vom Zeitpunkt  $k$  im Zeitpunkt  $k+1$  wieder zu finden und dessen Verschiebung zu bestimmen. Anhand der Signalmerkmale des Fensters  $\Omega_k$  wird ein Fenster  $\Omega_{k+1}$  gesucht, das im nächsten Zeitpunkt die gleichen Signalmerkmale aufweist. Die Suche erfolgt in der lokalen Umgebung von  $\Omega_k$  und ordnet das LK-Verfahren somit den *lokalen* Verfahren<sup>9</sup> zu. Für den zweidimensionalen Fall ergibt sich eine Kostenfunktion, die der HS-Bedingung entspricht. Die Verschiebungen  $u$  und  $v$  werden mit dem Vektor  $\mathbf{h}=(u \ v)^T$  zusammengefasst. Die Differenz  $I_k(x,y)-I_{k+1}(x,y)$  kann als zeitliche Grauwertverschiebung  $I_t(x,y)$  interpretiert werden. Die Ableitungen von  $I(x,y)$  werden mit dem Differenzoperator gebildet, so dass nun gilt:

$$F_{LK} = \sum_{\Omega} (I_2(x+u, y+v) - I_1(x, y))^2 \approx \sum_{\Omega} (I_2(x, y) + u \cdot \frac{\partial I_2}{\partial x} + v \cdot \frac{\partial I_2}{\partial y} - I_1(x, y))^2 \quad (2-1-3a)$$

$$F_{LK}(\mathbf{h}) = F_{LK}(u, v) = \sum_{\Omega} (I_x u + I_y v + I_t)^2 = \sum_{\Omega} (\nabla I^T \cdot \mathbf{h} + I_t)^2 \Rightarrow \min \quad (2-1-3b)$$

Setzt man die erste Ableitung Null, um das Minimum zu erhalten, folgt:

$$\frac{\partial F_{LK}}{\partial u} = 2 \sum_{\Omega} I_x (\nabla I^T \cdot \mathbf{h} + I_t) = 0 \quad \text{und} \quad \frac{\partial F_{LK}}{\partial v} = 2 \sum_{\Omega} I_y (\nabla I^T \cdot \mathbf{h} + I_t) = 0 \quad (2-1-3c)$$

Durch das Ausmultiplizieren der Summe lassen sich die Gradientenmatrix  $\mathbf{G}$  und der Bildfehlervektor  $\mathbf{b}$  definieren, so dass sich der Verschiebungsvektor aus der inversen Gradientenmatrix  $\mathbf{G}^{-1}$  und dem Bildfehlervektor berechnen lässt (Gl. 2-1-3e):

$$\sum_{\Omega} \begin{bmatrix} I_x^2 & I_x I_y \\ I_x I_y & I_y^2 \end{bmatrix} \cdot \begin{bmatrix} u \\ v \end{bmatrix} = - \sum_{\Omega} \begin{bmatrix} I_x I_t \\ I_y I_t \end{bmatrix} \quad \text{d.h.} \quad \mathbf{h} = \mathbf{G}^{-1} \cdot \mathbf{b} \quad (2-1-3d)$$

$$\mathbf{G} = \sum_{\Omega} \begin{bmatrix} I_x^2 & I_x I_y \\ I_x I_y & I_y^2 \end{bmatrix}, \quad \mathbf{b} = - \sum_{\Omega} \begin{bmatrix} I_x I_t \\ I_y I_t \end{bmatrix} \quad (2-1-3e)$$

<sup>9</sup> Wie bei dem Blockmatching-Verfahren entspricht die Verschiebung  $\mathbf{v}$  der Bewegung des Signals im Fenster  $\Omega_k$ , wenn die Fehlerfunktion minimal wird. Dabei wird im Gegensatz zum Blockmatching die Verschiebung der Intensitätsfunktion  $I(x,t)$  durch die Annahme der Linearität über die Taylorreihenentwicklung erster Ordnung geschätzt.

## 2. INTENSITÄTSBASIERTE ANSÄTZE

---

Die Gradientenmatrix  $\mathbf{G}$  verdeutlicht das Problem der differentiellen Verfahren. Der Flussvektor  $\mathbf{h}$  kann nur geschätzt werden, wenn die Matrix  $\mathbf{G}$  invertierbar ist, was wiederum nur dann gegeben ist, wenn  $\mathbf{G}$  nicht singular ist. Nach [Bronstein 01] ist die Berechnung der inversen oder reziproken Matrix  $\mathbf{G}^{-1}$  angewandt auf die Matrix  $\mathbf{G}$ , beschrieben durch:

$$\mathbf{G}^{-1} = \frac{1}{\sum_{\Omega} I_x^2 \cdot \sum_{\Omega} I_y^2 - \left(\sum_{\Omega} I_x I_y\right)^2} \sum_{\Omega} \begin{pmatrix} I_y^2 & -I_x I_y \\ -I_y I_x & I_x^2 \end{pmatrix} \quad (2-1-3f)$$

Ob die Matrix  $\mathbf{G}$  singular ist, lässt sich durch die Analyse ihrer Eigenwerte ableiten. Hier lassen sich drei Fälle für die beiden Eigenwerte  $\lambda_1$  und  $\lambda_2$  unterscheiden.

- Weisen die beiden Eigenwerte Werte größer als Null auf, ist  $\mathbf{G}$  nicht singular. Dies spricht für eine gute Merkmalsausprägung, die zur zuverlässigen Bestimmung der beiden Komponenten ( $u$ ,  $v$ ) des Flusses führt.
- Sind die beiden Eigenwerte gleich Null, ist  $\mathbf{G}$  singular. Hier lassen sich die beiden Komponenten des Flusses nicht berechnen. Dies ist der Fall, wenn keine ausgeprägte Primitiven/Gradienten vorhanden sind, also eine homogene Fläche betrachtet wird.
- Falls einer der beiden Eigenwerte größer als Null ist, ist  $\mathbf{G}$  auch singular. Dies spricht für eine Kante. Die Richtung der Kante ist orthogonal zum Eigenvektor des größeren Eigenwertes. Hier lässt sich nur eine Komponente des Flusses berechnen. Dies entspricht dem bekannten Apertur-Problem.

Durch die Betrachtung der Eigenwerte lässt sich also daraus auch schließen, ob das Apertur-Problem vorliegt oder ob die Struktur der Bilddaten eine zuverlässige Berechnung des optischen Flusses zulässt. Verglichen mit dem HS-Verfahren zeichnet sich das vorgestellte LK-Verfahren durch eine hohe Robustheit gegenüber Rauschen aus. Dies ist darauf zurückzuführen, dass das Fenster typischerweise ausgeprägte Primitiven beinhalten wird, und dadurch die Wahrscheinlichkeit für eine fehlende Textur in Realbildern mit wachsender Fenstergröße abnimmt. Dabei besitzt die Größe des Abtastfensters  $\Omega$  eine entscheidende Rolle, da dadurch die lokale Charakteristik einen Einfluss auf die Berechnung eines Vektors hat. Sind keine Primitiven im Abtastfenster (d.h.  $I_x = 0 \vee I_y = 0 \quad \forall I(x, y) \in \Omega$ ) vorhanden, so kann der Verschiebungsvektor aufgrund des sog. Apertur-Problems nicht berechnet werden.

Durch den lokalen Charakter dieses Verfahrens kann jedoch im Gegensatz zu dem HS-Verfahren keine 100%ige Vektorfelddichte erreicht werden. Dies kann dazu führen, dass eine genaue Objektbegrenzung durch die Analyse von Verschiebungsvektoren schwerer wird. Dies führt wiederum zu einer ungenauen Objektbegrenzung und in der praktischen Anwendung zu einer Verschmelzung von sich nah gelegenen bewegenden Objekten wie wir in den weiteren Abschnitten sehen werden. Darüber hinaus kommt die zunehmende Ungenauigkeit bei großen Pixelverschiebungen, wie auch *texturiertem Hintergrund*. Dieses beruht auf der Approximation der Bildwerttreue durch die Taylorreihenentwicklung, die die Linearität an dem zu entwickelnden Punkt fordert, sowie auf den fehlenden korrespondierenden Punkten. Um den Einfluss dieser genannten Probleme bei der Bestimmung des Verschiebungsvektors zu minimieren, ist die *Pyramidenimplementation* vom LK-Verfahren von großer Bedeutung, die im nächsten Abschnitt kurz erläutert und auf die Eignung und Robustheit auch untersucht wird.

### 2.1.3 Pyramidales LK-Verfahren

Die bisher beschriebenen Verfahren zur optischen Fluss schätzung benutzen die Approximation der Bildwerttreue zwischen zwei Bildern als Maß für die Bewegung. Zur Ermittlung der Flussvektoren ist dies richtig, solange die Helligkeit eines Objektes zwischen zwei Bildern sich nicht ändert und der Grauwertverlauf linear ist. Diese Annahmen sind jedoch nicht immer erfüllt, da in praktischen Anwendungen zum einen unvermeidliche Beleuchtungsänderungen auftreten und zum anderen größere Verschiebungen vorkommen können, die zum Korrespondenzproblem führen. Interessant ist die Einführung neuer Verfahren, welche die robuste Bestimmung der Flussvektoren eines Objektes bei diesen Effekten gewährleisten. Hier spielen so genannte *Coarse-to-Fine-Warping-Strategien* eine wichtige Rolle, da sie zuerst die Schätzung großer Verschiebungen erlauben und anschließend eine höhere Stabilitätsanforderung aufgrund der einbezogenen Primitiven stellen.

Während das Coarse-to-Fine-Konzept sicherstellt, dass die Verschiebungen auf der größten Auflösungsstufe sehr klein sind, sorgt die Warping-Strategie dafür, dass dies auch für die zu berechnenden Verschiebungsinkremente so bleibt. Da größere Verschiebungen ungenauer beim einfachen LK-Verfahren ermittelt werden können als kleinere, nimmt durch den Einsatz einer solchen Coarse-to-Fine-Warping-Strategie die Schätzqualität deutlich zu. Damit sind solche Strategien ein geeignetes Mittel, um den bewussten Verzicht auf eine Linearisierung der Konstanzannahmen im Term algorithmisch umzusetzen. Die Idee besteht darin die relativen Verschiebungen gering zu halten. Es ist nun aber nicht möglich die gegebene Bewegung zu beeinflussen, also verkleinert man das Bild. Somit verkleinern sich auch die relativen Pixelverschiebungen. Im Idealfall sollten die maximalen Pixelverschiebungen nicht größer als zwei bis drei Pixel sein. Der Ansatz wird mit der Stufenweisen Anwendung des optischen Fluss Algorithmus auf eine Bildpyramide, wie sie auch in [Jähn 02, Bouguet 00] beschrieben ist, verwirklicht. Die Pyramidenstruktur des LK-Verfahrens (PLK-Verfahren) wird rekursiv gebildet, sodass  $I^l$  aus  $I^0$  und  $I^2$  aus  $I^1$  gewonnen werden. Das Bild  $I^L$  sei weiterhin definiert durch die Breite  $w^L$  und die Höhe  $h^L$  des Bildes. Die Größe jedes Bildes ändert sich jeweils um  $w^L = (w^{L-1} + 1)/2$  und  $h^L = (h^{L-1} + 1)/2$ . Der Bewegungsvektor wird Top-Down durch das LK-Verfahren bestimmt, wobei die zu minimierende Funktion für das Level  $L$  durch den Verschiebungsvektor  $\mathbf{g}^L$  erweitert wird. Das Ziel ist die Ermittlung des Verschiebungsvektors  $\mathbf{h} = \mathbf{h}^0 + \mathbf{g}^0 = [u + g_x, v + g_y]^T$  in dem zweiten Bild  $J^L(\mathbf{h})$ , wenn  $I(\mathbf{h}^0)$  und  $J(\mathbf{h})$  ähnlich sind. Für Herleitungen und Details sei an dieser Stelle auf [Bouguet 00]<sup>10</sup> verwiesen.

$$F_{LK}^L(\mathbf{g}^L) = F_{LK}^L(g_x^L, g_y^L) = \sum_{\Omega} (I^L(x, y) - J^L(x + u^L + g_x^L, y + v^L + g_y^L))^2 \rightarrow \min \quad (2-1-4a)$$

$J^L$  ist das zweite Bild. Dieser ergibt sich aus der nächsthöheren Ebene durch Gleichung (2-1-4b) mit dem initialen Vektor  $\mathbf{g}^{Lm}$  der obersten Ebene. In der untersten Ebene  $L=1$  befindet sich der Verschiebungsvektor  $\mathbf{h}$ . Der durch die Pyramidenstruktur vergrößerte Bereich kann dennoch als ein lokaler Suchbereich betrachtet werden.

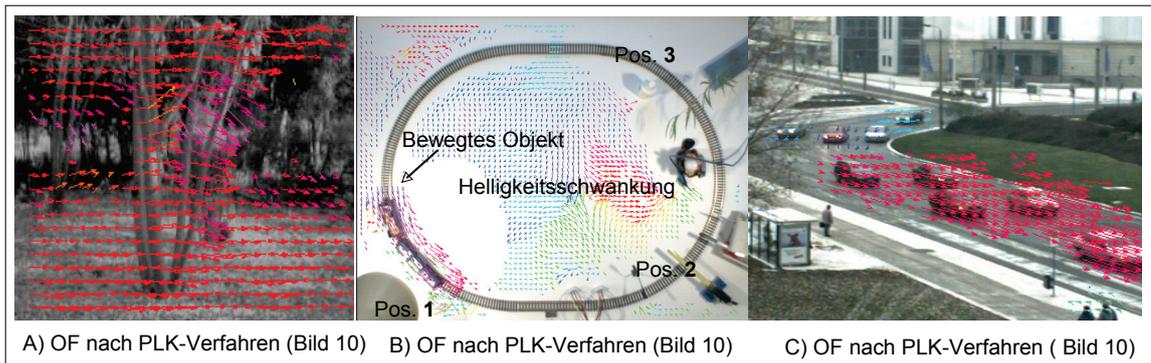
---

<sup>10</sup> Bouguet, J.-Y.: Pyramidal Implementation of the Lucas Kanade Feature Tracker. Part of OpenCV Documentation.

## 2. INTENSITÄTSBASIERTE ANSÄTZE

$$\mathbf{g}^{L-l} = 2(\mathbf{g}^L + \mathbf{h}^L), \quad \mathbf{g}^{Lm} = [0 \ 0]^T \quad \text{und} \quad \mathbf{h} = \mathbf{h}^0 + \mathbf{g}^0 \quad (2-1-4b)$$

Die pyramidale Lösungsstrategie ermöglicht es, das LK-Verfahren auch auf größere Verschiebungen anzuwenden ohne Abstriche an dessen Robustheit zu machen (s. Abb. 2-1-2). Jedoch bleibt der lokale Charakter erhalten. Da kleine Verschiebungen bei ausreichenden Texturen viel genauer berechnet werden können als größere, nimmt durch den Einsatz einer Coarse-to-Fine-Warping-Strategie die Schätzqualität im Allgemeinen deutlich zu (Abb. 2-1-2). Allerdings hängt die Schätzqualität von der Größe des Abtastfensters und der Anzahl der Level der Bildpyramide ab, die in dieser Arbeit optimiert wird.



**Abbildung 2-1-2:** Die Abbildungen zeigen die Vektorfelder für den optischen Fluss (OF) mit PLK-Verfahren. Für die Regularisierungsparameter wurde jeweils die Fenstergröße auf 10 und die Anzahl der Pyramidenstufen auf 3 gewählt. Ähnliche Ergebnisse erhält man mit Fenstergröße 5 und 3 Pyramidenstufen. Man erkennt den stark glättenden Effekt des PLK-Verfahrens. Für die Bildfolgen A), B) und C) sind die Ergebnisse des PLK-Verfahrens gut. An Position (1) ist eine Lampe, die eine starke Helligkeitsschwankung hervorruft. An Position (2) und (3) sind Teilverdeckung und Schattenwurfeffekte.

### 2.1.4 Bewertung der vorgestellten Verfahren

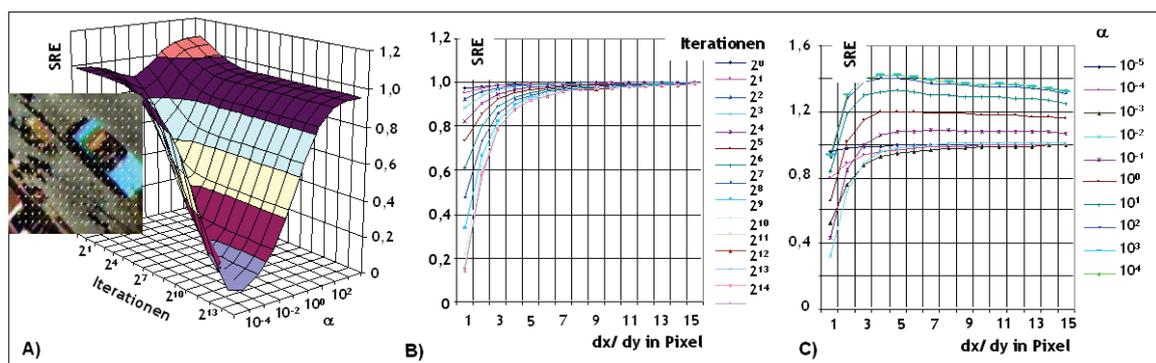
Nachdem in den beiden vorangegangenen Abschnitten die wesentlichen Eigenschaften der lokalen und globalen Verfahren besprochen worden sind, soll in diesem Abschnitt sowohl die Schätzgenauigkeit als auch die algorithmische Effizienz der vorgestellten Verfahren verdeutlicht werden. Eine Möglichkeit, um die Qualität der eingeschätzten Flussvektoren zu bewerten, besteht darin die Abweichung eines realen Verschiebungsvektors vom idealen Flussvektor zu ermitteln. Dafür können hier Ähnlichkeitskriterien als Fehlerkriterium zwischen dem Vektorfeld der Grundwahrheit  $\mathbf{v}_c$  und dem geschätzten Vektorfeld  $\mathbf{v}_e$  angewendet werden, die sich auf die Länge oder den Winkel zwischen Flussvektoren in aufeinander folgenden Bildern beziehen können. Da ein einzelner Flussvektor für sich allein gestellt nicht sehr aussagekräftig ist, werden in der Regel die normalisierten Summen der Winkelfehler (*Sum of all relative Angular Errors, SAE*) oder die normalisierten Summen aller relativen Fehler (*Sum of all Relativ Errors, SRE*) aus Gleichung (2-1-4a) von N-Vektoren angewendet [Klette 96].

Hier wird das SRE-Kriterium für die Bewertung der Genauigkeit des berechneten Vektorfeldes verwendet, da es Fehler bei kleinen Bewegungen sensibler wiedergibt, und die punktbasierten differentiellen Verfahren nur kleine Bewegungen genau erfassen können.

## 2. INTENSITÄTSBASIERTE ANSÄTZE

$$SRE = \frac{1}{N} \cdot \sum_N \frac{\|\mathbf{v}_e(x,y) - \mathbf{v}_c(x,y)\|}{\|\mathbf{v}_c(x,y)\|} \quad (2-1-4a)$$

Um die wesentlichen Verhaltensmerkmale des HS-Verfahrens zeigen zu können, wird in dem *ersten Szenario*<sup>11</sup> (gleichmäßige Verschiebung eines gesamten Bildinhaltes) ein Parameterdurchlauf der Iteration und des Wichtungsfaktors  $\alpha$  bei einer Verschiebung von einem Pixel durchgeführt (Abb. 2-1-3). Es zeigt sich, dass es sinnvoll ist, den Wichtungsfaktor  $\alpha$  klein zu halten, wenn ein homogenes optisches Feld angenommen werden kann und keine lokalen Bewegungen zu erwarten sind. Das Glättungskriterium aus Gleichung (2-1-1c/d) erhält somit höhere Bedeutung. Da die Bildwerttreue kaum noch Einfluss auf die Optimierung hat, werden dann jedoch mehrere Iterationen benötigt, um den Algorithmus konvergieren zu lassen. Der SRE-Fehler erreicht ein Minimum an der Stelle des Wichtungsfaktors ( $\alpha$ ) bei ca.  $10^{-4}$  (in Abb. 2-1-3), das sich mit steigender Iteration verkleinert. Die Verringerung des SRE-Fehlers konvergiert dabei in Abhängigkeit des Wichtungsfaktors. Mit zunehmenden Wichtungsfaktor konvergiert der SRE später. So ändert sich bei einem  $\alpha$  von 10 der Fehler nach 24 Iterationen und bei einem  $\alpha$  von  $10^{-4}$  nach  $2^{14}$  Iterationen kaum merkbar. Mathematisch kann allerdings für das Verfahren keine sichere Konvergenz nachgewiesen werden.



**Abbildung 2-1-3:** A) Der SRE-Fehler für das generierte Vektorfeld im dargestellten Bild mittels HS-Verfahrens ist bei unterschiedlichen Iterationen und Gewichtungsfaktor  $\alpha$  dargestellt, wobei die Anzahl der Iterationen und der Wichtungsfaktor dabei je für sich konstant gehalten werden. Bei einer Verschiebung von zwei Pixeln in  $x$ - und  $y$ -Richtung vergrößert sich der SRE-Fehler stark.

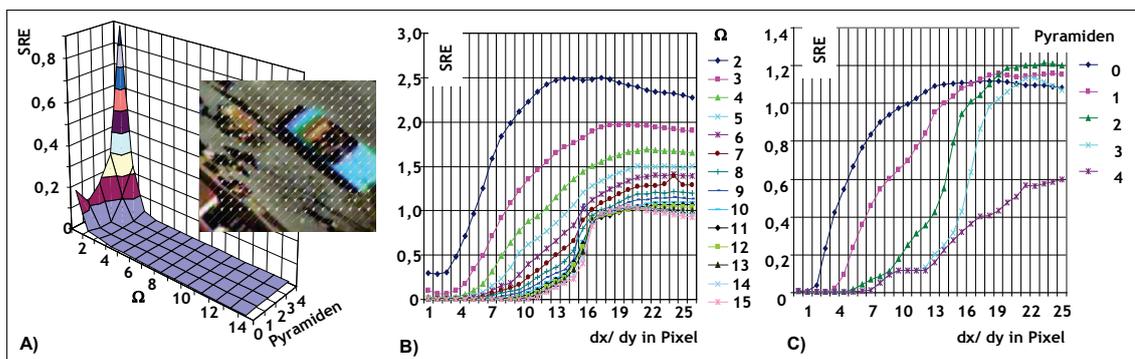
Interessant ist nun die Auswirkung der Verschiebung  $dx$  und  $dy$  auf das Messergebnis. Die Anzahl der Iterationen und der Wichtungsfaktor werden dabei jeweils konstant gehalten. Bei einer Verschiebung von zwei Pixeln in  $x$ - und  $y$ -Richtung vergrößert sich der relative Vektorfehler stark (Abb. 2-1-3B/C). Die Ursache dieses Fehleranstieges liegt in der Annahme der Linearität des Grauwertverlaufes der Bildfunktion  $I(x,y)$  und der dadurch getroffenen Approximation aus Gleichung (2-1-1a/b). Daraus lässt sich erkennen, dass mit dem HS-Verfahren nur recht kurze Verschiebungsvektoren korrekt bestimmt werden können, wenn eine ausreichend fein strukturierte Textur vorhanden ist. Diese gibt dem Verfahren lokale Stabilität, solange sich diese Textur nicht unterhalb der Verschiebungsdistanzen periodisch wiederholt.

Im Vergleich zum HS-Verfahren kommt das LK-Verfahren relativ besser mit Verschiebungen von mehr als einen Pixel zurecht (Abb. 2-1-4B). Eine sehr starke Fehlerzunahme

<sup>11</sup> Das erste Szenario stellt die gleichmäßige Verschiebung eines gesamten Bildes dar. Das Bild enthält die Aufnahme einer Verkehrsszene, sodass die Grauwertverteilung der eines natürlichen Geschehens entspricht.

## 2. INTENSITÄTSBASIERTE ANSÄTZE

me ist jedoch bei größeren Verschiebungen zu verzeichnen, sodass keine aussagekräftigen Ergebnisse zu erwarten sind. Die Fehlerschätzung wird durch die Approximation der Differentiation, die nur unter der Bedingung von kleinen Verschiebungen eine gute Näherung ist, hervorgerufen. Dadurch verfängt sich der Algorithmus oft in lokalen Minima, die nicht der Lösung entsprechen. Im Extremfall entstehen völlig unkorrelierte Vektoren. Zusätzlich hat die Größe des Abtastfensters<sup>12</sup>, von dem aus der Verschiebungsvektor bestimmt wird, einen Einfluss auf die Exaktheit und die Auflösung des resultierenden Verschiebungsvektorfeldes.



**Abbildung 2-1-4:** A) Das errechnete Vektorfeld im Bild mit 3 Pyramidenstufen und Abtastfenstergröße von 4 Pixeln. In B) und C) ist der SRE-Fehler bei unterschiedlichen Abtastfenstergrößen und Pyramidenstufen dargestellt. Der SRE-Fehler besitzt mit einer großen Anzahl von Pyramidenstufen und großem Abtastfenster sein Minimum. Es zeigt sich, dass der SRE-Fehler auch noch bei großen Pixelverschiebungen sehr gering ist, da in den höheren Pyramiden durch den homogenen Hintergrund, ein Objekt immer noch gut detektiert werden kann. Eine Pixelverschiebung, die größer als 2 Pixel ist, sollte beim einfachen LK-Verfahren (Pyramide 0) vermieden werden.

Die Erweiterung dieses LK-Verfahrens durch die Auflösungspyramiden und der optischen Flusschätzung nach einer Top-Down-Strategie löst das Problem bei großen Pixelverschiebungen, wie sich anhand des relativen SRE-Fehlers zeigen lässt (Abb. 2-1-4C). Wählt man ein kleines Abtastfenster bei einer hohen Anzahl an Pyramiden, wächst der Fehler wieder, wie in Abbildung 2-1-4 ersichtlich wird, an. Es ist zu beachten, dass mit dem Aufbau der Bildpyramide Ebene für Ebene auch die Primitiven und somit der eigentlichen Informationsgehalt des Bildes verloren gehen. Bei einem kleinen Abtastfenster kann es dann zum Apertur-Problem<sup>13</sup> kommen und der Flussvektor kann dort nicht bestimmt werden. Diese entspricht genau dem Fall, wenn die G-Matrix singular ist. Mit einer großen Pyramidenzahl und großem Abtastfenster besitzt der SRE-Fehler sein Minimum. Es zeigt sich, dass auch diesem Algorithmus Grenzen in der Erfassung von sehr großen Verschiebungen gegeben sind. Diese liegen bei 20 Pixeln (s. Abb. 2-1-4). Das jedoch deckt die meisten Fälle in natürlichen Szenen ab.

Wird ein Objekt mit Textur über einen homogenen Hintergrund (*Szenario II*) bewegt, wird durch den Hintergrund eine größere Luminanzdifferenz durch die Homogenität erzielt, da die Gradienten  $I_x$  und  $I_y$  auf einem homogenen Hintergrund gleich Null sind und das Objekt Bildbereiche mit ausgeprägten Grauwertgradienten beinhaltet. Dies hat aber auf den ersten Blick keinen Einfluss auf die Berechnung des optischen Flusses des

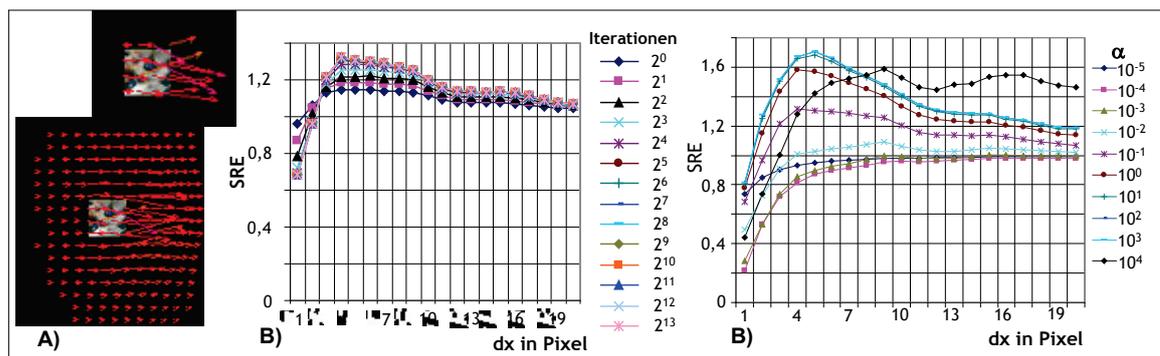
<sup>12</sup> Ein großes Abtastfenster verleiht dem Verfahren höhere Robustheit gegenüber Rauschen und erhöht die Anzahl der in ihm enthaltenen ausgeprägten Primitiven.

<sup>13</sup> Homogener Bildbereich ohne ausgeprägte Grauwertgradienten.

## 2. INTENSITÄTSBASIERTE ANSÄTZE

Objektes. Bei der Berechnung der Verschiebungsvektorfelder mittels HS-Verfahrens wird durch den homogenen Hintergrund die iterative Vektoranpassung nur durch das Glattheitskriterium bestimmt<sup>14</sup>. Es entsteht ein Vektorfeld, das sich in den informationslosen Raum (Bereiche, in denen es keine Gradienteninformation gibt (s. Abb. 2-1-5a)) ausbreitet.

In Hinblick auf die Erkennung von sich bewegenden Objekten anhand des optischen Flusses ist dieser vektorfeldausbreitende Effekt interessant, da dadurch eine genaue Bewegungssegmentierung oder eine lokale, detaillierte Bewegungserfassung erschwert wird. Obwohl die Ausrichtung der Vektoren mit zunehmender Iteration verbessert wird, beruht der Zuwachs der Fehlerwerte auf der Ausbreitung des Vektorfeldes in die informationslosen Bereiche. Dieser Effekt wird als *Vektorfeldausbreitung* oder auch auf Grund seiner Form als *koronaler Effekt* bezeichnet.



**Abbildung 2-1-5a:** A) Das errechnete Vektorfeld mit 10 Iterationen (Oben) und unten mit 1000 Iterationen. In B) und C) ist der SRE-Fehler des HS-Verfahrens bei einer Verschiebung um einen Pixel, mit  $\alpha = 0.1$  (links) und 100 Iterationen (rechts) für das Szenario II dargestellt. Es fällt sofort auf, dass der Fehler mit ansteigender Iteration oder Wichtung größer wird. Das liegt an dem sich mit zunehmender Optimierung ausbreitenden Vektorfeld. Durch die 4er-Nachbarschaftsabhängigkeit breitet sich so das Vektorfeld mit fortlaufender Iteration aus. Die Ausrichtung der Vektoren wird jedoch mit zunehmender Iteration verbessert. Da jedoch die Bewegungsvektoren an Stellen erzeugt werden wo keine sind, steigt der Fehler an.

Der Effekt der Vektorfeldausbreitung wurde an einem realitätsnahen Beispiel untersucht (*Szenario III*)<sup>15</sup>. Hier zeigt sich bei dem HS-Verfahren, dass die Textur dem Ausweiten des Vektorfeldes entgegenwirkt, da nun auch Informationen im Hintergrund enthalten sind, die im Gegenzug zu homogenen Flächen den Ruhezustand beinhalten (Abb. 2-1-5). Wird jedoch der Faktor  $\alpha$  zu klein gewählt, hebt sich die hemmende Wirkung der Textur auf, was durch die Gleichungen (2-1-2a/b) gezeigt werden kann. Strebt der Wert  $\alpha$  gegen 0, so wird der positiv hemmende Einfluss der horizontalen und vertikalen Gradienten geschwächt und das Gesamtvektorfeld ist durch seine Glattheit und nicht durch seine Bildwerttreue definiert. Es ist auch nicht von Bedeutung, ob der Hintergrund Textur enthält oder homogen ist. Der HS-Algorithmus liefert bei Pixelverschiebungen von mehr als zwei Pixeln extrem schlechte Werte.

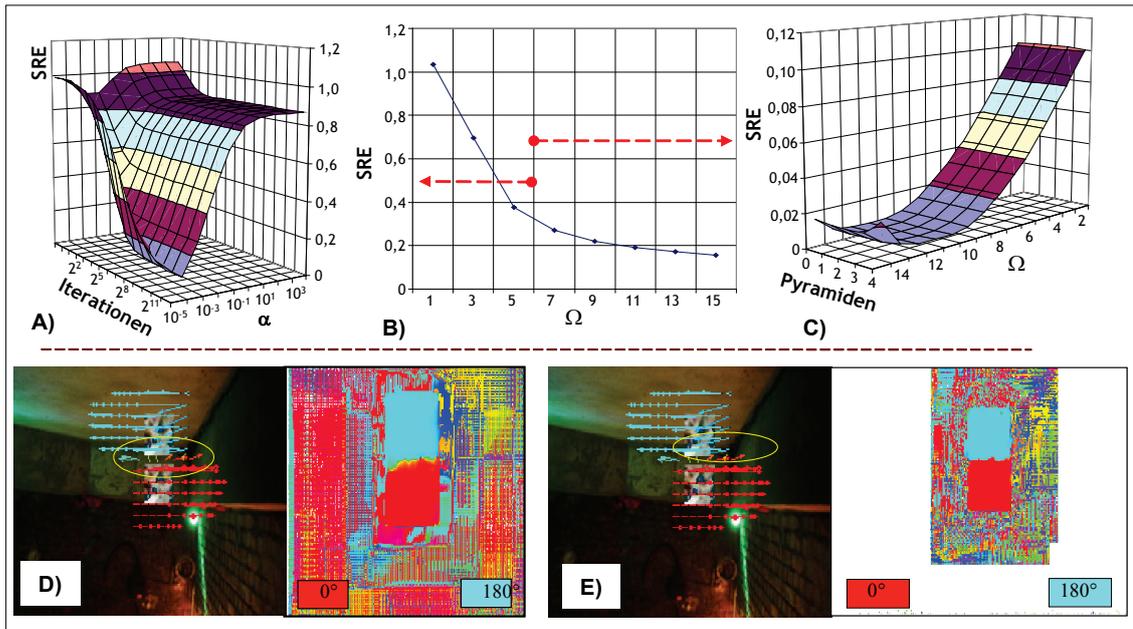
Nicht nur bei dem HS-Verfahren tritt diese Problematik der Vektorfeldausbreitung auf, sondern auch bei dem LK-Verfahren. Diese Ausbreitung von Vektoren ist aber im LK-

<sup>14</sup> Durch die Vierernachbarschaftsabhängigkeit breitet sich das Vektorfeld mit fortlaufender Iteration aus.

<sup>15</sup> Da es nur sehr selten homogene Hintergründe in Straßenszenen gibt, soll nun der Hintergrund mit einem Bild gefüllt werden, auf dem sich, dann das mit Textur versehene Objekt (aus Abbildung 2-1-5a) bewegt.

## 2. INTENSITÄTSBASIERTE ANSÄTZE

Verfahren durch die Größe des Abtastfensters begrenzt. Die Ursache für die Vektorfeldausbreitung liegt wieder im Fehlen von Gradienten im Hintergrund, da auch bei dem LK-Verfahren die Gradienteninformation als Grundlage für die Verschiebungsschätzung dient. Durch die Pyramidenstruktur tritt noch eine zusätzliche Verstärkung dieses Effektes ein. Bei der Übergabe der Verschiebungsvektoren von einer Auflösungsstufe zur nächstgrößeren wird der Fehler jeweils übernommen und entsprechend verstärkt. Das PLK-Verfahren weist ebenfalls ein ähnliches Verhalten auf. Die Abbildung 2-1-6 enthält zwei Objekte (*Szenario IV*), die sich mit einem Abstand von sechs Pixeln in entgegengesetzter Richtung bewegen. Der Fehlerverlauf, für das HS- und LK-Verfahren bleibt unverändert (Abbildung 2-1-6).



**Abbildung 2-1-6:** A) Der SRE-Fehler für das HS-Verfahren, LK-Verfahren B) und PLK-Verfahren C) für das Szenario IV. D) und E) Das errechnete Vektorfeld durch den PLK-Algorithmus, bei einer Abtastfenstergröße von 15 (links) und 5 (rechts), linkes Bild zeigt das Vektorfeld und das rechte den Winkel der Vektoren im Falschfarbenbild. Es zeigt, dass die Textur dem Ausweiten des Vektorfeldes entgegenwirkt, da an ihr keine Bewegung festgestellt werden kann, im Gegensatz zu homogenen Flächen. Auch bei dem PLK-Verfahren erkennt man die fast lineare Abhängigkeit des SRE von der Größe des Abtastfensters. Da pro Fenster nur ein Vektor geschätzt wird, wirkt sich dies als Mittelung aus und es können fehlerhafte Richtungsvektoren geschätzt werden, wie in D) und E) ersichtlich ist. Im linken Bild tritt eine falsche Bewegungsschätzung an der Grenzlinie im gelb markierten Bereich durch ein Abtastfenster der Größe 15 auf. Durch die Verkleinerung des Abtastfensters auf eine Größe von 6 kann dieser Fehler verkleinert werden.

Im Falle des PLK-Algorithmus zeigt sich jedoch die für ihn typische Vergrößerung des relativen Fehlers bei größer werdendem Abtastfenster. Man kann die Kurve aus Abbildung 2-1-6B in zwei Bereiche teilen. Der erste beschreibt eine Fenstergröße von zwei bis sechs und der zweite eine Fenstergröße von  $6 \times 6$  (6 in der Abbildung eingetragen) bis  $15 \times 15$ . Es ist dort ersichtlich, dass bei geringer Abtastfenstergröße von bis zu sechs Pixeln die Verschiebungsfelder in den Randregionen wesentlich nicht beeinflussen. Ein Abtastfenster von mehr als sechs Pixeln Breite enthält in der Region zwischen den beiden Quadraten Merkmale von beiden. Es kommt lokal gesehen zu einer

Interpolation der beiden einzelnen Vektorfelder (blau und rot) an ihren gemeinsamen Grenzen (Abb. 2-1-6D). Die Verschiebungsfelder beeinflussen sich gegenseitig. Es tritt neben dem koronalen Fehler ein zweiter auf, der die Verwischung zweier Verschiebungsvektorfelder verursacht.

### 2.1.5 Verbesserung des PLK-Verfahrens

In diesem Abschnitt soll der Effekt der Vektorfeldausbreitung des optischen Flusses näher untersucht werden, da die Genauigkeit der Bewegungsschätzung und die Separabilität der lokalen Bewegungen davon abhängen. Der Effekt der *Vektorfeldausbreitung* wirkt störend auf die weitere Verarbeitung<sup>16</sup>. Der entscheidende Faktor, der diesen auftretenden Effekt verursacht, ist die Größe des Abtastfensters. Da das Abtastfenster eine minimale Größe benötigt, um das Apertur-Problem zu vermeiden, muss ein Mittelweg zwischen der fehlerhaften Berechnung der Vektoren in Bezug auf ihre Ausrichtung und Größe, vorkommend bei kleinem Abtastfenster, und der fehlerhaften Berechnung der Objektträger durch die Vektorfeldausbreitung, hervorgerufen durch große Abtastfenster, gefunden werden. Zu diesem Zweck wird das PLK-Verfahren durch eine *Adaption der Abtastfenstergröße* und eine Reduzierung des Vektorfeldausbreitungseffektes zur Konturverbesserung erweitert. Diese Erweiterungen werden dann auf ihre Tauglichkeit hinsichtlich der Segmentierung von sich bewegenden Objekten analysiert.

#### 1. Adaption der Abtastfenstergröße

Bei der ersten Erweiterung werden die Gradienten eines Fensters einbezogen. Diese Herangehensweise wurde durch die Arbeit von Shi und Tomasi [Shi 94] motiviert. Da die genaue Vektorschätzung durch die Erfassung der Ecken des Bildes und nicht der Kanten bei der lokalen Betrachtung bestimmt ist, wird der minimale Eigenwert betrachtet, um ein Kriterium für den minimalen Informationsgehalt in einem Abtastfenster<sup>17</sup> zu erhalten. Es sollte also möglich sein, eine minimale Informationsgrenze zu bestimmen, die für eine vorgegebene Genauigkeit der Verschiebungsschätzung ausreicht, um damit ein minimales Abtastfenster festzulegen. Der Informationsgehalt eines Fensters ist in der **G**-Gradientenmatrix aus Gleichung (2-1-3) enthalten. Ihre Eigenwerte  $\lambda_1$  und  $\lambda_2$  geben Anhaltspunkte über die Gradienten des Abtastfensters, aus denen sich folgende Eigenschaften des Bildes ableiten lassen:

- Kleines  $\lambda_1$  und kleines  $\lambda_2$  deuten auf homogene Flächen ohne ausgeprägte Grauwertgradienten. Die **G**-Matrix ist in diesem Fall singulär. Dies ist ein verlässliches Indiz dafür, dass kein Flussvektor eingeschätzt werden kann.
- Kleines  $\lambda_1$  und großes  $\lambda_2$  (oder umgekehrt) deuten auf einen unidirektionalen Kantenverlauf hin.
- Großes  $\lambda_1$  und großes  $\lambda_2$  deuten auf einen ausgeprägten Grauwertgradienten (z.B. Ecken) hin.

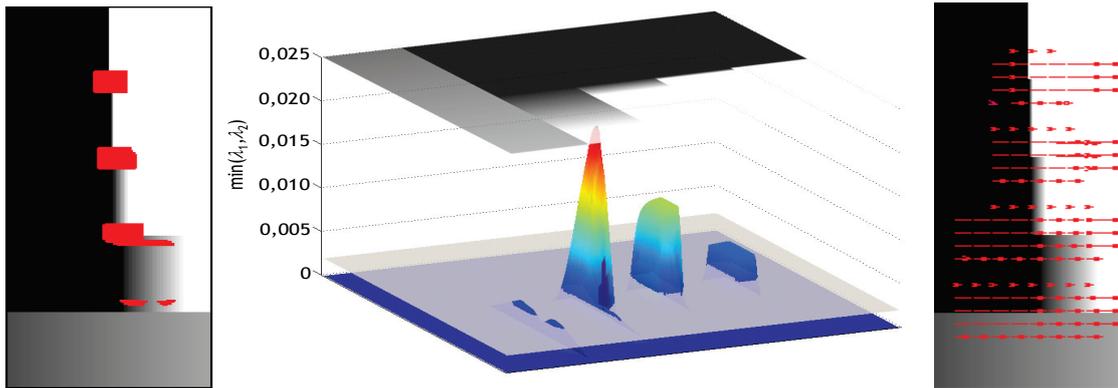
---

<sup>16</sup> z.B. Segmentierung von sich bewegenden Objekten oder Erfassung von lokalen detaillierten Bewegungen, wenn sich zum Beispiel zwei Objekte nah aneinander vorbei bewegen

<sup>17</sup> Durch ein Abtastfenster tritt auch der Effekt der Vektorfeldausbreitung auf, wenn ein Eigenwert einen minimalen Wert aufweist. Das sind die unidirektionalen Kanten, an denen keine Verschiebung berechnet werden kann. Somit zeigt sich, dass das Minimum der Eigenwerte als Merkmal für den minimalen Informationsgehalt in einem Abtastfenster geeignet ist.

## 2. INTENSITÄTSBASIERTE ANSÄTZE

Das Prinzip kann anhand eines Testbildes einfach verdeutlicht werden, das aus fünf untereinander angeordneten Grauwertübergängen besteht (Abb. 2-1-7), die von unten nach oben zunehmend steiler werden. Aus der Darstellung des minimalen Eigenwertes wird ersichtlich, dass die Information zur Berechnung von Intensitätsverschiebungen nicht entlang der Grauwertübergänge in waagerechter Richtung enthalten sind, sondern in den Übergangsbereichen zwischen den Stufen, in denen sowohl waagerechte als auch senkrechte Grauwertübergänge stattfinden. Dieses Problem ist in der Literatur als *Apertur-* oder *Blendenproblem* bekannt, dem aber mit der Vergrößerung des Abtastfensters entgegengewirkt werden kann.



**Abbildung 2-1-7:** Bestimmung der zu verfolgenden Punkte anhand eines Testbildes (links) nach Shi und Tomasi. Über das Minimum der Eigenwerte der Gradientenmatrix, die mit einem 8x8 Pixel großen Abtastfenster bestimmt wurden (mitte), werden über einem Schwellwert (graue Ebene), die gut zu verfolgenden Regionen, rot gekennzeichnet, bestimmt (links). Rechts ist das Kantenbild mit errechneten Verschiebungsvektoren.

Wie in der Abbildung 2-1-7 ersichtlich ist, können durch ein entsprechendes Fenster auch Bereiche in der Nähe starker Ecken, die selber wenig Information enthalten (im linken Bild rot markiert), verfolgt werden. In Anlehnung daran und an die Arbeit von Shi und Tomasi [Shi 94] wird ein Kriterium ( $E_\lambda$ ) definiert, um die optimale Größe des Abtastfensters für jeden Bildpunkt zu bestimmen. Diese erfolgt durch die Schrittweise Vergrößerung des Abtastfensters bis das Kriterium  $E_\lambda$  eine vordefinierte Schwelle erfüllt. Es gilt:

$$E_\lambda = \begin{cases} \frac{\lambda_S}{2} - \min(\lambda_1, \lambda_2) & , \text{ für } \frac{\lambda_S}{2} - \min(\lambda_1, \lambda_2) \geq 0 \\ 0 & , \text{ sonst} \end{cases} \quad (2-1-4b)$$

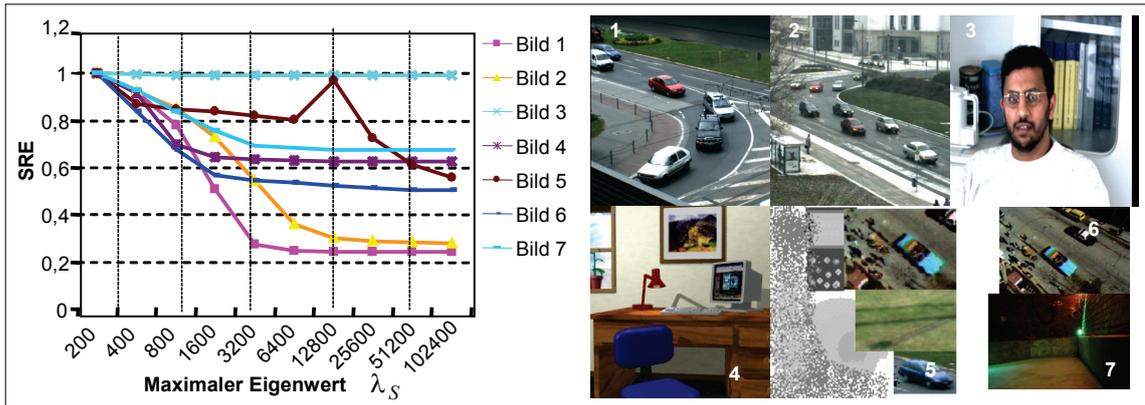
$$\Omega(x, y) = m \times m \text{ , für } \min_{m: m < m_{max}} [\lambda]$$

mit  $\lambda_1$  und  $\lambda_2$  als Eigenwerte der **G**-Matrix für ein Fenster der Größe  $m \times m$  und  $m_{max}$  der maximalen Fenstergröße. Der Grenzwert  $\lambda_S$  (der auch als Schwellwert zu betrachten ist) ist dem minimal benötigten Informationsgehalt in einem Abtastfenster zugeordnet, und wird als maximaler Eigenwert bezeichnet. Die Schwierigkeit besteht in der Bestimmung des Grenzwertes (z.B. minimaler Informationsgehalt, der maximale Eigenwert), der eine gute Verschiebungsschätzung erlaubt. Für die Applikation hier wurde der *maximale Eigenwert*  $\lambda_S$  anhand der Analyse von Testbildern eingeschätzt, in denen mehrere Testbilder um kleinere Abstände verschoben wurden. Durch die dabei gemessenen

## 2. INTENSITÄTSBASIERTE ANSÄTZE

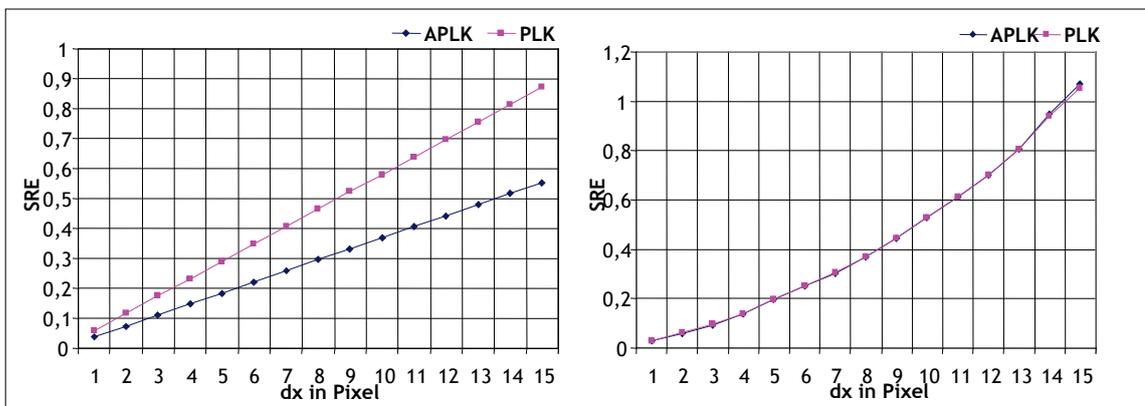
SRE-Fehler, die durch die Verschiebungsschätzung mit verschiedenen maximalen Eigenwerten entstanden sind, kann aus dem Diagramm der Parameter  $\lambda_S$  durch die Schwellenfestlegung geschätzt werden.

Bei einer bestimmten Fenstergröße zeichnet sich eine Konvergenz gegen das Fehlerminimum ab. Daraus lässt sich schließen, dass ab einem maximalen Eigenwert von dieser erfassten Größe (in unserem Fall 10000, Abb. 2-4-2) keine gravierenden Verbesserungen in der Verschiebungsschätzung auftauchen und das Abtastfenster bei dieser Größe belassen werden kann. Der Grenzwert drückt aus, welche Fenstergröße dem maximalen Wert entspricht. Das Prinzip kann für die Mehrzahl von Versuchsbildern (Abb. 2-1-8) gezeigt werden.



**Abbildung 2-1-8:** (links) SRE bei der Verschiebung der Testbilder 1 bis 7 (rechts) um drei Pixel mit verschiedenen maximalen Eigenwerten.

Besonders das synthetisch zusammengesetzte Bild 5 in Abbildung 2-1-8 weist Schwankungen auf, die zu Schwierigkeiten bei der Ermittlung des optischen Flusses führen können. Denn durch die regelmäßige Textur im linken Bereich des Bildes kommt es zu lokalen Minima in der Berechnung des Optischen Flusses durch das LK-Verfahren. Dadurch ist die Annahme der Grauwertstabilität nicht mehr zu gewährleisten. Für die restlichen Bilder ist jedoch eine angestrebte Konvergenz der SRE bei ca. 10000 zu erkennen.



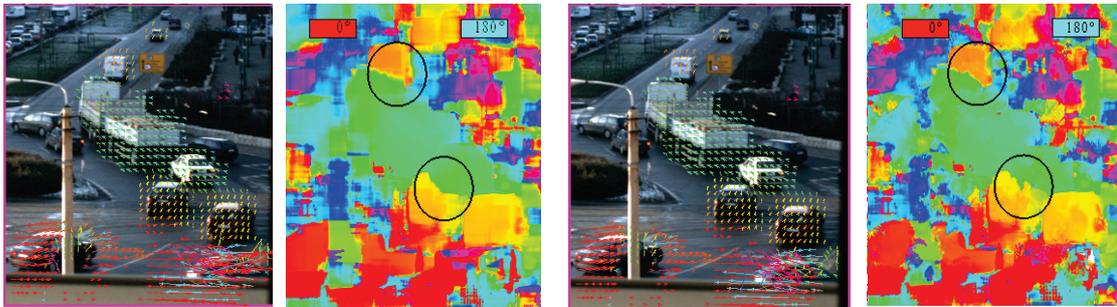
**Abbildung 2-1-9a:** SRE für das PLK-Verfahren mit einer Abtastfenstergröße von 10 und das adaptive PLK-Verfahren (APLK) bei der Verschiebung eines Quadrates mit Textur auf einem schwarzen Hintergrund (links) und einem Hintergrund mit Textur (rechts).

## 2. INTENSITÄTSBASIERTE ANSÄTZE

Vergleichend mit PLK-Verfahren ohne Adaption zeigt die Adaption vom PLK-Verfahren (APLK) eine Verringerung des Fehlers um die 30% bei dem texturarmen Hintergrund (Abb. 2-1-9a). Dies ist darauf zurückzuführen, dass hier ein größerer Dynamikgewinn aufgrund der Grauwertdifferenz zwischen Vordergrundobjekt und dem Hintergrund „Helligkeitsübergänge“ erreicht wird und dadurch die kleineren Fenster auch mehr Informationsgehalt beinhalten.

Bewegt sich das Objekt über einen Hintergrund mit Textur (wie in Szenario IV), so wird diese Differenz kleiner und somit werden die Fehlerkurven geradezu identisch sein (s. Abb. 2-1-9a). Die Hintergrundtextur enthält in diesem Beispiel wenig Ecken und schwächt die Ecken vom Übergang des bewegten Objektes zum Hintergrund ab. Dadurch werden die Abtastfenstergrößen um das Objekt wenig adaptiert und es kommt zu keiner signifikanten Verbesserung.

Die Verbesserungen in den realen Verkehrsszenen fallen nur geringfügig aus (s. Abb. 2-1-9b). Es zeigt sich, dass die Verschiebungsfelder des dichten Straßenverkehrs in den markierten Bereichen durch die Adaption besser aufgelöst werden. Jedoch ist der Nutzen für die nachfolgende Segmentierung zu gering, als das es den erhöhten Rechenaufwand rechtfertigt.



**Abbildung 2-1-9b:** In der Verkehrsszene ist die Berechnung des Verschiebungsfeldes mit (erstem Bild) dem PLK-Verfahren und dem adaptiven PLK-Verfahren (dritten Bild) dargestellt. Dazu sind die Winkel der Vektoren im Falschfarbendarstellung zur Verdeutlichung dargestellt.

### 2. Reduzierung des Vektorfeldausweitungseffektes

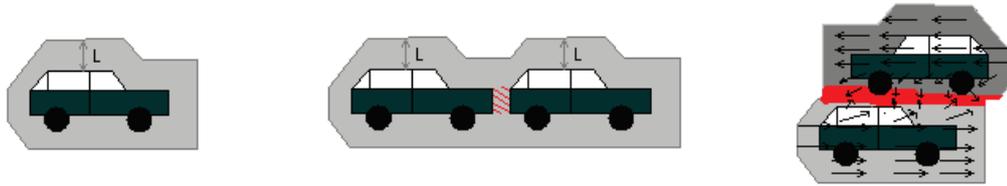
Der zweite Schritt der vorgeschlagenen Adaption befasst sich nicht mit dem gesamten Bildinhalt, sondern mit der durch eine Bewegungssegmentierung entstandenen Kontur der bewegten Regionen. Das Hauptproblem, das bei dem PLK-Verfahren auftritt, ist die zu grobe Abtastung, die zum Verschwinden von kleinen Bewegungsregionen aufgrund der Ausdehnung des Verschiebungsvektorfeldes über die sich bewegenden Objekte führt. Durch die Fensteradaption wird dieser Effekt teilweise minimiert, es sind jedoch weitere Verbesserungen notwendig. Der Vektorfeldfehler ist wie beschrieben abhängig von der Objektbewegung und den enthaltenen bidirektionalen Gradienten eines Abtastfensters. Dadurch kommt es zu drei für die Bewegungssegmentierung interessante Effekte:

- koronaförmige Ausbreitung des Vektorflussfeldes um das sich bewegende Objekt,
- Verschmelzung von benachbarten Vektorflussfeldern durch den Koronaeffekt und
- Verwischen von scharfen Verschiebungsfeldrändern, wenn sie sich überlappen.

Der Hauptfehler wird dabei von der Korona verursacht, die durch die Überlappungsbreite  $L$  charakterisiert ist (Abb. 2-1-10). Allerdings ist die Überlappungsbreite bzw. die Breite der Korona nicht überall konstant. Sie hängt stark von den im Hintergrund

## 2. INTENSITÄTSBASIERTE ANSÄTZE

enthaltenen Primitiven ab. Dabei erreicht die Überlappungsbreite ihr Maximum, wenn sich ein Objekt über einen homogenen Hintergrund bewegt. Enthält der Hintergrund keine bidirektionalen Gradienten, so entspricht die Überlappungsbreite  $L$  genau der Abtastfenstergröße  $(\Omega-1)$ . Um die Korona zu beseitigen, wurden verschiedene Strategien entwickelt, die auf der Berechnung zweier Vektorfelder durch den PLK-Algorithmus mit verschiedenen Fenstergrößen ( $\Omega_L$  und  $\Omega_S$ ) basieren.



**Abbildung 2-1-10:** Die printipielle Darstellung der auftretenden Korona-Effekte durch große Abtastfenster mit der Überlappungsbreite  $L$ .

Die Aufgabe, der anschließenden Strategien, besteht in der Vereinigung dieser beiden Vektorfelder, so dass sich der koronale Effekt verringert, die Verschiebungsschätzung sich jedoch nicht marginal verschlechtert. Dazu wird das Abtastfenster  $\Omega_L$  durch die Fensterbreite  $m_L$  definiert. Das mit ihm errechnete optische Vektorflussfeld  $\mathbf{h}_L(x,y)$  hat die Eigenschaft, dass es grob aufgelöst und eine robuste Schätzung der Objektbewegungen liefert (in den Versuchen ist  $m_L = 10$ ). Ein zweites Abtastfenster  $\Omega_S$  ist durch die Fensterbreite  $m_S$  gegeben. Obwohl das mit ihm errechnete optische Flussfeld  $\mathbf{h}_S(x,y)$  *feiner* aufgelöst ist, kann die Bewegungsinformation rauschempfindlicher und ungenauer sein. Der nächste Schritt beinhaltet die Abtrennung der sich bewegenden Bereiche vom Hintergrund. Dies kann durch eine Schwellwertfestlegung ( $H$ ) geschehen. Dabei werden alle Vektoren  $\mathbf{h}_L(x,y)$  gelöscht, die eine zu geringe Verschiebung aufweisen:

$$\forall \mathbf{h}_L(x,y), \|\mathbf{h}_L(x,y)\| < H \rightarrow \mathbf{h}_L(x,y) = (0 \ 0)^T \quad (2-1-5a)$$

Es entsteht eine Bewegungsmaske  $M$  mit den Punkten  $p_m(x,y)$ :

$$\forall p(x,y), \exists \mathbf{h}_L(x,y) \wedge \|\mathbf{h}_L(x,y)\| > 0 \rightarrow p_m(x,y) = 255 \quad (2-1-5b)$$

(füge Punkt  $p_m(x,y)$  der Bewegungsmaske  $M$  hinzu)

Die Korona  $K$  mit den Punkten  $p_k(x,y)$  ergibt sich dann aus der verbreiterten Konturlinie mit den Konturpunkten  $p_c(n)$  der Bewegungsmaske:

$$\forall p(x,y), \exists p_m(x,y) \wedge \|p_m(x,y) - p_c(n)\| \leq m_L \rightarrow p_k(x,y) = 255 \quad (2-1-5c)$$

(füge Punkt  $p_k(x,y)$  der Koronamaske  $K$  hinzu)

Es scheint auf dem ersten Blick ratsam, die Bewegungsvektoren der Korona *auf* Null zu setzen. Jedoch ist zu beachten, dass die Dicke der Korona entsprechend der Texturinformation des Hintergrundes variiert. Objektteile, die sich in der Korona befinden, würden dadurch entfernt. Daher wird die Strategie verfolgt, die in der Bewegungsmaske enthaltenen Verschiebungen mit einem kleinen Abtastfenster  $\Omega_S$  (in den Versuchen  $m_S = 5$ ) mit dem PLK-Verfahren neu zu bestimmen. Es wird dazu das Vektorfeld  $\mathbf{h}_S$  verwendet. Dieses Vektorfeld enthält die feiner aufgelöste Bewegungsinformation.

## 2. INTENSITÄTSBASIERTE ANSÄTZE

Durch die feinere Auflösung von Bewegungsstrukturen können die Stellen detektiert werden, an denen fälschlicherweise eine Verschiebung errechnet wurde. In den folgenden *Versuchen* (Strategien) werden Möglichkeiten vorgestellt, die beiden Bewegungsfelder  $\mathbf{h}_L$  und  $\mathbf{h}_S$  zu vereinen.

V1. Das Bewegungsfeld in der Korona wird durch die neu berechneten Vektoren  $\mathbf{h}_S$  bestimmt. Auf diese Weise kann das Bewegungsfeld dort feiner aufgelöst werden.

Es gilt:

$$\mathbf{h}(x,y) = \begin{cases} \mathbf{h}_S(x,y) , & \text{wenn } p(x,y) \in K \\ \mathbf{h}_L(x,y) , & \text{sonst} \end{cases} \quad (2-1-6a)$$

Es stellt sich aber heraus, dass gehäuft sehr viele Ausreißer in der Korona enthalten sind (s. Abb. 2-1-12a) und die Bewegungsschätzung am Rand besonders bei homogenen Flächen sehr ungenau werden kann.

V2. Um die Ausreißer in der Korona zu beseitigen, wird eine weitere Randbedingung hinzugefügt. Demnach werden alle Vektoren, deren relativer SRE-Vektorfehler zwischen den  $\mathbf{h}_S$  und  $\mathbf{h}_L$  Vektoren gemessen wird, der größer als eins ist, auf Null gesetzt. Dies beruht auf der Annahme, dass jene Vektoren, deren SRE größer als eins ist, Ausreißer sind und im Randgebiet falsch geschätzt wurden und somit zum Hintergrund gehören (s. Abb. 2-1-11a). Somit können durch ein zu kleines Abtastfenster entstandene Ausreißer entdeckt und eliminiert werden.

$$\mathbf{h}(x,y) = \begin{cases} \mathbf{h}_S(x,y) , & \text{wenn } p(x,y) \in K \wedge SRE(\mathbf{h}_S(x,y), \mathbf{h}_L(x,y)) \leq 1 \\ (0,0)^T , & \text{wenn } p(x,y) \in K \wedge SRE(\mathbf{h}_S(x,y), \mathbf{h}_L(x,y)) > 1 \\ \mathbf{h}_L(x,y) , & \text{sonst} \end{cases} \quad (2-1-6b)$$

V3. Um nun die fehlerhaften Schätzungen des feinen Vektorfeldes zu glätten, wird das Bewegungsfeld der Korona durch eine gewichtete Summe zwischen den Feldern  $\mathbf{h}_S$  und  $\mathbf{h}_L$  interpoliert. Dabei wird der SRE-Fehler als Wichtungsfaktor  $\gamma$  genommen. In diesem Schritt findet nun eine Interpolation der beiden Vektorfelder gemäß Gleichung (2-1-6c) statt und Fehler, die durch Berechnung von  $\mathbf{h}_S$  entstanden sind, können beseitigt werden.

$$\mathbf{h}(x,y) = \begin{cases} \mathbf{h}_S(x,y) \cdot \gamma(x,y) + \mathbf{h}_L(x,y) \cdot (1 - \gamma(x,y)) , & \text{wenn } p(x,y) \in K \wedge \gamma(x,y) \leq 1 \\ (0,0)^T , & \text{wenn } p(x,y) \in K \wedge \gamma(x,y) > 1 \\ \mathbf{h}_L(x,y) , & \text{sonst} \end{cases} \quad (2-1-6c)$$

mit  $\gamma(x,y) = SRE(\mathbf{h}_S(x,y), \mathbf{h}_L(x,y))$

V4. Mit den bisher gewonnenen Verfahren wird das Problem der Vektorfeldausbreitung beseitigt (A und B Terme in Gleichung 2-1-6d). Um die Verschmelzung und Vermischung der Vektorfelder zu verbessern, müssen auch die Bereiche innerhalb der von der Korona eingeschlossenen Bewegungsmaske durch das fein aufgelöste Vektorfeld  $\mathbf{h}_S$  neu berechnet werden.

$$\mathbf{h}(x,y) = \begin{cases} \mathbf{h}_S(x,y) \cdot \gamma(x,y) + \mathbf{h}_L(x,y) \cdot (1 - \gamma(x,y)) , & \text{wenn } p(x,y) \in K \wedge \gamma(x,y) \leq 1 & (A) \\ (0,0)^T , & \text{wenn } p(x,y) \in K \wedge \gamma(x,y) > 1 & (B) \\ \mathbf{h}_S(x,y) , & \text{wenn } p(x,y) \notin K \wedge \lambda(x,y) \geq \lambda_s & (C) \\ \mathbf{h}_S(x,y) \cdot \beta(x,y) + \mathbf{h}_L(x,y) \cdot (1 - \beta(x,y)) , & \text{sonst} & (D) \end{cases}$$

$$\text{mit } \gamma(x,y) = SRE(\mathbf{h}_S(x,y), \mathbf{h}_L(x,y)) \text{ und } \beta(x,y) = \frac{\lambda(x,y)}{\lambda_s} \quad (2-1-6d)$$

Die Zuverlässigkeit eines Verschiebungsvektors aus dem Vektorfeld  $\mathbf{h}_S$  wird am einfachsten durch den aus Abschnitt 2.1.5 besprochenen minimalen Eigenwert  $\lambda$  der Gradientenmatrix an der Stelle  $(x,y)$  des zugehörigen Vektors bestimmt. Übersteigt dieser den Grenzwert  $\lambda_s$ , so kann die Bewegungsschätzung mit dem Vektor  $\mathbf{h}_S(x,y)$  als wahrscheinlich angenommen werden (C). Je kleiner der minimale Eigenwert ist, desto geringer ist auch der Einfluss des Vektors  $\mathbf{h}_S(x,y)$  gegenüber dem Vektor  $\mathbf{h}_L(x,y)$  auf den resultierenden Vektor  $\mathbf{h}(x,y)$  (Teil D). Mit dieser Modifikation können auch falsche Bewegungen innerhalb der Bewegungsmaske bei Sequenzen mit eckenreichen Regionen entfernt werden. Diese Analyse liefert eine Zuverlässigkeitsbewertung, die dazu eingesetzt werden kann, unzuverlässige Verschiebungsvektoren zurückzuweisen. Allerdings berücksichtigen die bisher beschriebenen Schritte nicht den Ghosting-Effekt, der nicht nur bei der Differenzbildtechnik auftritt, sondern auch bei der Berechnung des optischen Flusses durch die gradientenbasierten Verfahren.

### 3. Differenzbildtechnik und Anti-Ghost-Adaption

Es gelingt mit einer Bewegungselektion durch ein auf Differenzbildtechnik basierendes Verfahren, die Vektorfeldausweitung zu eliminieren. Allerdings erscheinen im Differenzbild  $D_k$  auch Regionen, in denen der Hintergrund durch die Bewegung eines Objektes wieder freigegeben wird. Dieser Effekt wird als *Ghosting-Effekt* bezeichnet. Um diesen zu beseitigen, wurde die Differenzbildtechnik modifiziert und erweitert (s. Kapitel 3). Die Modifikationen beziehen sich auf doppelte Differenzbildtechnik in Verbindung mit einem Nachverarbeitungsalgorithmus (s. Kapitel 3), die zur genauen Segmentierung der bewegten Regionen [Al-Hamadi 03a] führt. Problematisch sind weiterhin texturlose Regionen, wie sie oft an Fahrzeugen durch Überbelichtung auftreten können. Aus diesem Grund werden in einem weiteren Vorverarbeitungsschritt, Löcher in der Bewegungsmaske mit dem morphologischen Operator „Closing“ geschlossen.

Ein zweiter Verarbeitungsschritt die *Kontur-Erosion*<sup>18</sup>, die in [Al-Hamadi 04c, 05b] eingeführt wurde. Durch diesen zweiten Schritt wird die Kontur der Objekte geglättet und Regionen, die durch einen dünnen Pixelstreifen mit einander verbunden sind, voneinander getrennt. Durch weitere auf dem Differenzbild  $D$  angewandte Verarbeitungsschritte erhält man eine Bewegungsmaske MD. Diese lässt sich auf das Vektorfeld  $\mathbf{h}$  anwenden. So lässt sich die doppelte Differenzbildtechnik (MD) auch für das PLK-Verfahren anwenden. Zunächst werden zwei Bewegungsmasken berechnet. Das Differenzbild  $D^1$  wird nach Gleichung (2-1-6e) bestimmt, welches sich wiederum aus den Eingangsbildern  $I_{k-1}$  und  $I_k$  ergibt. Die Bewegungsmaske  $D_2$  wird aus den Bildern  $I_k$  und  $I_{k+1}$  berechnet und mittels einer festgelegten Schwellenwertes  $H$  binärisiert. Es ergibt sich eine Bewegungsmaske MD aus  $D_1$  und  $D_2$  somit das resultierende Vektorfeld  $\mathbf{h}$  nach der Anti-Ghost-Adaption entsprechenden Gleichung (2-1-6g).

$$D_1(x,y) = \begin{cases} 1 & , \text{wenn } |I_{k-1}(x,y) - I_k(x,y)| > H \\ 0 & \text{sonst} \end{cases} \quad (2-1-6e)$$

---

<sup>18</sup> Dieser Operator arbeitet auf der Kontur einer Region und entfernt alle Pixel der Bewegungsmaske, die auf einer Sekante dieser Kontur liegen und deren Strecke, kleiner als eine vordefinierte Schwelle ist.

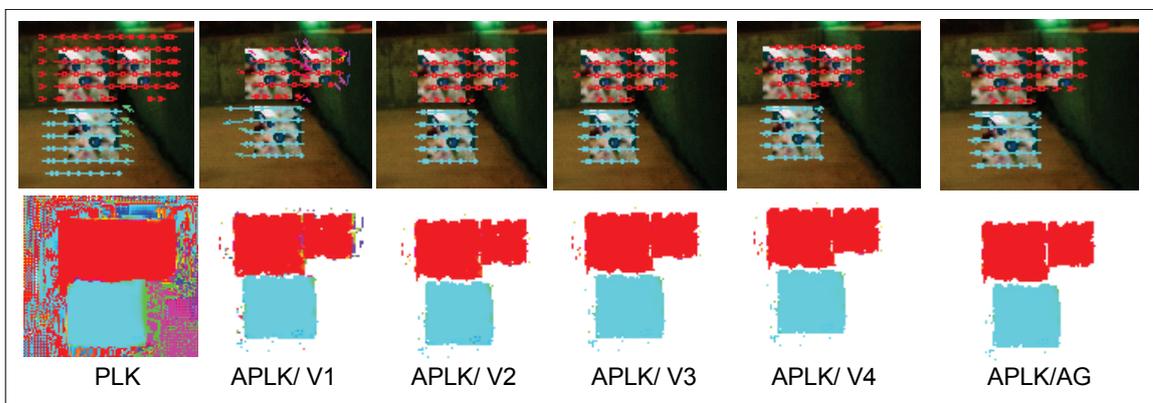
## 2. INTENSITÄTSBASIERTE ANSÄTZE

$$D_2(x,y) = \begin{cases} 1 & , \text{wenn } |I_k(x,y) - I_{k+1}(x,y)| > H \\ 0 & \text{sonst} \end{cases} \quad \text{und MD} = D_1 \wedge D_2 \quad (2-1-6f)$$

$$\mathbf{h}(x,y) = \begin{cases} \mathbf{h}_L(x,y) & , \text{wenn } p(x,y) \in \text{MD} \\ (0,0)^T & , \text{sonst} \end{cases} \quad (2-1-6g)$$

### 4. Bewertung und Vergleich

Als Testszene wurde ein Hintergrund mit Textur gewählt, auf denen sich 3 Quadrate bewegen, um auftretende Fälle aus Abbildung 2-1-11a zu behandeln. Die oberen zwei Quadrate bewegen sich nach rechts, während sich das untere Quadrat nach links verschiebt. Im ersten Bild, dem normalen PLK-Algorithmus, sind deutlich die (roten) sich überlappenden Vektorfelder, so wie die sich beeinflussenden Felder rot und blau zu erkennen.



**Abbildung 2-1-11a:** Versuchsreihen, von oben nach unten, PLK, V1, V2, V3, V4, Anti Ghost, links Winkel (Farbe) und normierte Vektorlänge (Helligkeit), rechts Vektorfeld.

Durch die Neuabastung mit einem kleinen Fenster (V1) gelingt es die Überlappung der Vektorfelder zu minimieren. Jedoch wird auch deutlich, dass sich die Bewegungsschätzung im zweiten Bild verschlechtert. Die Versuche V2 und V3 sind sich im Ergebnis sehr ähnlich und man erkennt die verbesserten Randregionen, sowie eine gute Verschiebungsberechnung im inneren der Objekte. Es zeigt sich, dass der Versuch V4 keine weiteren Verbesserungen hervorruft. Das beste sichtbare Ergebnis liefert die Anti-Ghost Adaption (s. Abb. APLK/AG), es gelingt ihr auch die an den Rändern bei Versuch V2 und V3 punktförmigen Vektoransammlungen, die die Verbesserungen überlebt haben, zu beseitigen.

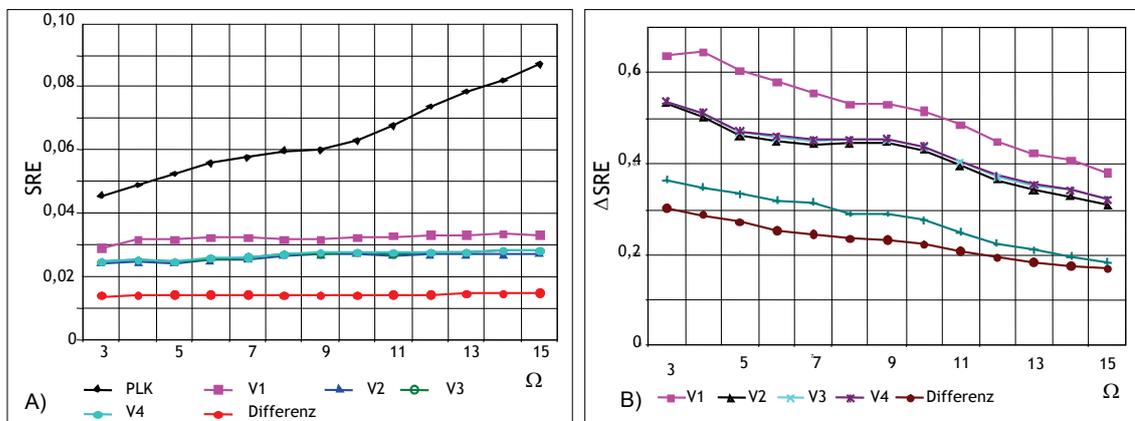
In der Abbildung 2-1-12 ist das Fehlverhalten der Adaptionen bei einer realistischen Sequenz zu sehen. Es zeigt sich das der SRE-Fehler des PLK-Verfahrens nur noch annähernd linear von der Abtastfenstergröße abhängt. Der Versuch V1 liefert die geringste Verbesserung, da durch die reine Neuberechnung der Korona mit sehr kleinem Abtastfenster das Vektorfeld fehlerhaft geschätzt wird. Es zeigt sich das eine Interpolation mit dem grob aufgelösten Vektorfeld notwendig ist. Die Anti-Ghost-Adaption zeigt mit größer werdendem Abtastfenster eine besser werdende Verschiebungsfeldschätzung. Versuch 4 hebt sich im Gegensatz zu den idealen Szenen nicht von den Versuchen 1 bis 3 ab, was durch eine schwache Texturierung des Hintergrunds zu erklären ist. Der Unterschied zwischen den Versuchen V2 bis V4 erweist sich bei realen Szenen als minimal und wird nur in Bereichen mit einer großen Dichte von Ecken und

## 2. INTENSITÄTSBASIERTE ANSÄTZE

Primitiven deutlich werden. Die Anti-Ghost-Adaption liefert sehr gute Ergebnisse auch im Bezug auf die Beseitigung der Vektorfeldverschmelzungen und erweist sich im Hinblick auf die Geschwindigkeit der Fehlerminimierung als bestes Verfahren.

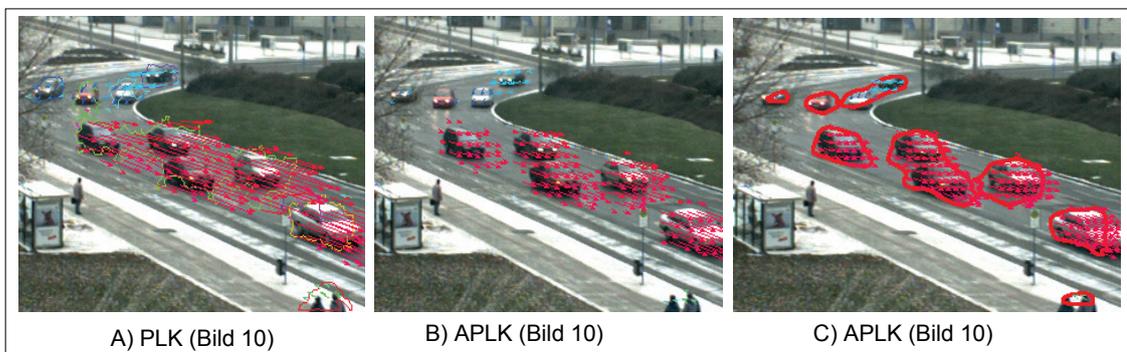
**Tabelle 2-1-1:** SRE-Fehler der Versuchreihen der realistischen Sequenz im Verhältnis zum PLK-Verfahren

$\Omega$	3	5	7	9	11	13	15
V1	36,29%	39,56%	44,39%	47,03%	51,50%	57,71%	61,87%
V2	46,77%	53,77%	55,65%	55,37%	60,47%	65,81%	68,95%
V3	46,34%	53,00%	54,80%	54,76%	59,69%	64,84%	67,74%
V4	46,33%	52,87%	54,78%	54,57%	59,52%	64,60%	67,78%
Differenz.	69,72%	72,89%	75,56%	76,68%	79,07%	81,65%	83,07%



**Abbildung 2-1-12:** SRE-Fehler der Versuchreihen bei dem Parameterdurchlauf mit der realistischen Sequenz (links) absoluter Fehlerverlauf, (rechts) Fehler der Adaptionen im Verhältnis zum SRE des PLK-Verfahrens.

In Abbildung 2-1-13 wurden die Vektorfelder des normalen PLK-Verfahrens für reale Bildsequenzen dargestellt. Es stellt sich heraus, dass ohne die Verwendung der Differenzbildtechnik keine Verbesserung bezüglich der Segmentierung von sich bewegenden Objekten erreicht werden kann.



**Abbildung 2-1-13:** Das berechnete Verschiebungsvektorfeld mit der daraus resultierenden Segmentierung mittels PLK-Verfahren ohne Adaption in A). In B) und C) ist das Verschiebungsvektorfeld mit Adaption durch den MD-Ansatz, das zu einer genauen Segmentierung von sich bewegenden Regionen führt (s. C).

Die Ergebnisse weichen sehr stark von der tatsächlichen Objektbegrenzung ab. Unter Verwendung der Differenzbildtechnik in B und C erkennt man die signifikante Verbesserung bei der Eliminierung der Vektorfeldausweitung. Die Ursachen dafür sind darauf zurückzuführen, dass durch den MD-Ansatz ein Ghosting-Effekt eliminiert wird und die Berechnung nur für die genaue Bewegungsmaske erfolgt. Diese wird in Kapitel 5 ausgenutzt, um eine genaue Beschreibung von Regionen zu ermöglichen. Durch die Verwendung des Anti-Ghost-Verfahrens konnte eine kleine Verbesserung erlangt werden, die sich aber in realen Verkehrsszenen durchaus bemerkbar machen.

### 2.1.6 Kurzfassung und Diskussion

Im Rahmen der bisher dargestellten Untersuchungen wurden drei gradientenbasierte Verfahren für die Berechnung des optischen Flusses aus der Literatur vorgestellt und untersucht. Dabei konnten die grundlegenden Eigenschaften und Charakteristika aufgezeigt werden. Es stellte sich heraus, dass das HS- und das LK-Verfahren Probleme mit großen Pixelverschiebungen in realen Bildsequenzen aufweisen. Es wurde das PLK-Verfahren als Lösung für dieses Problem vorgestellt. Der koronale Effekt, der sich dabei als Problemstellung gezeigt hat, wurde mit verschiedenen vorgeschlagenen Adaptionen des PLK-Verfahrens verringert. Dabei zeigte sich, dass eine prinzipielle Adaption des Abtastfensters schwierig zu bewerkstelligen ist und es wurde eine konturbezogene Adaption mittels des MD-Ansatzes [Al-Hamadi 05] bevorzugt, der mit einem niedrigen Rechenaufwand verbunden ist.

Von weiterem Interesse ist die Untersuchung des von Bruhn und Weickert vorgestellten Algorithmus [Bruhn 02], oder das von Brox und Bruhn in [Brox 04] vorgestellte Verfahren, welches das HS-Verfahren mit dem LK-Verfahren vereint. Die Auswirkung der Kombination beider Verfahren auf die Vektorausweitung sollte sich jedoch in Grenzen halten. Ebenfalls von Interesse wäre das Benutzen eines gewichteten Fensters, z.B. eines Gaußfensters, anstelle des Rechteckfensters für die Berechnung eines Vektors beim LK-Verfahren. Beide Möglichkeiten konnten jedoch auf Grund des begrenzten Zeitraumes nicht implementiert werden. Es zeigt sich, dass der LK Algorithmus in seiner pyramidenartigen Implementierung ein robustes und akkurate Verfahren für die Berechnung des optischen Flusses gerade bei großen Verschiebungen ist. Dieser Algorithmus stellt eine Grundlage für die im Kapitel 5 beschriebene Merkmalsextrahierung dar.

### 2.2 Korrelationsverfahren

In diesem Abschnitt soll eine adaptive Systemstruktur auf der Grundlage eines Korrelationsverfahrens als Alternative<sup>19</sup> entwickelt werden, die die Bestimmung der Bewegungsparameter eines Objektes unter dem Einfluss von gestörten Bildsituationen gewährleistet. Das Prinzip der Korrelation beruht auf einen Vergleich von Mustern, gewöhnlicherweise Blöcke. Die Verknüpfung dieser Blöcke miteinander erfolgt durch ein Ähnlichkeitsmaß. Dieses liefert einen Extremwert je nachdem, welches Kriterium man verwendet.

Als Ähnlichkeitsmaße werden Korrelationsfunktionen im allgemeinen Sinne, wie bspw. die MAD-Funktion (Mittlere Absolute Differenz), die MSE-Funktion (Mean Square Error) oder die normierte Kreuzkorrelationsfunktion (nKKF), verwendet [Aschwanden 93]. Durch die Lokalisierung des Extremums lässt sich dann die genaue Position des gesuchten Merkmals (Referenzbereich) im zweiten Bildausschnitt (Suchbereich) angeben [Musmann 85, Al-Hamadi 00a, 01a, 02a]. In der Regel erfolgt zur Aufwandreduzierung eine Einteilung der Bilder in gleichgroße, rechteckige Blöcke der Dimension ( $N \times M$ ) mit dem Ziel, für alle Blockelemente einen Verschiebungsvektor zu ermitteln [Al-Hamadi 01a]. Hier spricht man von einer Vollsuchstrategie<sup>20</sup>.

Allerdings sollte die Bestimmung der Bewegung eines Objektes oder Blockes weitgehend störungsunabhängig sein. Das hängt aber nicht nur von der Größe und Wahl des Referenz- und Suchbereiches, sondern wesentlich von dem angewendeten Ähnlichkeitsmaß ab, mit dem die Verknüpfung zwischen Blöcken hergestellt wird. Werden zum Beispiel störanfällige Grauwerte anstatt anderer weniger problematischer Merkmale des Bildes zum Matchen benutzt, sollte das Verfahren möglichst rauschunabhängig sein. Entscheidet man sich dagegen für ein geeignetes Merkmal, werden eventuell schon bei dessen Detektion in der Vorverarbeitung manche Störungen "eliminiert". Segmentiert man z.B. die Kanten in einem Bild, ist es hinterher für das Matching unwesentlich, ob das Bild stark verrauscht war, solange die Kanten und Primitiven selbst noch gut zu erkennen waren.

Ein größeres Problem dieses Matchingsverfahrens ist der vergleichsweise höhere Rechenaufwand, da die Berechnung von Verschiebungsvektoren durch ein flächenbasiertes Ähnlichkeitsmaß, oft verbunden mit voller Suche, erfolgt. Hinzu kommen noch Probleme unter dem Einfluss von nichtkooperativen Messsituationen [Al-Hamadi 00b, 02b]. Diese Situationen führen oft zu einer falschen Lokalisierung des Extremwertes des Ähnlichkeitsmaßes und dadurch zu einem falschen Verschiebungsvektor, der die tatsächliche Bewegung nicht mehr beschreibt. Zur Verbesserung der Bewegungsschätzung wird anschließend eine Systemstruktur auf Grundlage des Matchings vorgeschlagen, bei der die Bewegungsschätzung von Objekten nicht wie bisher überwiegend in der Schwarz-Weißdomäne, sondern in der Farbdomäne durchgeführt wird [Al-Hamadi 01a, 02b, 03g, 03f].

---

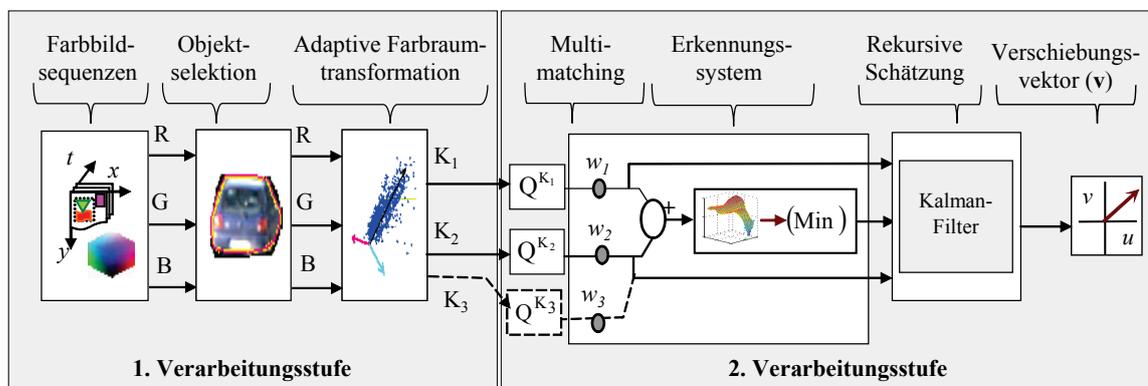
<sup>19</sup> Die folgenden Ausführungen beziehen sich generell auf flächenhafte Ausschnitte, die miteinander verglichen werden - im Gegensatz zu verschiedenen anderen Methoden, die zunächst Merkmale aus den Bildern extrahieren und diese dann einander zuordnen.

<sup>20</sup> Man findet verschiedene Suchstrategien, wie bspw. Die *2D-logarithmische Suche* [Jain 81], die *Drei-Schritt-Suche* [Koga 81] und die *Konjugierte Richtungssuche* [Srinivasan 84]

### 2.2.1 Vorgeschlagene Systemstruktur zur Bewegungsschätzung

Ausgehend von den Ausführungen in den vorigen Abschnitten soll nun ein vorgeschlagenes System beschrieben und insbesondere im nachfolgenden Abschnitt die Eignung zur Bewegungsanalyse anhand realer Bildsequenzen nachgewiesen werden. Besonders wird hierbei auf die Zuverlässigkeit und Robustheit des Systems beim Auftreten bildspezifischer Störungen eingegangen. Die Verwendung dieses Systems zur Bestimmung von Verschiebungsvektoren in Bildbereichen, in denen bisher keine eindeutige Messung (nichtkooperative Messsituationen) möglich ist, unterstützt die Aussagen über die Effizienz dieses Systems [Al-Hamadi 03f, 03g].

Die vorgeschlagene Systemstruktur<sup>21</sup> zur Bewegungsschätzung wird durch zwei Verarbeitungsstufen beschrieben (Abbildung 2-2-1), wobei die erste sich mit der *Aufbereitung der Farbinformation* befasst. Diese enthält die Farbbilddaufnahme mit einer Farbkamera, die Objektselektion mittels Verschiebungsvektorfeldern und die adaptive Farbraumtransformation für die selektierten Objekte. In der zweiten Verarbeitungsstufe des Systems wird *die Bewegungsanalyse für die selektierten Objekte* aus der ersten Verarbeitungsstufe durchgeführt. Das intensitätsbasierte Multi-Matching, die Auswertung des kombinierten Gesamtkriteriums mit einem neuronalen bildspezifischen Erkennungssystem zur Bestimmung der Verschiebungsvektoren bzw. der Bewegungstrajektorien bilden die zweite Verarbeitungsstufe des Systems. Dazu kommt auch ein rekursives Schätzverfahren zur Filterierung von gestörten Größen. Die Verknüpfung der Bildregionen miteinander erfolgt beim Multi-Matching durch ein  $Q^c$ -Ähnlichkeitsmaß (bspw. MAD-Funktion). Dieses liefert bei bestem Match einen Extremwert (Minimum oder Maximum) je nachdem, welches Kriterium man verwendet hat.



**Abbildung 2-2-1:** Die vereinfachte adaptive Systemstruktur zur Bewegungsanalyse durch die Einbeziehung der Farbinformation [Al-Hamadi 03f].

#### 2.2.1.1 Aufbereitung der Farbinformation

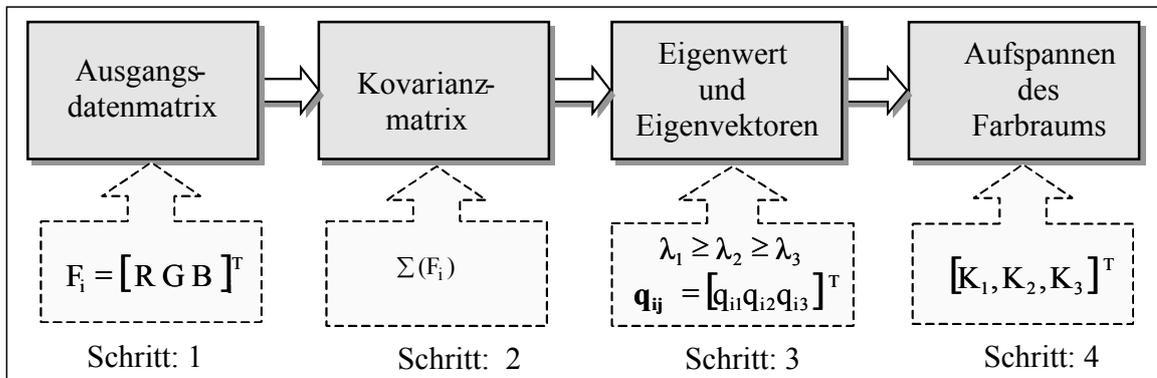
Für die automatische Objektselektion wurden Zahlreiche Verfahren entwickelt, die auf einem Vollsuche-Blockmatching [Al-Hamadi 00a], einen farbklassenbasierten Blockmatching [Al-Hamadi 00b, 01a] oder auf Differenzbildtechnik in Verbindung mit Nachverarbeitungsprozess gemäß [Al-Hamadi 03a] basieren. Beim *Vollsuche-Blockmatching* wendet man das Blockmatching-Verfahren zur Bestimmung des Verschiebungsvektorfeldes an. Im Ergebnis dieses ersten Matchings erhält man für jeden Referenzblock einen Verschiebungsvektor, der die ermittelte Bewegung des

<sup>21</sup> Die vorgeschlagene Systemstruktur kommt ohne Voraussetzungen bzw. Wissen über die äußeren Bedingungen (z.B. Szene, Lichtquelle, Kameras) aus.

## 2. INTENSITÄTSBASIERTE ANSÄTZE

repräsentierten Bild- bzw. Objektausschnittes beschreibt. Die resultierenden Bewegungsvektorfelder der einzelnen Blöcke des Gesamtbildes können ausgewertet werden, indem Blöcke, die die gleichen Verschiebungsvektoren aufweisen, zu einem Objekt zusammengefasst werden, das den Ausgangspunkt für das Tracking darstellt [Al-Hamadi 01a]. Da aber die Vollsuche-Blochmatching zur Objektselektion rechenaufwendig und gegenüber nichtkooperativen Messsituationen empfindlich sind, werden sie nur zur automatisierten initialen Objektselektion angewendet.

Die resultierenden Objektkandidaten aus der Objektselektion bilden nun den Ausgangspunkt zur adaptiven Farbraumtransformation. Dafür wird ein adaptiver Farbraum (Abbildung 2-2-1) vorgeschlagen, der jeweils an den aktuellen Bildinhalt angepasst ist [Al-Hamadi 00a,02b]. Dieser Ansatz resultiert aus der Tatsache, dass für Bildbereiche, die ausgeprägte Intensitätsgradienten enthalten, eine genaue und zuverlässige Bewegungsbestimmung realisiert werden kann. Somit stellt die adaptive Farbraumtransformation ein wichtiges Grundelement der verbesserten Bewegungsanalyse unter dem Einfluss von nichtkooperativen Messsituationen dar. Diese ist motiviert durch die Tatsache, dass die Erregungen der Rezeptoren im menschlichen Sehsystem aufgrund der großen Überlappung der spektralen Empfindlichkeitskurven der unterschiedlichen Rezeptoren hoch korreliert sind. Diese Korrelation hängt im Detail von der spektralen Verteilung des Lichtes in der betrachteten Szene ab. Somit kann prinzipiell hier eine Transformationsmatrix berechnet werden, deren Anwendung dazu führt, dass die Erregungen der Farbkanäle weitgehend unkorreliert sind, d.h. die höchstmögliche Informationsmenge wird transportiert. Diese Überlegung entspricht genau dem eingeführten adaptiven Farbraum in dieser Systemstruktur [Al-Hamadi 01a, 02c] (s. Abb. 2-2-2a).



**Abbildung 2-2-2a:** Die Rechenschritte zur Bestimmung des adaptiven Farbraumes

Zur mathematischen Bestimmung der *Transformationsmatrix*  $[q_{ij}]$  zum adaptiven  $K_1K_2K_3$ -Farbraum eignet sich die Karhunen-Loeve-Transformation [Al-Hamadi 00a,00b,01a], bei der die Komponenten des adaptiven Farbraumes im Sinne eines möglichst großen Kontrastes in Richtung größter Varianzen ausgerichtet werden. Eingangsdaten sind die Farbwerte  $F_i = [R \ G \ B]^T$  der selektierten Bildregionen. Dies geschieht durch drei Schritte:

- 1). Im ersten Schritt enthält die *Ausgangsdatenmatrix* die Daten für die weitere Verarbeitung, die die Farbinformationen im RGB-Format enthält, wobei die Kanäle Rot, Grün und Blau, die Eingangsmerkmalsvektoren beschreiben. Das Ziel ist die Bestimmung von einer Hauptkomponente und den zwei Nebenkomponten, die

## 2. INTENSITÄTSBASIERTE ANSÄTZE

ausreichend sind, um die Beschreibung der Gesamtvarianz der erfassten Eingangsdaten im dreidimensionalen Raum zu ermöglichen.

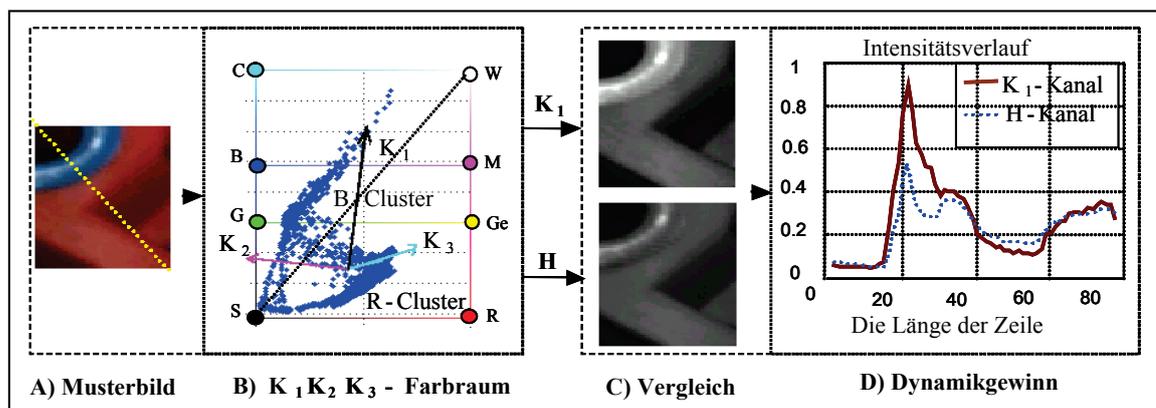
- 2). In dem zweiten Schritt wird der statistische Zusammenhang zwischen den angegebenen Variablen beschrieben. Dies geschieht in der Form der Kovarianzmatrix, die die statistische Varianz in den Hauptdiagonalen enthält. Die Dimension der Matrix entspricht der Anzahl der Merkmale in der Ausgangsdatenmatrix.

$$\Sigma = E\{(\mathbf{F}_i - E\{\mathbf{F}_i\}) \cdot (\mathbf{F}_i - E\{\mathbf{F}_i\})^T\} \quad (2-2-1a)$$

- 3). Nachdem die Kovarianzmatrix aus der Datenmatrix gemäß Gleichung (2-2-1a) berechnet wird, sind anschließend die Eigenwerte  $\lambda_i$  dieser Kovarianzmatrix  $\Sigma$  zu ermitteln. Die Eigenwerte  $\lambda_1 \geq \lambda_2 \geq \lambda_3$  entsprechend die zugehörigen Eigenvektoren und werden ihrer Größe nach geordnet, wobei schließlich solche Eigenvektoren mit kleinen Eigenwerten nur noch wenig Information beinhalten. Diese Eigenvektoren bilden nun die Transformationsmatrix, die hier als Wichtungsmatrix bei der Transformation der RGB-Kanäle auf neue  $K_1K_2K_3$ -Komponenten dient. Im mathematisch-statistischen Sinne stellen diese Gewichte nichts anderes dar, als ein Maß für den Zusammenhang zwischen Komponenten und Variablen.

$$[K_1 \ K_2 \ K_3]^T = [q_{ij}] \cdot [R \ G \ B]^T \quad (2-2-1b)$$

Der adaptive Farbraum in Gleichung (2-2-1b) wird nun durch die drei Komponenten  $K_1$ ,  $K_2$  und  $K_3$  aufgespannt. Diese Komponenten sind zueinander orthonormal. Dadurch ergibt sich, dass die erste Komponente die Projektionsgerade beschreibt. Da sie den größten Eigenwert aufweist, sind die Gewichte in der Projektionskomponente generell positiv. Dies ist auf die Informationsstreuung, die diese Komponente enthält, zurückzuführen.



**Abbildung 2-2-2b:** Dargestellt ist hier die Verteilung der Farbdaten eines Musterbildes a), das einen Farbübergang von Blau (B-Cluster) nach Rot (R-Cluster) enthält. Im Bild ist eine Ansicht im 3-dimensionalen Raum dargestellt, die die einfache Mittelwertbildung und den Projektionsvektor zeigt. d) zeigt der Dynamikgewinn, wobei man gemäß der markierten blauen Linie (für die gelb markierte Linie im Musterbild a) den Unterschied anhand der Intensitätsprofile zwischen dem  $K_1$  und H Kanal eines HUV-Farbraumes erkennt.

Während die erste Hauptkomponente ( $K_1$ ) in diesem Farbraum die Luminanz enthält, beschreiben die anderen Komponenten ( $K_2$  und  $K_3$ ) die Chrominanz (Farbinformation), die für die Unterdrückung des Einflusses von Schatten und Helligkeitsschwankungen

## 2. INTENSITÄTSBASIERTE ANSÄTZE

---

verwendet werden. Dieses Ziel wird erreicht, da diese beiden Komponenten helligkeitsunabhängig sind und als Differenzrelation vorliegen. Die Komponenten  $K_2$  und  $K_3$  verhalten sich nahezu konstant, wenn sich die RGB-Farbwerte aufgrund von Helligkeitsschwankungen ändern [Al-Hamadi 01,02b]. Dieses Verhalten entspricht somit der visuellen Farbwahrnehmung, dass bei Variation der Sättigung oder der Helligkeit einer Farbe der Farbton in etwa konstant bleibt. Die in den transformierten Bildern enthaltene Gesamtvarianz lässt sich mittels dieses adaptiven Farbraumes durch die Eigenwerte der Kovarianzmatrix der RGB-Daten gemäß folgender Relation definieren:

$$\lambda_{Gesamt} = \lambda_1 + \lambda_2 + \lambda_3 \hat{=} 100\% \quad (2-2-1c)$$

Außerdem wird die Varianz für jede Komponente durch ihren zugehörigen Eigenwert angegeben. Durch die achromatische Projektionskomponente werden die meisten Informationen beschrieben. Diese entsprechen einer Teilvarianz von mehr als 92% der Gesamtvarianz. Diese prozentuale Angabe hängt jedoch stark von der jeweiligen betrachteten Bildregion ab. Zur Bestimmung der Teilvarianz führt man hier ein Qualitätsmaß als Gütekriterium für die Daten in der jeweiligen Komponente ein [Al-Hamadi 01a].

$$E_i = \lambda_i \cdot \left( \sum_{i=1}^{n=3} \lambda_i \right)^{-1} \quad \text{und} \quad E_3 = \frac{\lambda_1 + \lambda_2}{\lambda_1 + \lambda_2 + \lambda_3} \quad (2-2-1d)$$

Die Größe  $E_3$  beschreibt die Bildvarianz, die bei fehlender Komponente  $K_3$  vorhanden sind, während die Größe  $E_1$  die in der  $K_1$ -Komponente enthaltene Bildinformation beschreibt. Wenn der Wert des Qualitätsmaßes  $E_3$  in der Nähe von 100% liegt, dann sind die meisten Informationen in den beiden Komponenten  $K_1$  und  $K_2$  enthalten.

### 2.2.1.2 Bewegungsanalyse für die selektierten Objekte

Nach der Transformation der Farbbilder in den adaptiven Farbraum erfolgt die *zweite Verarbeitungsstufe* im System. Durch die Erweiterung vom intensitätsorientierten Blockmatching bei den Grauwertbildern zu einem intensitätsorientierten Multi-Blockmatching bei der Verwendung von Farbinformationen erhält man eine Verbesserung der Matchgüte, insbesondere bei nichtkooperativen Messsituationen. Wesentlich dafür ist die Tatsache, dass durch die Verwendung von Farbinformationen insbesondere bei geringeren Grauwertgradienten Verbesserungen auftreten in Hinsicht auf die Lagedetektion des Extremums gegenüber den bisherigen Graubildinformationen, die dazu führen, das Q-Ähnlichkeitsmaß (in unserem Fall haben wir das MAD-Kriterium als Q-Ähnlichkeitsmaß aufgrund der Einfachheit und Formeigenschaften angewendet) zu verfälschen. Dies liegt daran, dass in dem Graubild nicht mehr die nötigen Informationen enthalten sind, da die Helligkeitsübergänge bzw. Primitiven durch Grauwertbildung geschwächt werden bzw. verschwinden.

Als Erweiterung wurde mit beliebigen Blockgrößen gearbeitet, die an die Grenzen des sich bewegenden Objektes angepasst sind. Die Anpassung kann entweder durch den MD-Ansatz oder durch Vollsuche Blockmatching in den ersten beiden Bildern erfolgen [Al-Hamadi 01a]. Es wird nun angenommen, dass zusammenhängende Verschiebungsvektoren eine bewegte Region beschreiben. Aufgrund dessen, dass die Farbkomponenten unterschiedliche Ausprägung aufweisen, wurde hier ein kanalspezifisches Matching durchgeführt. Dies führt zu der Überlegung, aus den kanalspezifischen Ähnlichkeitsma-

## 2. INTENSITÄTSBASIERTE ANSÄTZE

---

Ben des vorgeschlagenen adaptiven Farbraumes ein kombiniertes Gesamtkriterium abzuleiten, um die optimale Ausnutzung der Farbinformationen und somit die Ausprägung der Intensitätswertgradienten zu erreichen. Bei diesem Gesamtkriterium wird über die Faktoren  $w_i$  eine Wichtung der einzelnen kanalspezifischen Ähnlichkeitsmaße entsprechend ihrer Zuverlässigkeit ermöglicht [Al-Hamadi 02d, 02a].

$$Q_k^{K_1 K_2 K_3}(u, v) = \left( \sum_{i=1}^3 w_i \right)^{-1} \cdot \left( w_1 \cdot Q_k^{K_1}(u, v) + w_2 \cdot Q_k^{K_2}(u, v) + w_3 \cdot Q_k^{K_3}(u, v) \right) \quad (2-2-2a)$$

wobei  $Q_k^{K_1}, Q_k^{K_2}$  und  $Q_k^{K_3}$  Ähnlichkeitsmaße der Komponenten  $K_1, K_2$  und  $K_3$  des adaptiven Farbraumes zum Zeitpunkt  $k$  sind. Dabei stellen  $(u, v)$  die horizontale und vertikale Bewegung dar und  $w_i$  die Wichtungsfaktoren.

Zur Bestimmung der Wichtungsfaktoren eignet sich ein künstliches neuronales Erkennungssystem [Al-Hamadi 01a, 02d] auf der Grundlage von bildspezifischen Merkmalen [Mecke 99]. Die Merkmale für das neuronale Netz werden in diesem System aus dem MAD-Ähnlichkeitsmaß abgeleitet [Al-Hamadi 01a], da sich beim MAD-Kriterium die bildspezifischen Störungen gut beschreiben. Beispielsweise verursacht eine homogene Beleuchtungsänderung eine Anhebung des Wertebereiches und des Minimums des MAD-Kriteriums, während bei inhomogener Beleuchtungsänderung eine zusätzliche Deformation an dem MAD-Kriterium auftritt. Überlagerte Bewegungen sowie partielle und vollständige Objektverdeckungen zeigen sich oft durch mehrere erkennbare Minima und die Anhebung des Minimums. Die extrahierten Merkmale werden aus einem einzelnen MAD-Ähnlichkeitsmaß (absolute Merkmale) und /oder aus zwei zeitlich aufeinander folgenden MAD-Ähnlichkeitsmaßen (Differenzmerkmale) ermittelt.

Die Berechnung der Merkmale aus zwei aufeinander folgenden Ähnlichkeitsmaßen hat den Vorteil [Al-Hamadi 01a], dass die Differenzmerkmale von den möglichen Offsets unabhängig sind. Dies liegt daran, dass die Offsetgrößen<sup>22</sup> in beiden Ähnlichkeitsmaßen durch die Differenzbildung eliminiert werden. Darüber hinaus kommen je nach Struktur des betrachteten Bildinhaltes verschiedene Formeigenschaften des MAD-Kriteriums vor, die bei absoluten Merkmalen meist zu einem großen Variationsbereich und somit zu Problemen bei der Interpretation der Merkmale führen. Die Einbeziehung von weiteren Differenzmerkmalen ist erforderlich, wenn die bildspezifischen Störungen mittels der verwendeten absoluten Merkmale nicht detektierbar sind. Demgemäß werden einige relevante Merkmale, die sich auf die absoluten und Differenzmerkmale beziehen, für die weiteren Untersuchungen verwendet [Al-Hamadi 01a, 02c, 02d].

Die bildspezifischen Merkmale  $f^{(1)}$  bis  $f^{(4)}$  in Tabelle 2-2-1 stellen Differenzmerkmale dar, da diese aus Differenzen von jeweils zwei zeitlich aufeinander folgenden Ähnlichkeitsmaßen berechnet werden. Somit beschreiben sie die Änderungen der Ähnlichkeitsmaße zwischen zwei Aufnahmezeitpunkten. Bei der Beschränkung auf eine Merkmalsart (absolute oder Differenzmerkmale) entsteht das Problem, dass die Störung nicht zuverlässig detektiert werden kann. Abhilfe erfolgt hier durch die Kombination von beiden relevanten Merkmalsarten miteinander [Al-Hamadi 01a]. Da für die Detektion der bildspezifischen Störungen im Zusammenhang mit dieser Arbeit die Merkmale in der Tabelle 2-2-1 ausreichend sind, soll auf weitere Untersuchungen bzw.

---

<sup>22</sup> Solche Offsetgrößen treten im Allgemeinen aufgrund des Kamerarauschens auf, das zur Anhebung des Wertebereiches der MAD-Funktion führt.

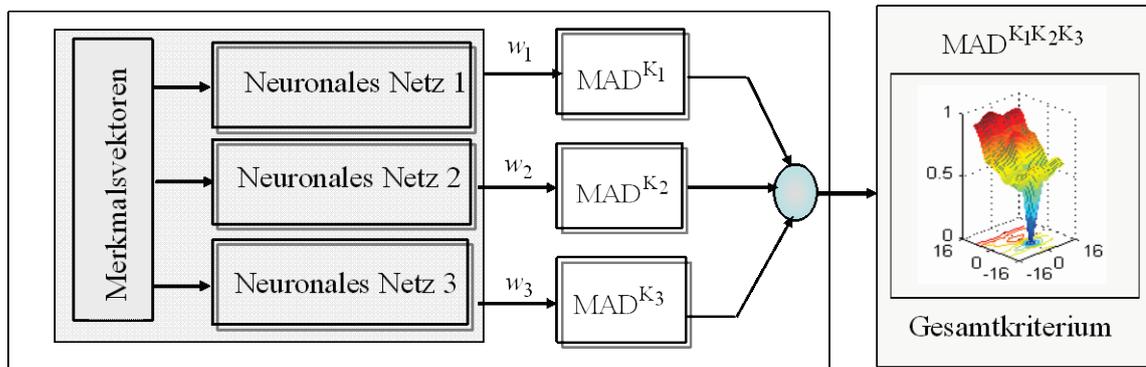
## 2. INTENSITÄTSBASIERTE ANSÄTZE

die Definition weiterer Merkmale verzichtet werden. Diese Merkmale bilden den Merkmalsvektor für das neuronale Erkennungssystem in Abbildung 2-2-3.

**Tabelle 2-2-1:** Merkmale zur Charakterisierung der Ähnlichkeitsmaße und zur Beschreibung der korrespondierenden Ausprägung der Ähnlichkeitsmaße.

Merkmale	Beschreibung
$f^1$	Änderung des Minimumwertes
$f^2$	Wertebereichsänderung
$f^3$	Mittelwertänderung
$f^4$	Änderung der Randbereichswerte
$f^5$	Wertebereich

Für die Realisierung des *neuronalen Erkennungssystems* wurde der Einsatz von überwacht angelegten Feed-Forward-Netzen eines Multi-Layer-Perzeptrons eingesetzt [Al-Hamadi 01a,03g,03f]. Für die beschriebene Aufgabe wird ein dreischichtiges Netz zur Verarbeitung der Merkmalsvektoren  $f$  in jeder Farbkomponente des adaptiven Farbraumes (s. Abb. 2-2-3) eingesetzt. Die Eingangsknoten gehören zur so genannten Eingabeschicht (Input-Layer), die verborgenen Knoten zur verborgenen Schicht (Hidden-Layer) und die Ausgangsknoten zur sogenannten Ausgabeschicht (Output-Layer). Die Übertragungseigenschaften dieser künstlichen neuronalen Strukturen werden unter Verwendung von Trainingsbeispielen über ein Lernverfahren durch die Einstellung von Gewichten antrainiert. Die Trainingsdaten werden aus einer Menge von Merkmalsvektoren gebildet, die man aus Testsequenzen generiert, welche hinreichend signifikante Störsituationen enthalten.



**Abbildung 2-2-3:** Kombiniertes Gesamtkriterium im Fall einer MAD-Funktion, wobei Netz 1 die achromatische MAD-Komponente bewertet. Netz 2 und 3 weisen die gleiche Netztopologie wie Netz 1 auf und sind den chromatischen Komponenten zugeordnet. Diese drei Netze haben nun die Aufgabe, die Wichtungsfaktoren in dem kombinierten Gesamtkriterium (Gl. 2-2-2a) zu ermitteln.

Die grundlegende Idee bei der Generierung des Gesamtkriteriums besteht darin, dass das neuronale Netz eine Ausgangsaktivierung ( $w_i$ ) liefert (s. Abb. 2-2-3), die die kanalspezifischen Ähnlichkeitsmaße  $MAD_k^{K_i}(u,v)$  einer Komponente des adaptiven Farbraumes anhand der Zuverlässigkeit wichtet. Somit werden die Ausgangsaktivierungen ( $w_i$ ) des jeweiligen kanalspezifischen Ähnlichkeitsmaßes unterdrückt, wenn das Messergebnis nicht zuverlässig ist. Daraus ergibt sich, dass das Ähnlichkeitsmaß in

Bezug auf die Lokalisierung des Minimums verbessert wird und dieses Konzept die Reduktion des Einflusses von nicht zuverlässigen Komponenten gewährleistet. Die detaillierte Beschreibung dieses Verfahrens würde leider den Rahmen dieser Arbeit sprengen (siehe auch [Al-Hamadi 01a]).

Aus den ermittelten Wichtungsfaktoren können gewisse Eigenschaften und Informationen in Bezug zu den bildspezifischen Störungen in der Szene abgeleitet und zwecks Objektverdeckung mittels *rekursiven Schätzverfahrens* auf Grundlage vom Kalman-Filter analysiert werden [Al-Hamadi 03f, 03g]. Die Wichtungsfaktoren können beispielsweise auch einen Wert von Null aufweisen, da reale Szenen von der Tatsache geprägt sind, dass sie nicht nur von den Effekten Schattenwurf und Beleuchtungsänderung überlagert sein können, sondern auch andere Effekte, wie z.B. die partielle Objektverdeckung, auftreten. Anhand dieser Wichtungsfaktoren wurden die Störeinflüsse erkannt.

Eine partielle Objektverdeckung in der betrachteten Szene kann vorliegen, falls die Wichtungsfaktoren der drei Komponenten gleichzeitig einen Wert nahe Null annehmen. Eine derart gestörte Szene erzwingt die qualitative Beurteilung der Messergebnisse. Es ist erforderlich, den Einfluss dieser Störungen zu minimieren, um aus den gestörten Messergebnissen möglichst sinnvolle Resultate zu gewinnen. Dabei bilden die berechneten Vektoren die Messvektoren sogenannten Trajektorie, die die tatsächliche Bewegung des Objektes beschreibt. Diese Messvektoren wurden als Eingangsdaten für ein *rekursives Schätzverfahren* (s. Abb. 2-2-1), das um bestimmte Erkennungseigenschaften erweitert wurde [Al-Hamadi 03f, 03g], verwendet. Dieses generiert bei Ausfall der Messinformation durch Verdeckungen o.ä. eine Ersatztrajektorie, die in einen begrenzten Bereich gültig ist (Linierarität). Aufgrund der rekursiven Struktur des Algorithmus geht jedoch nicht nur der jeweils aktuelle Messwert in den Algorithmus ein, vielmehr werden auch die zeitlich zurückliegenden Messwerte wirksam [Al-Hamadi 03g].

### 2.2.2 Experimentelle Ergebnisse mit der Systemstruktur

Um das Verhalten der Komponente des adaptiven Farbraumes bei der Bewegungsanalyse analysieren zu können, wird von dem Verlauf des MAD-Ähnlichkeitskriteriums<sup>23</sup> ausgegangen. Die Aufgabe der Projektionskomponente  $K_1$  muss es sein, ein möglichst ausgeprägtes Minimum des MAD-Ähnlichkeitskriteriums zu realisieren. Zudem ist bei relativ schwach ansteigenden Verläufen des MAD-Kriteriums ein steilerer Anstieg im Bereich der Umgebung des Extremums wünschenswert. Dies gilt insbesondere bei texturiertem Vordergrund und homogenen Beleuchtungsänderungen. Obwohl im ersten Fall das MAD-Ähnlichkeitskriterium deformiert wird, wird die Lage des Minimums des MAD-Kriteriums nicht wesentlich beeinflusst, da der Verlauf des MAD-Ähnlichkeitskriteriums im Bereich des Minimums einen größeren Anstieg aufweist. Die Texturen bewirken hier nur eine Deformation der Form des MAD-Kriteriums. Im zweiten Fall wird das Minimum angehoben.

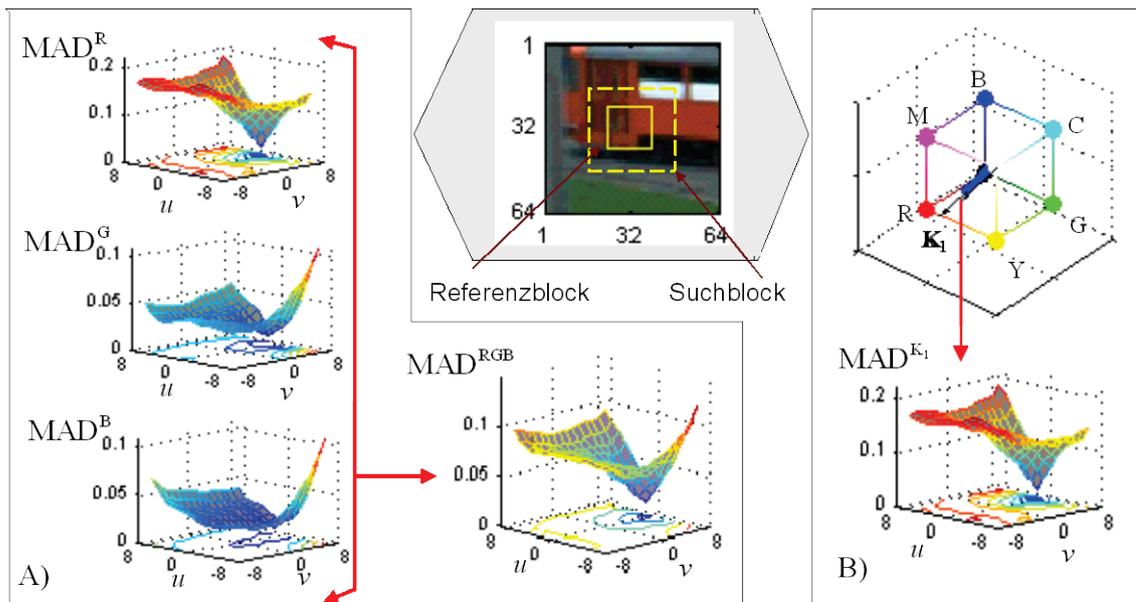
Vergrößert man nun die Dimension der herangezogenen Bildbereiche in dem adaptiven Farbraum, so werden sich die Koeffizienten der Projektionskomponente bei endlicher

---

<sup>23</sup> Die Verwendung des herkömmlichen Gesamtähnlichkeitskriteriums für RGB-Farbbilddaten unterscheidet sich sehr gering von dem mittelwertbasierten H-Kanal des HUV-Farbraumes. Dies liegt daran, dass das Gesamtkriterium auf dem Mittelwert basiert, welches einer Konvertierung der RGB-Farbdaten in eine höhere Ebene der Verarbeitung der aufeinanderfolgenden Bilder entspricht. Die Komponente  $K_1$  geht bei endlicher Dimension in den H-Kanal des HUV-Farbraumes über.

## 2. INTENSITÄTSBASIERTE ANSÄTZE

größerer Dimension angleichen. Als Folge davon ist eine Verbesserung anhand des Dynamikgewinns nicht mehr erzielbar [Al-Hamadi 01]. Vergleicht man nun das aus der Komponente  $K_1$  berechnete MAD-Kriterium für ungestörte Situationen mittels einer realen Bildsequenz mit dem Gesamtkriterium für die RGB-Daten, erkennt man einen deutlich steileren Minimumverlauf des  $MAD^{K_1}$ -Kriteriums (Abb. 2-2-2B). Diese Verbesserung beruht auf der Tatsache, dass sich die größte Streuung der Farbverteilung der betrachteten Bildregion in Richtung der Varianz „Rot“ zeigt, welche einen größeren Wichtungsfaktor in der Transformationsmatrix aufweist, während die anderen Komponenten keine dominanten Wichtungsfaktoren besitzen. a, 02a, 03f].



**Abbildung 2-2-2:** Gesamtkriterium  $MAD^{RGB}$  (A) und aus der Komponente  $K_1$  berechnetes Ähnlichkeitsmaß (B) für einen Referenzblock. Die Komponente  $K_1$  ist oben in B) dargestellt. Sie zeigt in Richtung der roten Komponente.  $MAD^{RGB}$ -Kriterium wird als Mittelwertbildung ( $w_i=0,3$ ) von allen drei kanalspezifischen Ähnlichkeitsmaßen in RGB-Kanälen (A) gebildet.

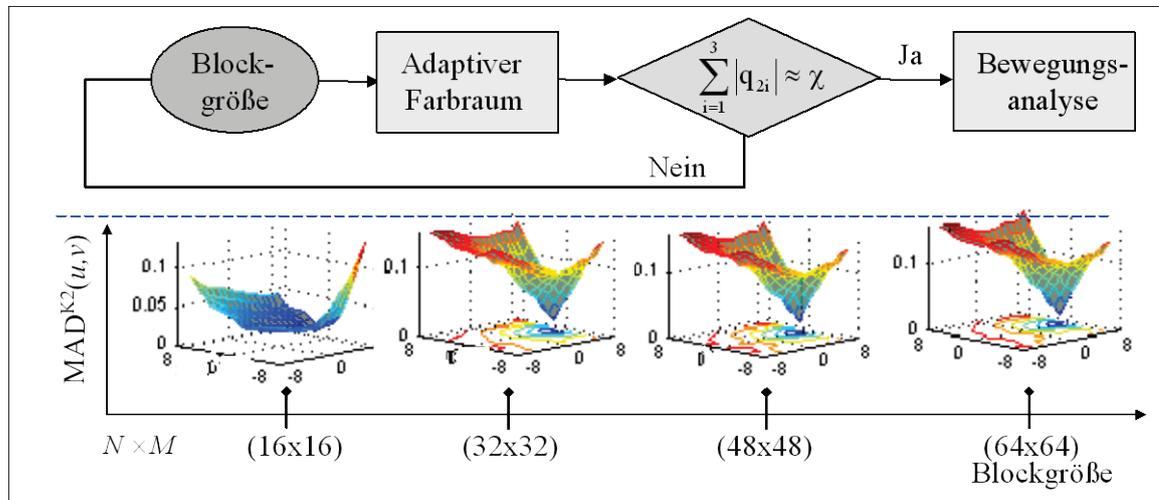
Der Verlauf des MAD-Kriteriums der achromatischen Komponente in dem adaptiven Farbraum entspricht in diesem Fall nahezu dem Verlauf des MAD-Kriteriums des roten Kanals des RGB-Farbraumes (Abb. 2-2-2B). Der Abstand zwischen Maximalwert und Minimalwert verdoppelt sich bei der  $K_1$ -Komponente verglichen mit dem  $MAD^{RGB}$ -Gesamtkriterium. Dieser Abstand beschreibt die Luminanzdifferenz zwischen Objektvordergrund und Hintergrund. Dadurch werden die geringeren Störungen wie beispielsweise texturierter Vordergrund (z. B. Gitter oder die Äste eines Baumes) und geringere Grauwertgradienten die Lage des Minimums des Ähnlichkeitsmaßes nicht verfälschen.

Bei einer Betrachtung der Komponenten  $K_2$  und  $K_3$  scheint es, dass die beiden Komponenten ein ähnliches Verhalten in Bezug auf die Differenzbildung aufweisen. Sie unterscheiden sich nur dadurch, dass sie einerseits die Beziehung zwischen unterschiedlichen Farbkomponenten beschreiben und andererseits, dass die meisten Farbinformationen in der Komponente  $K_2$  enthalten sind, da diese eine größere Teilvarianz als  $K_3$  aufweist. Interessante Aussagen folgen, wenn man die passende Blockdimension<sup>24</sup>

<sup>24</sup> Durch die Einbeziehung größerer Blöcke in der  $K_2$ -Komponente wird das Resultat in Richtung eines steileren Minimums verbessert. Interessant scheint daher die Optimierung der Blockdimension, die die Forderung an die

## 2. INTENSITÄTSBASIERTE ANSÄTZE

bestimmt, welche für die Transformation in den adaptiven Farbraum herangezogen wird. Unter diesem Aspekt wird davon ausgegangen, dass der adaptive Farbraum eine vordefinierte *Verlustfunktion* erfüllen soll, um die transformierten Daten optimal auszunutzen.



**Abbildung 2-2-3:** Die Optimierung des adaptiven Farbraumes anhand einer Verlustfunktion. Abhängigkeit des MAD-Verlaufes in dem chromatischen  $K_2$ -Kanal des adaptiven Farbraumes von der Größe des betrachteten Blockes.

Vergrößert man anschließend die Blockgröße schrittweise (s. Abb. 2-2-3), so erkennt man, dass das berechnete MAD-Kriterium einen besseren signifikanten Verlauf im Bereich des Minimums zeigt. Dies ist darauf zurückzuführen, dass die Ausgangsdaten optimal auf  $K_2$  projiziert sind und dazu die Intensitätswerte der Farbe (höchst mögliche Farbinformation wird dadurch erhalten) verstärken. Ausgehend von der Tatsache, dass die Summation der Beträge der Gewichte der chromatischen Komponente  $K_2$  in der Transformationsmatrix einen Wert  $\chi$  in der Nähe von Null ergibt, kann eine Verlustfunktion abgeleitet werden, die die Aufgabe hat, die ideale Blockdimension beim adaptiven Farbraum zu ermitteln.

Die Abbildung 2-2-3 verdeutlicht die Funktionsweise der Verlustfunktion ( $q_{2i}$  sind die Gewichte in der Transformationsmatrix). Erfüllt diese Verlustfunktion die Bedingung in der Abbildung 2-2-3, dann kann die Bewegungsanalyse erfolgen, ansonsten muss die Blockdimension noch vergrößert werden. Bei einer optimalen Nutzung des Farbraumes kann das MAD-Kriterium in der  $K_2$ -Komponente einen signifikanten Verlauf annehmen, welcher zur Verbesserung des Ergebnisses bezüglich der Genauigkeit beiträgt [Al-Hamadi 00b, 01a].

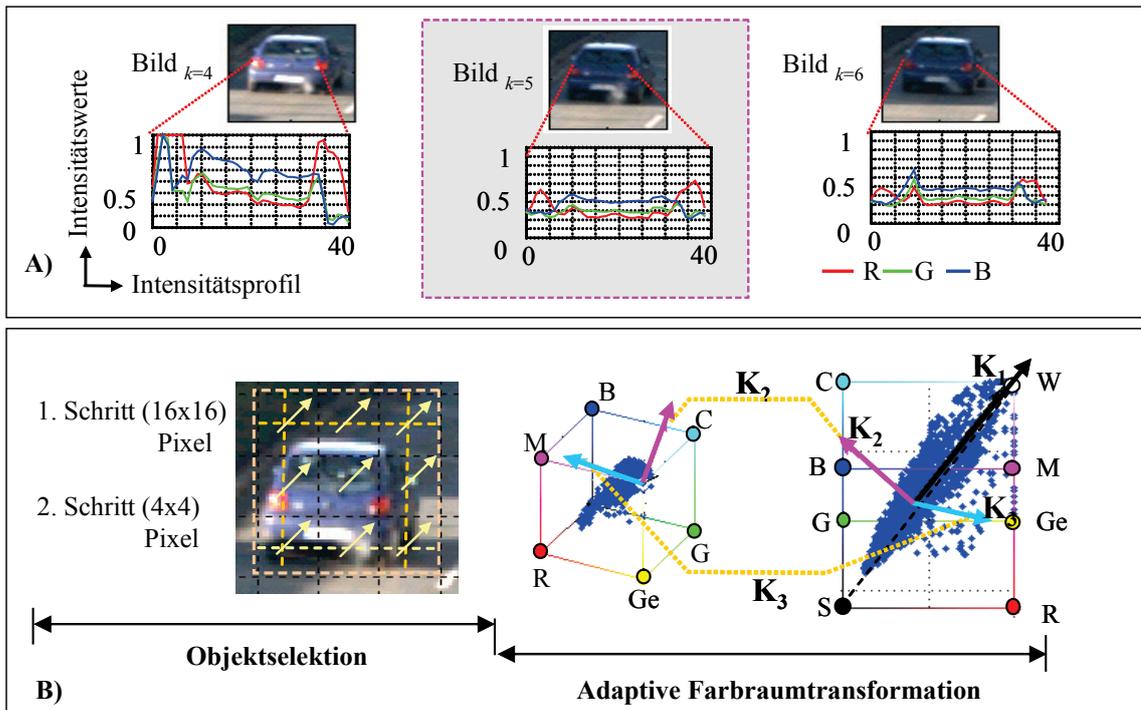
### 2.2.2.1 Einfluss bildspezifischer Störungen

Im Weiteren werden die Eigenschaften des adaptiven Farbraumes anhand der Bewegungsanalyse unter besonderer Berücksichtigung bildspezifischer Störungen interpretiert. In der Abbildung 2-2-4a ist hierfür ein Beispiel aus einer realen Bildsequenz dargestellt, das diesen Effekt verdeutlicht. Diese enthält ein bewegtes Objekt, das vom Schattenwurf der Bauten in Fahrbahnnähe überlagert wird. Die Überlagerung führt zu

Genauigkeit des Verlaufes der MAD-Funktion realisiert, ohne dass man die Ergebnisse subjektiv zu beurteilen braucht.

## 2. INTENSITÄTSBASIERTE ANSÄTZE

starken Modifikationen der Intensitätswerte (s. Abb. 2-2-4a) und somit zu einem Korrespondenzproblem. Für die betrachtete Bildregion enthält die Komponente  $K_1$  Informationen, die etwa 90% der Gesamtvarianz der Bildregion entsprechen. Diese prozentuale Angabe ändert sich natürlich mit der jeweils betrachteten Region. Bei der Vergrößerung der Bildregion, wie in Abbildung 2-2-4b, gleicht sich die  $K_1$ -Komponente dem H-Kanal des HUV-Farbraums immer stärker an. Bezüglich kleiner bzw. objektangepasster Bildregionen verstärkt  $K_1$  die Primitiven, während sie bei dem H-Kanal geschwächt werden.



**Abbildung 2-2-4a:** A) Ausschnitt aus einer Bildsequenz mit dem Einfluss eines Schattenwurfs im Bild  $k=5$  und einer Beleuchtungsänderung im Bild  $k=10$  (s. Anhang 8.1). Zusätzlich sind die Intensitätsprofile (RGB-Kanäle) für die markierte Zeile zwischen den gekennzeichneten Punkten in der Sequenz angegeben. B) Angepasstes Objekt mittels farbklassenbasierten Blockmatchings und b) die mittels Verschiebungsvektoren (Vollsuche-Blockmatching) mit den jeweiligen adaptiven Farbraumkomponenten. Die adaptiven Farbkomponenten haben die gleiche Richtung.

Die  $K_2$ - und  $K_3$ -Komponenten im adaptiven Farbraum werden mit dem Ziel der Minimierung des Einflusses von Helligkeitsschwankungen und Schattenwurf ausgewertet [Al-Hamadi 01, s. Anhang 8.1]. Berechnet man für die Bildregion diese beiden Farbkomponenten, so erhält man die folgenden Gleichungen (2-2-3a/3b). Addiert bzw. subtrahiert man zu den Farbwerten in dieser Gleichung eine Konstante  $\Delta h$  als Maß für die Helligkeitsschwankung, dann zeigt sich, dass diese den Wert von  $K_3$  innerhalb der Bildregion über die Sequenz kaum beeinflusst.

$$K_3 = 0,32 \cdot R - 0,81 \cdot G + 0,48 \cdot B \approx 0,32 \cdot (R + \Delta h) - 0,81 \cdot (G + \Delta h) + 0,48 \cdot (G + \Delta h)$$

$$K_3(R + \Delta h, G + \Delta h, B + \Delta h) \approx K_3(R, G, B) + \Delta h \cdot \underbrace{(0,32 - 0,81 + 0,48)}_{\varepsilon} \quad (2-2-3a)$$

Aus der Gleichung (2-2-3a) ergibt sich nämlich, dass die Summe der Faktoren  $\varepsilon$  näherungsweise Null (-0.01) wird. Diese Gleichung gibt also die Beziehung wieder, die

## 2. INTENSITÄTSBASIERTE ANSÄTZE

---

zwischen den beiden Komponenten  $M = (R+B)$  und  $G$  in Bezug auf die betrachtete Bildregion in Abbildung 2-2-4a besteht. Ähnliches gilt für die  $K_2$ -Komponente.

$$K_2 = -0,74 \cdot R + 0,10 \cdot G + 0,66 \cdot B = -0,74 \cdot (R + \Delta h) + 0,10 \cdot (G + \Delta h) + 0,66 \cdot (B + \Delta h)$$

$$K_2(R + \Delta h, G + \Delta h, B + \Delta h) \approx K_2(R, G, B) + \Delta h \cdot \underbrace{(-0,74 + 0,10 + 0,66)}_{\varepsilon} \quad (2-2-3b)$$

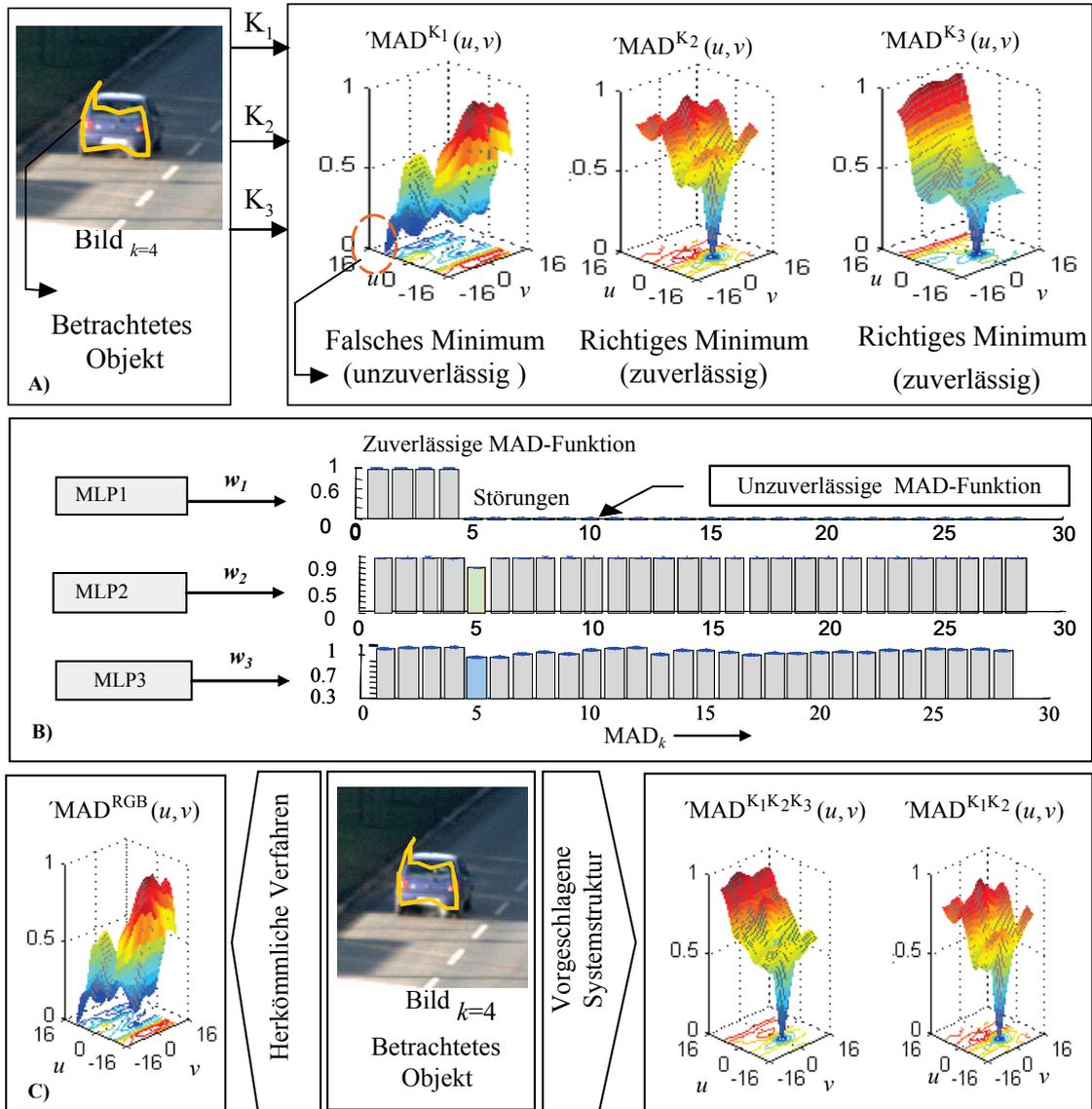
Anhand der Gleichung (2-2-3b) erkennt man die Beziehung, die zwischen den beiden Farbwerten  $R$  und  $B$  in Bezug auf die betrachtete Bildregion besteht. Die  $G$ -Komponente ist hier nicht dominant und bei der Interpretation vernachlässigbar. Wie in den Gleichungen (2-2-3a/3b) zu sehen ist, ändern sich die  $K_2$ - und  $K_3$ -Komponenten nur wenig, wenn von jedem Farbwert eine Konstante  $\Delta h$  abgezogen bzw. addiert wird. Dieses Ergebnis ist dem Verhalten der  $U$ - bzw.  $V$ -Komponente des HUV-Farbraumes ähnlich und es entspricht der visuellen Farbwahrnehmung - variiert die Sättigung oder die Helligkeit einer Farbe, so bleibt der Farbton annähernd konstant.

Nach der Transformation der Farbbilder in den adaptiven Farbraum werden diese in die zweite *Verarbeitungsstufe* des Systems weitergeleitet (s. Abb. 2-2-1). Dort wird die Bewegung der Bildregion (Bewegungsvektor  $\mathbf{v}$ ) über die Lagekoordinaten des MAD-Minimums ermittelt. Durch die Verwendung von objektangepassten Bildregionen aus der *ersten Verarbeitungsstufe* des Systems geht man zum ikonischen Matching über. Dieses arbeitet ähnlich wie das Blockmatching-Verfahren.

Der Vorteil des Blockmatching ist die Einfachheit und die dichte Verschiebungsvektorfelder. Die wesentlichen Nachteile des Blockmatchings sind der hohe Rechenaufwand, die *feste Wahl* des Blockrasters, das Fehlen eines *eindeutigen Minimums* innerhalb homogener Gebiete und das *Blendenproblem* aufgrund rein lokaler Sicht. Die letzten drei Nachteile sind beim *ikonischen Matching* (Regionsbasiertes Matching in Intensitätsebene) nicht mehr vorhanden. Es liefert sehr gute Ergebnisse im Vergleich zum Blockmatching-Verfahren, da die Blockdimension beim ikonischen Matching an die Objektform angepasst ist. Die Objektgrenzen wurden in dieser Arbeit entweder durch die Verwendung von Verschiebungsvektorfeldern [Al-Hamadi 00a], die Verwendung des farbklassenbasierten Blockmatching [Al-Hamadi 00b] oder die Verwendung des MD-Ansatzes gemäß [Al-Hamadi 03a] festgelegt. Die ersten beiden Verfahren basieren auf dem Blockmatching. Das erste ist einfach und fordert nur die Erweiterung auf das hierarchische Matching, um die angepasste Grenze des Objektes zu extrahieren. Der Suchbereich ergibt sich aus den maximal in horizontaler und vertikaler Richtung zu erwartenden Verschiebungen, die als A-priori Information vorliegen.

In der Abbildung 2-2-4b ist dies am Beispiel einer realen Bildsequenz dargestellt. Die Problemstellung ist die Überlagerung eines bewegten Objektes durch den Schattenwurf. Dieser Effekt führt zu deutlichen Änderungen der Intensitätswerte (z.B. Abdunklung beim Schattenwurf) im Bild  $k=5$  des betrachteten Objektes. Im Ergebnis dieser Überlagerung treten im Ähnlichkeitsmaß Deformationen und mehrere Minima auf. Dies geschieht sowohl in der helligkeitsabhängigen Komponente  $K_1$  sowie im kanalspezifischen mittelwertbasierten RGB-Gesamtkriterium gemäß [Al-Hamadi 01a], wenn das Matching unter Verwendung der Bildinhalte im RGB-Farbraum durchgeführt wird. Daraus resultierende fehlerhafte Ergebnisse bei der Bestimmung der Bewegungsvektoren und bei der Sequenzanalyse führen dazu, dass die betrachtete Bildregion von der ursprünglich verfolgten Region wegdriftet (s. Anhang 8.1).

## 2. INTENSITÄTSBASIERTE ANSÄTZE



**Abbildung 2-2-4b:** A) Die berechneten MAD-Kriterien für das betrachtete Objekt (bei der Überlagerung von einem Schattenwurf) in den jeweiligen Kanälen des adaptiven Farbraumes. B) Der zeitliche Verlauf der Ausgangsaktivierung des neuronalen Netzes 1 (für die  $K_1$ -Komponente), Netzes 2 (für die  $K_2$ -Komponente) und des Netzes 3 ( $K_3$ -Komponente) für die Testsequenz „Schatten“. C) MAD-Analyseergebnisse für das betrachtete Objekt aus der Testsequenz „Schatten“ bei der Überlagerung von inhomogenen Beleuchtungsänderungen (Schattenwurf) mittels des vorgeschlagenen Systems (rechts) im Vergleich zum herkömmlichen Matchingverfahren (links).

Die Ähnlichkeitsmaße der beiden Komponenten  $K_2$  und  $K_3$  ermöglichen aufgrund der guten Minimumsausprägung hingegen eine fehlerfreie Bewegungsbestimmung (Abb. 2-2-4b). Dies ist darauf zurückzuführen, dass die Berechnung der Korrespondenzen mit Hilfe des ikonischen Matching erfolgt und andererseits die beiden Komponenten des adaptiven Farbraums optimal durch die objektangepasste Transformation der Bildregion eingestellt sind. Ausgehend davon nimmt unabhängig von dem Ähnlichkeitsmaß die Wahrscheinlichkeit für eine Fehlpositionierung mit der objektangepassten Bildregion ab. Darüber hinaus ist der Einfluss des Blendenproblems und der geringeren Grauwertgradienten aufgrund rein lokaler Sicht nicht mehr auf der Gesamtauswertung beträchtlich. Es erscheint analog zum Blockmatching sinnvoll, aus den kanalspezifischen

## 2. INTENSITÄTSBASIERTE ANSÄTZE

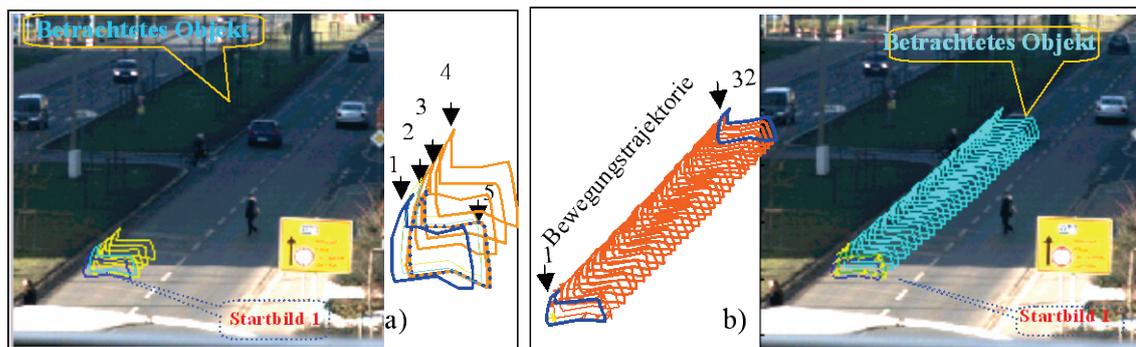
Ähnlichkeitsmaßen des adaptiven Farbraumes das kombinierte Gesamtkriterium wie in Gleichung (2-2-2a/2b bei der Komponentenreduzierung) zu berechnen (Table 2-2-1).

**Tabelle 2-2-1:** Die qualitative Bewertung des MAD-Kriteriumes unter dem Einfluss von bildspezifischen Bildstörungen [Al-Hamadi 03f].

	Geringere Grauwertgradient	Schattenwurf und Beleuchtungsänderungen	Teilverdeckung	Andere Störungen
$MAD_{K_1}$	zuverlässig	unzuverlässig	unzuverlässig	zuverlässig
$MAD_{K_2}$	zuverlässig	zuverlässig	unzuverlässig	unzuverlässig
$MAD_{K_1K_2K_3}$	zuverlässig	zuverlässig	unzuverlässig	zuverlässig

Durch eine Kombination von diesen absoluten Merkmalen und Differenzmerkmalen, die dazu beitragen, bildspezifische Störungen zu beschreiben, entsteht der Eingangsmerkmalsvektor<sup>25</sup> des neuronalen Erkennungssystems. Dieser Merkmalsvektor bildet die Voraussetzung, das MAD-Kriterium als zuverlässig oder unzuverlässig zu bewerten. Da das Objekt von einem Schattenwurf ab dem fünften Bild überlagert ist, bewertet das neuronale Netz 1 das berechnete MAD-Kriterium in der achromatischen Komponente  $K_1$  ab dem fünften Bild als unzuverlässig (Abb. 2-2-4b rechts).

Die neuronalen Netze 2 und 3 im Erkennungssystem verhalten sich anders als das Netz 1, wobei sie eine Ausgangsaktivierung im Bereich 0.8 und 0.9 bei der bildspezifischen Störung im Bild  $k=5$  liefern. Sie bewerten die MAD-Kriterien in den Komponenten  $K_2$  und  $K_3$  als zuverlässig. Als Resultat wird das  $MAD_{K_1}^{K_1}$ -Kriterium durch die Ausgangsaktivierung  $w_1$  mit Null multipliziert, während das  $MAD_{K_2}^{K_2}$ - und  $MAD_{K_3}^{K_3}$ -Kriterien durch die Ausgangsaktivierungen  $w_2$  und  $w_3$  als zuverlässiges MAD-Kriterium bewertet werden und in das kombinierte Gesamtkriterium  $MAD^{K_1K_2K_3}$  eingehen.



**Abbildung 2-2-4c:** Die Analyseergebnisse bei der Betrachtung längerer Sequenzen als Bewegungstrajektorie a) beim herkömmlichen intensitätsbasierten Matchingsverfahren und b) gemäß der vorgeschlagenen adaptiven Systemstruktur.

Das so ermittelte Gesamtkriterium zeigt exemplarisch erreichbare Verbesserungen im Vergleich zu den herkömmlichen Verfahren in der Bewegungsanalyse (Abb. 2-2-4b links). Bei der Verwendung des herkömmlichen Verfahrens beschreibt die Bewegungs-

<sup>25</sup> Merkmale zur Charakterisierung der Ähnlichkeitsmaße und zur Beschreibung der korrespondierenden Ausprägung der Ähnlichkeitsmaße sind Änderung des Minimumwertes, Wertebereichsänderung, Mittelwertänderung, Änderung der Randbereichswerte und Wertebereich [Al-Hamadi 01a, Al-Hamadi 02a]

trajektorie beim Eintritt des Objektes in den Schattenbereich nicht mehr die tatsächliche Objektbewegung. Dahingegen ermöglicht die vorgeschlagene Systemstruktur die Verfolgung der betrachteten Bildregion über die gesamte Bildsequenz (Abb. 2-2-2c). Die Bewegungstrajektorie bricht ab dem Bild  $k=5$  beim herkömmlichen Verfahren (Schatteneffekt) ab, während sie bei dem vorgeschlagenen System einen stabilen Verlauf aufweist, der die tatsächliche Bewegung des Objektes beschreibt.

Weitere Untersuchung zeigt, dass der Wert der ersten beiden Komponenten des adaptiven Farbraumes meist in der Nähe von 98% der Gesamtvarianz liegt. Dies ist auf die Tatsache zurückzuführen, dass reale Bildszenen vorwiegend Grautöne und wenig gesättigte Farbe enthalten. Deshalb kann die Achse  $K_3$  häufig ohne signifikanten Verlust an Information vernachlässigt werden, da die beiden ersten Komponenten die Merkmale enthalten, die zur Beschreibung der Primitiven des Objektes ausreichen. Dadurch geht die Gleichung (2-2-2a) in Gleichung (2-2-2b) [Al-Hamadi 03f] über.

$$Q_k^{K_1K_2}(u,v) = \left( \sum_{i=1}^2 w_i \right)^{-1} \cdot \left( w_1 \cdot Q_k^{K_1}(u,v) + w_2 \cdot Q_k^{K_2}(u,v) \right) \quad (2-2-2b)$$

### 2.2.3 Schlussfolgerung /Diskussion

Der Vorteil des Matching ist die Einfachheit und die Tatsache, dass dieses stabil arbeitet und relativ dichte Verschiebungsvektorfelder liefert, die ausgenutzt werden können, um zusätzliche Anhaltspunkte über die Objektform abzuleiten. Die wesentlichen Nachteile des herkömmlichen Blockmatchings sind der hohe Rechenaufwand, die feste Wahl des Blockrasters, das Fehlen eines eindeutigen Minimums innerhalb homogener Gebiete und das Apertur-Problem aufgrund rein lokaler Sicht. Die letzten drei Nachteile sind bei dem entwickelten *Matchingverfahren* aufgrund des vorgeschlagenen ikonischen Multi-Matching nicht mehr vorhanden. Es liefert signifikante Verbesserungen im Vergleich zum einfachen Blockmatching-Verfahren, da die Blockdimension beim ikonischen Matching an die Objektform optimal angepasst ist.

Die automatische initiale Objekterfassung wurde in dieser Arbeit durch die Verwendung von Verschiebungsvektorfeldern [Al-Hamadi 00a] realisiert. Dieses Verfahren basiert auf dem einfachen Vollsuche-Blockmatching zwischen den ersten beiden Bildern. Das erste ist einfach und fordert nur die Erweiterung auf das hierarchische Matching, um die angepasste Grenze des Objektes optimal zu verfeinern. Allerdings wird dadurch der Rechenaufwand erhöht. Besonders interessant wäre hier der Einsatz des modifizierten differenzbildbasierten Ansatzes [Al-Hamadi 04c], da dieser einerseits *einfach* und nicht *rechenaufwendig* ist und andererseits eine starke *Datenreduzierung* und *genaue* Objekterfassung liefert [Al-Hamadi 04c]. Die dadurch erfassten Regionen bilden die Objektkandidaten für die Transformation in den adaptiven Farbraum, die auch wiederum die Grundlage für weitere Bewegungsanalyse.

Durch den Einsatz des adaptiven Farbraumes könnte man den Einfluss von Beleuchtungsänderungen bei der Berechnung der Verschiebungsvektoren minimieren. Allerdings gibt es eine andere Art von Schattenwurf<sup>26</sup>, der die Berechnung des Vektorfeldes nicht beeinflusst aber die Erfassung und Beschreibung von Objektbegrenzungen erschwert. Dies führt zu einer Verschmelzung von bewegten Regionen und somit auch zum Korrespondenzproblem (aufgrund von Vorkommen von mehreren Objekten in einer

---

<sup>26</sup> Man unterscheidet hier self shadow und shadow casting.

## 2. INTENSITÄTSBASIERTE ANSÄTZE

---

Region). Dies zeigt die sinnvolle Integration eines Schattenfilters bei einem MD-Ansatz gemäß [Al-Hamadi 05d], der auch auf der Farbinformation basiert. Dadurch wäre eine genaue Initialisierung gesichert.

Durch die achromatische Projektionskomponente des adaptiven Farbraumes werden die meisten Informationen beschrieben. Diese entsprechen einer Teilvarianz von mehr als 92% der Gesamtvarianz [Al-Hamadi 01a, 02b und 03g]. Dies ist auf die Tatsache zurückzuführen, dass reale Bildszenen vorwiegend Grautöne und wenig gesättigte Farben enthalten. Deshalb wurde im weiteren Arbeiten die Komponente  $K_3$  des adaptiven Farbraumes häufig ohne signifikanten Verlust an Information vernachlässigt [Al-Hamadi 03f], da die beiden ersten Komponenten die Merkmale enthalten, die zur Beschreibung der Primitiven des Objektes ausreichen. Die Entfernung von Komponente  $K_3$  des adaptiven Farbraumes hat zu einer Reduzierung des Multi-Matchings auf zwei Komponenten [Al-Hamadi 03f und 03g] geführt, bei denen auch der Rechenaufwand um ein Drittel reduziert werden konnte. Die erste Komponente enthält die Luminanzinformationen. Es wird ein größerer Dynamikgewinn erzielt und gleichzeitig wird durch die zweite Komponente die Chrominanzinformation ausgedrückt, so dass anschließend eine gewisse homogene Farbabbildung entsteht. Durch diese Eigenschaften unterscheidet sich der adaptive Farbraum von dem wahrnehmungsorientierten HUV-Farbraum, d.h. der adaptive Farbraum kann meist auf einen zweidimensionalen Farbraum ohne signifikanten Informationsverlust reduziert werden.

Bewertungskriterium für die Konvertierung von Farbdaten in Grauwertdaten durch die  $K_1$ -Komponente ist hierbei, in welchem Umfang man die Primitiven bzw. Helligkeitsübergänge in allen drei Farbkomponenten möglichst ungeschwächt auf die Projektion überträgt. Da jedoch die Projektionsvektoren für jeden Block des unterteilten Bildes im Allgemeinen unabhängig berechnet werden, können von Block zu Block starke Diskontinuitäten in der Projektionsrichtung auftreten. So kann im Extremfall ein Projektionsvektor, der etwa senkrecht auf dem Mittelwert der Daten steht, von Block zu Block „kippen“ und so zufällig die eine oder die andere Seite des Helligkeitsüberganges dunkel bzw. hell erscheinen lassen.

Für die Beurteilung der Messwertzuverlässigkeit wurde z.B. ein neuronales Erkennungssystem eingesetzt. Dieses erhält als Eingangsdaten Merkmalsvektoren, die aus den beim Matching berechneten Ähnlichkeitsmaßen gewonnen werden. Das Erkennungssystem besitzt die Fähigkeit, typische bildspezifische Störungen anhand der korrespondierenden Merkmalsausprägung zu detektieren. Außerdem realisiert es eine Adaption des Schätzverfahrens. Als Ergebnis wird eine hohe Störsicherheit und Robustheit des Gesamtsystems erreicht. Dazu ermöglicht das angewandte Schätzverfahren in [Al-Hamadi 03f] die Einbeziehung von A-priori-Wissen über die Objektbewegung in Form eines kinematischen Modells [Mecke 99], wodurch eine hohe Effizienz und Genauigkeit in Fall von Teilverdeckung erzielt wird. Durch das eingesetzte Schätzverfahren könnte eine wesentliche Minimierung des Einflusses statistischer Störungen auf das Analyseergebnis erreicht werden.

Durch die Erweiterung auf Farbbilder/3D können zusätzliche Aussagen erwartet werden. Insbesondere wird der Bereich der Szenenauswertung ohne zusätzliche strukturierte Beleuchtung erweitert. 3D-Informationen könnte auch hier für eine weitere Verbesserung der Erfassung von bewegten Objekten bei der bewegten Kamera genutzt werden. Eine Möglichkeit dafür ist die Auswertung von Bildern, die zeitgleich von zwei räumlich versetzten Kameras aufgenommen wurden (Stereobildpaar) [Ohashi 97, Zhang 92, Dror 95, Tornow 03]. Eine Weiterverarbeitung wird nur vorgenommen, wenn die Korrelation

einen gewissen Grenzwert überschreitet. Durch Triangulation werden die den relevanten Disparitäten zugeordneten Koordinaten berechnet. Im zweiten Schritt, der Bildsegmentierung, werden die mit dem Stereosystem ermittelten 3D-Punkte zu sog. „Clustern“ zusammengefasst [Al-Hamadi 06e]. Dazu wird nach 3D-Punkten gesucht, die ähnliche Ortskoordinaten aufweisen und eine bestimmte Höhe ergeben. Ist eine solche Konstellation gefunden, so kennzeichnet den Bildbereich eines erhabenen Objektes [Al-Hamadi 06e, Elmezian 08a, 09a]. Generell bewirken die Objekte im Tiefenbild eine Häufung eines bestimmten Disparitätswertes. Ausgehend davon könnte hier der Aufbau eines Tiefenhistogrammes realisiert und ausgewertet werden [Al-Hamadi 06a]. Die Zweckmäßigkeit einer Kombination der Ergebnisse ist noch ein Forschungsbedarf.

### 2.2.4 Zusammenfassung

Mit der Arbeit in diesem Teilkapitel wurde die Entwicklung einer adaptiven Systemstruktur zur Bewegungsanalyse in Farbvideosequenzen unter der Berücksichtigung von nichtkooperativen Messsituationen beschrieben. Dabei wurden sowohl Methoden der Bildverarbeitung als auch der Einfluss von Farbinformationen betrachtet. Daraus resultierten zwei Verarbeitungsschritte: die Aufbereitung der Farbinformation zur Objekterkennung und die eigentliche Bewegungsanalyse in Farbbildsequenzen. Ausgehend von den grundlegenden Überlegungen wurde ein Farbraum in Abhängigkeit von den Bildinhalten definiert. Dieser basiert auf der Grundidee, dass die Erregungen der Rezeptoren im menschlichen Sehsystem durch die starke Überlappung der spektralen Empfindlichkeitskurven der unterschiedlichen Rezeptoren hoch korreliert sind. Somit berechnet man hier eine Transformationsmatrix, deren Anwendung dazu führt, dass die Erregungen der Kanäle weitgehend unkorreliert sind, d.h. es wird die höchstmögliche Informationsmenge transportiert wird. Dieser adaptive Farbraum gewährleistet in Verbindung mit einem Korrelationsverfahren die Bestimmung der Bewegungsgrößen eines Objektes auch in ungünstigen Situationen wie beispielsweise nichtkooperativen Messsituationen.

Für die Realisierung des bildspezifischen Erkennungssystems wurde der Einsatz von überwacht angelegten Feed-Forward-Netzen vorgeschlagen. Die Übertragungseigenschaften dieser künstlichen neuronalen Strukturen werden unter Verwendung von Trainingsbeispielen über ein Lernverfahren durch die Einstellung von Gewichten antrainiert. Diese neuronalen Netze haben die Aufgabe, die Wichtung der Kanäle für das Multi-Kanal-Matching einzustellen. Die beschriebenen Überlegungen legen ein intelligentes Bildverarbeitungssystem nahe, das, ähnlich dem visuellen System des Menschen, ohne Modellierung der Szene auskommt und adaptiv auf Veränderungen im Bild reagiert. Dabei wird die Bewegungsbestimmung von Objekten nicht nur in der Schwarzweiß-Domäne, sondern auch in der Farbdomäne durchgeführt. Es wurde die Bedeutung der einzelnen Komponenten des vorgeschlagenen Systems und deren Zusammenwirken beschrieben.

Die Eignung der vorgeschlagenen Systemstruktur zur robusten Bewegungsbestimmung und Objektverfolgung wurde anhand realer Farbbildsequenzen nachgewiesen, die häufig vorkommende bildspezifische Störungen enthalten (s. Anhang 8.1). So standen im Mittelpunkt der Untersuchungen ausgewählte bildspezifische Störungen wie geringere Grauwertgradienten, Schattenwurf und inhomogene Beleuchtungsänderungen. Es wurde gezeigt, dass mittels der vorgeschlagenen Systemstruktur eine höhere Störsicherheit bei der Bewegungsanalyse beim Auftreten dieser Störeffekte erreicht

## **2. INTENSITÄTSBASIERTE ANSÄTZE**

---

wird. Es können Messsituationen mit dem entwickelten System erfasst werden, die mit anderen derzeit verfügbaren Verfahren nicht auswertbar sind.

Die durchgeführten Untersuchungen und die beschriebenen Überlegungen legen eine robuste Systemstruktur nahe, die ohne Modellierung der Szene auskommt und adaptiv auf Veränderungen im Bild reagiert. Konkret legt die funktionelle Beschreibung des Systems zur Bewegungsanalyse zunächst weder ein Bewegungsmodell noch den verwendeten Farbraum fest. Dabei wird die Bewegungsbestimmung von Objekten nicht nur wie bisher überwiegend in der Schwarz-Weißdomäne sondern in der Farbdomäne durchgeführt.

## 3. Matching-Paradigma

Merkmalsbasierte Verfahren bieten eine weitere Möglichkeit zur Analyse von bewegten Bildern. Das primäre Ziel besteht darin, pro Bild der Sequenz die Lage und Geschwindigkeit der verwendeten Objektprimitiven zu bestimmen. Im Gegensatz zu intensitätsbasierten Ansätzen, welche die Folgebilder Bildpunkt für Bildpunkt vergleichen, werden hier Bildmerkmale auf höherer *Ebene* zur Korrespondenzermittlung herangezogen. Die Wahl der verwendeten Bildprimitiven erfolgt generell entsprechend den Formeigenschaften der zu verfolgenden Objekte. Neben einfachen Objektmerkmalen (z.B. *Ecken* und *Linien*) werden auch häufig komplexere geometrische Primitiven, wie beispielsweise starre oder flexible Konturmodelle [Blake 98], verwendet. Ein wesentlicher Nachteil dieser Verfahren ist die große Anzahl von Bildprimitiven, die öfter zu Mehrdeutigkeiten bei der Korrespondenzermittlung führen. Daher ist meist eine vollständige Suche nach Bildprimitiven im Bild uneffizient, da sich beispielsweise bei einer durchschnittlichen Anzahl von nur 60 Bildprimitiven im Bild schon ( $2^{60} \times 2^{60}$ ) verschiedene Möglichkeiten der Zuordnung ergeben und somit die Wahrscheinlichkeit für ein unvermeidliches Korrespondenzproblem aufgrund der Mehrdeutigkeit signifikant steigt.

Durch eine Merkmalsreduzierung bzw. -selektion kann eine Minimierung der Mehrdeutigkeit erzielt werden, in dem man bspw. eine Scatter-Matrix<sup>27</sup> verwendet [Theodoridis 03], die wiederum eine Auskunft über die Merkmalsverteilung innerhalb oder zwischen den Klassen angibt. Mit Hilfe der Scatter-Matrix können Abhängigkeiten zwischen den Merkmalen aufgedeckt werden. Diese erfordert allerdings die Entwicklung eines robusten Ansatzes, der nicht nur feste gewählte Objekte, sondern auch beliebige veränderliche Objekte in Bildfolgen automatisch segmentiert und die Merkmale in ihrem zeitlichen Kontext analysiert. Dadurch wird eine fundamentale Annahme zur Zuordnung und somit zum anschließenden Tracking von sich bewegenden Objekten in Bildsequenzen abgeleitet. Allerdings sind beim Tracking viele Probleme zu lösen, bis sich die bekannten merkmalsbasierten Analyseverfahren vollautomatisch anwenden lassen. Dies hat einfache Gründe:

- Eine typische Verkehrssituation beinhaltet beispielsweise Teilverdeckung und Beleuchtungsänderung sowie Schattenwurf, die zur Verschmelzung von Objekten oder zur Aufspaltung eines Objektes in zwei (oder mehrere) Teile führen können. Diese verursachen ein Korrespondenzproblem und repräsentieren somit bis dato Herausforderungen im Bereich der Analyse von bewegten Objekten.
- Bewegungen verdeckter Objekte bzw. Objektteile lassen sich oft nicht oder nur mit erheblichen Einschränkungen statistisch rekonstruieren. Die damit verbundenen unterschiedlichen Trajektorien von Merkmalen bzw. die erneute Vereinigung von Merkmalen zu einem Cluster ist ein weiteres Problem von herkömmlichen Verfahren.

Daher wird in diesem Kapitel ein neues Matching-Paradigma auf der Grundlage von Farbinformation entwickelt, welches durch eine Merkmalsextrahierung und Merkmals-

---

<sup>27</sup> Die Within-Class-Scatter-Matrix gibt dabei Auskunft über die Merkmalsverteilung innerhalb einer jeden Klasse. Die Between-Class-Scatter-Matrix beschreibt die Merkmalsverteilung zwischen den Klassen. Mit Hilfe der Mixture-Class-Scatter-Matrix können Abhängigkeiten zwischen den Merkmalen aufgedeckt werden.

zuordnung eine hohe Effizienz bei gleichzeitiger Robustheit in verschiedenen Situationen gestörter Bildinhalte erreicht. Die dadurch erreichte Robustheit wird zum einen aus der Anwendung eines bewegungsbasierten Ansatzes zur automatischen Erfassung von bewegten Regionen erreicht, welcher eine starke Reduzierung der Eingangsdaten auf relevante Bildinhalte ermöglicht, und zum anderen aus der optimalen Nutzung von Farbinformationen, die es in Kombination mit einem geeigneten Farbsegmentierungsverfahren ermöglichen, effizient hochwertige Bildprimitiven zu ermitteln und diese sehr zuverlässig zu korrelieren. Die Korrelation von Bildprimitiven wird hier hierarchischen erfolgen, welche erstens zur signifikanten Verbesserung der Matchinggüte führt und zweitens zur Lösung des Korrespondenzproblems in gestörten Bildsituationen beiträgt. Dieses Konzept des hierarchischen Matchings segmentiert, verfolgt und erkennt nicht nur feste gewählte Regionen, sondern auch beliebige *veränderliche* Objekte in Bildfolgen auf Grundlage kleinerer Regionen automatisch [Al-Hamadi 04a, 04b, 05e, 06b]. Aufgrund der Anzahl der erfassten Merkmale werden wir hier ein Multi-Matching<sup>28</sup> einfügen, das sich von dem klassischen Matching durch seine hierarchische Struktur und Adaptivität unterscheidet.

#### 3.1 Klassisches Multi-Matching

In der Literatur wurde der Begriff des Multi-Matchings auf der Basis von Objektregionen geprägt [Rehmann 98b, Melzer 02, Al-Hamadi 03a, 3d, 3b, 3e]. In diesem Konzept besteht der Ansatz zur Lösung des Korrespondenzproblems in der Zuordnung von Objektsegmenten. Dazu erfolgt erst eine vollständige Farbsegmentierung der Einzelbilder der untersuchten Sequenz. Anschließend werden auf Grundlage der Farbsegmente Merkmale ermittelt, um Korrespondenz durch regionsbasiertes Matching herauszufinden. Da die angewendeten Farbsegmente ausgeprägte Charakteristiken beinhalten, wird eine signifikante Verbesserung der Matchingqualität erwartet. Allerdings hängt die Verbesserung von der Segmentbeschreibung, der zeitlichen Segmentstabilität wie auch von der Kombination der Ähnlichkeitsmaße der jeweiligen Merkmale ab. Da die angewandten Farbregionen oder die Primitiven die Grundlage des klassischen und des hier entwickelten merkmalsorientierten Matchings darstellen, motiviert dies eine genauere Betrachtung der Eigenschaften der gebräuchlichen Farbsegmentierungsalgorithmen.

##### 3.1.1 Farbsegmentierungsalgorithmen

Zur Gewinnung der Farbsegmente gibt es eine Vielzahl von Segmentierungsverfahren, die flächenorientiert arbeiten. Sie lassen sich in lokale (Split-and-Merge) und globale Verfahren (Recursive-Histogram-Splitting) unterteilen. Das erste hat den Vorteil, dass lokale Bildzusammenhänge erhalten bleiben und somit kann auf diese Weise die Farbinformation effektiv zur räumlichen Primitivgewinnung genutzt werden. Während die globalen Verfahren statistische Daten über das ganze Bild in Form von Histogrammen sammeln, betrachten die lokalen Verfahren Zellen, Pixel oder kleine Regionen und lassen diese durch sequentielle Verschmelzung mit benachbarten ähnlichen Regionen wachsen. Obwohl die histogrammbasierten Verfahren eine globale Sicht auf die Daten

---

<sup>28</sup> Es gibt mittlerweile einige Arbeiten, die sich an der dort vorgeschlagenen Methode orientieren [Rehmann 98, Melzer 02].

reflektieren, bieten sie keine räumliche Regionsbeschreibung im Bild. Darüber hinaus überlappen die Verteilungen der Einzelregionen in komplexen Szenen und dadurch können Cluster von Regionen verdeckt sein. Somit sind die Entscheidungen bezüglich der Clusterdefinition anfällig für Fehler. Allerdings liefern die globalen Verfahren wie bei Recursive-Histogram-Splitting (RHS) für einfache gesättigte Bilder gute Ergebnisse. Die lokalen Verfahren wie z.B. Split-and-Merge (S&M) lassen sich einfach und elegant implementieren und haben auch eine weite Verbreitung gefunden. Hier findet die Kombination einer Aufteilung und Verschmelzung von Regionen statt. Solche Verfahren arbeiten für gewöhnlich auf einer Quadtree-Datenstruktur. Die Aufteilung findet an allen inneren Knoten der Baumstruktur statt. Die Verschmelzung der homogene Quadranten/Regionen erfolgt mittels eines Homogenitätskriteriums. Über die definierten Schwellwertparameter für Aufteilung und Verschmelzung kann die Segmentierung sehr genau gesteuert werden. Die lokalen Verfahren besitzen ein gutes Laufzeitverhalten verglichen mit dem globalen Verfahren. Die Schwächen liegen in den ungenauen, blockähnlichen Grenzen der Segmente und der Abhängigkeit von der festgelegten Schwelle zur Aufteilung. Wählt man beispielsweise die Schwelle der Aufteilung großzügig, wird die Laufzeit dadurch stark absinken (aber die Begrenzungen der Segmente werden blockähnlich).

Beim globalen Verfahren (z.B. *Recursive Histogram Splitting*) wird in einem Top-Down-Verfahren das Bild in disjunkte Regionen unterteilt. Diese Regionen werden dann, falls sinnvoll, weiter zerlegt. Dies hängt davon ab, ob ein Histogramm der einzelnen Farbkanäle der Region I einen deutlichen Peak besitzt. Ist dies der Fall, so wird ein Intervall ( $j$ ) um diesen Peak berechnet. Anschließend wird die Region in zwei Klassen gruppiert. Wenn eine Klasse aus einigen größeren Subregionen besteht, so wird I in diese zergliedert. Dies wird so lange wiederholt, bis man keine weiteren Teilregionen findet. Der Hauptnachteil von diesem Verfahren ist, dass es ein rein globales Top-Down-Verfahren ist, ohne jeglichen Mergeanteil (Zusammenfügen von ähnlichen Regionen). Das hat zur Folge, dass die Segmente „ausgefranst“ aussehen.

#### 3.1.1.1 *Color-Structure-Code-Algorithmus*

Der CSC-Algorithmus ist ein erweitertes Regionenwachstumsverfahren, welches die vorteilhaften Eigenschaften lokaler und globaler Verfahren verbindet, d.h. Einfachheit und Schnelligkeit mit Robustheit und Qualität [Rehrmann 94, Priese 98, Al-Hamadi 06c, Al-Hamadi 06d, Al-Hamadi 04c]. Beim Regionenwachstum werden in einer *Bottom-Up* Strategie Pixel sukzessiv, entsprechend eines festgelegten Ähnlichkeitskriteriums, zu einer Region zusammengefasst. Häufig auftretende Probleme beim Regionenwachstum sind Verkettungsfehler, die durch lokale Ähnlichkeiten entstehen. Durch wiederholte lokale Ähnlichkeiten kann es dazu kommen, dass entfernte Pixel, die verschiedene Farb- bzw. Grauwerte haben, trotzdem einem Segment zugeordnet werden. Ein weiteres ungünstiges Merkmal ist, dass das Segmentierungsergebnis von der Wahl des Startpunktes und der Reihenfolge der einzelnen Verschmelzungsschritte abhängig ist.

Der CSC-Algorithmus vermeidet derartige Nachteile durch die Verwendung einer hierarchisch hexagonalen Inselstruktur. Dazu wird angenommen, dass alle Pixel in einer hexagonalen Anordnung vorliegen. Dies lässt sich leicht für herkömmliche orthogonale Bilder simulieren (Abb.3-1-1b). Innerhalb einer hexagonalen Struktur besitzt jede Komponente sechs äquidistante Nachbarn. Jedes zweite Pixel pro Reihe und jedes

### 3. MATCHING-PARADIGMA

zweite Pixel pro Zeile repräsentiert im Ausgangsbild eine sog. "Insel der Ebene 0". Diese Insel enthält neben dem Mittelpunkt sechs Nachbarpixel (Abb. 3-1-1a). Daraus folgt für ein Bild der Größe  $N \times N$  eine Anzahl von  $N/4$  Inseln in der Ebene 0, von denen jede genau sieben Pixel enthält. Außerdem gehören aufgrund der Überlappung, alle Pixel außer den Mittelpunkten zu genau zwei Inseln der Ebene 0. Im generellen Fall gilt, dass jede vierte Insel der Ebene  $i$  genau den Mittelpunkt einer Insel der Ebene  $i+1$  repräsentiert, welche wiederum aus sieben Inseln der Ebene  $i$  besteht (Abb. 3-1-1a). Die Effizienz des CSC-Algorithmus beruht auf den zahlreichen Überlappungen der Inseln [Priese 93, Rehrmann 94, Rehrmann 96, Priese 98].

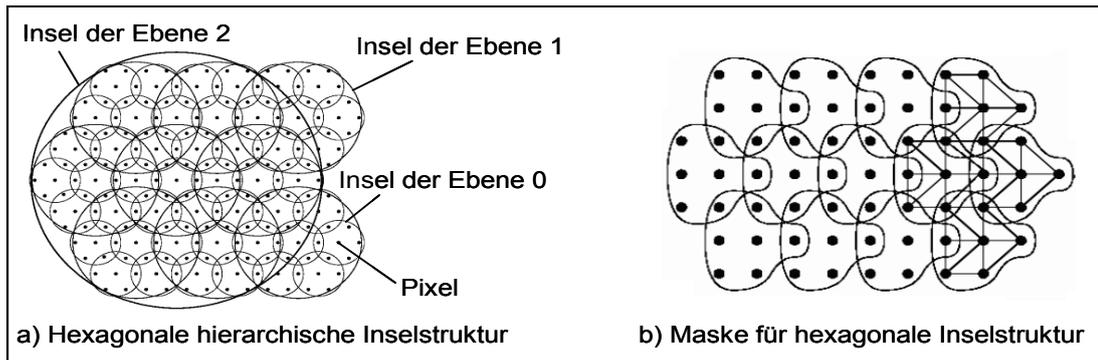


Abbildung 3-1-1: CSC-Inselstruktur [Rehrmann 94]

In der Anfangsphase wird eine Farbsegmentierung auf jeder Insel der Ebene 0, d.h. auf den jeweils sieben Eingangspixeln, durchgeführt. Danach findet die hierarchische Abarbeitung in den einzelnen Ebenen statt, welche in der letzten Ebene  $k$  stoppt, in der die zugehörige Insel das gesamte Bild erfasst. Nach dem CSC-Prinzip erfolgt zuerst eine Verschmelzung von Überlappungsregionen. Diese kann jedoch im Anschluss wieder rückgängig gemacht werden, falls es zur Verfehlung eines bestimmten Farbähnlichkeitskriteriums kommt. Häufig für den CSC verwendete Farbähnlichkeitskriterien sind der Euklidische Abstand im RGB-Farbraum, der den Segmentierungsgrad gut kontrollierbar macht, oder eine von Rehrmann definierte Farbähnlichkeitstabelle, die im FSH-Farbraum<sup>29</sup> arbeitet [Rehrmann 94]. Nach [Rehrmann 94, Priese 98, Al-Hamadi 04c, 05b] ist der CSC-Algorithmus schnell, robust und liefert gute Segmentierungsergebnisse. Leichte Störungen können demnach höchstens in künstlich konstruierten Bildern mit allmählichen vertikalen Farbübergängen entdeckt werden, bei denen die hexagonale Inselstruktur in den Segmentierungen zum Vorschein kommt. Dies liegt an der nicht vorhandenen Möglichkeit vertikale Kanten auf die hexagonalen Masken zu überführen.

**Fakt:** Die vorgestellten Farbsegmentierungsverfahren weisen eine unterschiedliche Eignung zur Erzeugung von Bildprimitiven auf, welche in einem merkmalsorientierten Matching verwendet werden können. Die Anforderungen an die Eigenschaften eines segmentierten Bildprimitives werden durch die Erfordernisse des Matchings bestimmt. Da die Qualität der Primitiven entscheidend für die Funktionsfähigkeit des Matchings ist, ist es sinnvoll, die Farbsegmentierungsverfahren bzgl. der Eignung zur Primitiver-

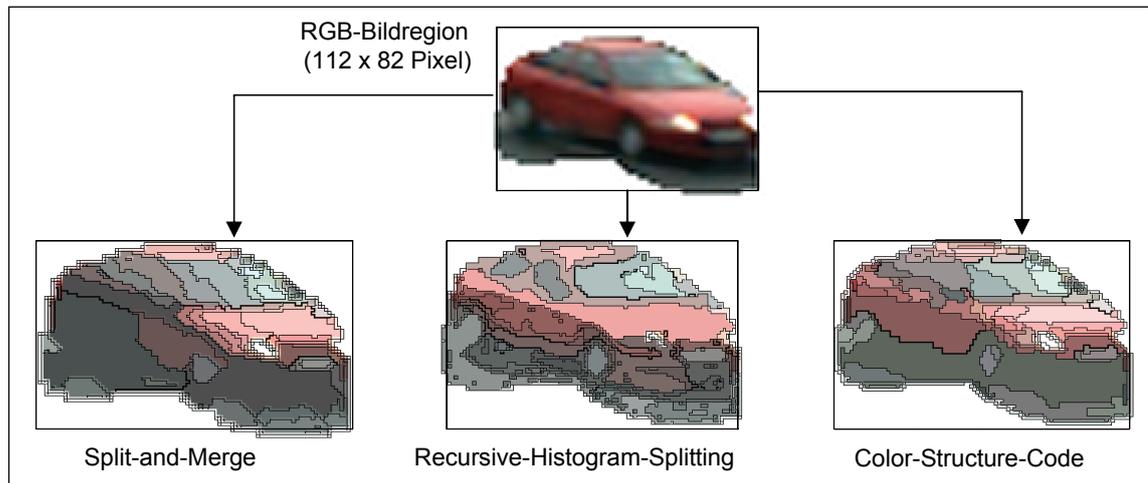
<sup>29</sup> Die Abkürzung steht für Farbton, Sättigung und Helligkeit. Andere Bezeichnungen in der Literatur sind auch HSV, BHS, YHS und HSI. Farbvalenzen, die in diesem pyramidenförmigen Farbraum auf einer Ebene liegen, werden als gleich hell definiert. Daher errechnet sich die Helligkeit aus der Mittelwertbildung der Farbwerte. Die drei Parameter des Farbraumes lassen sich durch eine nichtlineare Transformation von RGB-Werten berechnen.

### 3. MATCHING-PARADIGMA

zeugung zu untersuchen. Ein solcher Test wurde durchgeführt, die Ergebnisse werden im folgenden Abschnitt vorgestellt.

#### 3.1.1.2 Forderungen und Vergleiche an die Segmentierungsverfahren

Eine wichtige Forderung an die vorgestellten Segmentierungsverfahren ist, dass die ermittelten Segmente eine möglichst gleichmäßig runde und wenig zerklüftete Form aufweisen. Eine solche Eigenschaft kann genutzt werden, um stark formentartete Segmentprimitive zu eliminieren, welche meist nicht für ein Matching geeignet sind und daher möglichst selten segmentiert werden sollten. Eine weitere Forderung ist, dass durch ein solches Verfahren bei Anwendung eines bestimmten Segmentierungsschwellwertes vergleichbar große Segmente mit einer Mindestgröße erzeugt werden. Zu diesem Zweck wurde eine Untersuchung durchgeführt, bei der Farbsegmente im Hinblick auf relevante Matchingparameter geprüft worden sind. Zu den verglichenen Verfahren gehören Split-and-Merge, Recursive-Histogram-Splitting und Color-Structure-Code (CSC). Eine Untersuchung zur Eignung der drei Segmentierungsverfahren wurde anhand einer realen, durchschnittlichen Testsequenz durchgeführt. Die Berechnung erfolgte bei einem in etwa vergleichbaren Segmentierungsgrad.



**Abbildung 3-1-2:** Segmentierungsvergleich anhand einer RGB-Bildregion

Zum Vergleich der Segmentierungsverfahren wurde der Anteil der Segmente bestimmt, die für das Matching relevante Eigenschaften aufweisen, d.h. eine bestimmte Kompaktheit ( $K_o$ ) und Mindestgröße  $A$  besitzen. Die folgende Darstellung der Ergebnisse (Tabellen 3-1-1, 3-1-2 und 3-1-3) beinhaltet:

*For1).* gesamte Segmentanzahl

*For2)* die Anzahl der Segmente mit Kompaktheit  $K_{o_{max}} \leq 5$

*For3)* Forderung (*For2*) und Mindestgröße  $A_{min} \geq 10$

*For4)* Anteil der Segmente aus (*For3*) bzgl. der Gesamtanzahl (*For1*)

**Tabelle 3-1-1:** Segmentierung durch Split-and-Merge

	Bild 11	Bild 12	Bild 13	Bild 14	Bild 15	Bild 16
<i>For1</i>	240	233	237	244	257	233
<i>For2</i>	232	227	228	220	233	215
<i>For3</i>	215	209	220	215	212	190
<i>For4</i>	<b>89.6 %</b>	<b>89.7%</b>	<b>92.8%</b>	<b>88.1%</b>	<b>82.4%</b>	<b>81.5%</b>

### 3. MATCHING-PARADIGMA

**Tabelle 3-1-2:** Segmentierung durch Recursive-Histogram-Splitting

	<i>Bild 11</i>	<i>Bild 12</i>	<i>Bild 13</i>	<i>Bild 14</i>	<i>Bild 15</i>	<i>Bild 16</i>
<i>For1</i>	310	280	287	312	295	313
<i>For2</i>	262	233	214	277	245	260
<i>For3</i>	112	91	85	122	140	154
<i>For4</i>	<b>36.1%</b>	<b>32.5%</b>	<b>29.3%</b>	<b>39.1%</b>	<b>47.4%</b>	<b>49.2%</b>

**Tabelle 3-1-3:** Segmentierung durch Color-Structure-Code

	<i>Bild 11</i>	<i>Bild 12</i>	<i>Bild 13</i>	<i>Bild 14</i>	<i>Bild 15</i>	<i>Bild 16</i>
<i>For1</i>	272	274	302	293	311	297
<i>For2</i>	268	265	292	287	294	289
<i>For3</i>	243	245	274	261	269	259
<i>For4</i>	<b>89.3%</b>	<b>89.4%</b>	<b>90.7%</b>	<b>89.1%</b>	<b>86.4%</b>	<b>87.2%</b>

Auskunft über die Matching-Tauglichkeit des jeweiligen Segmentierungsverfahrens gibt der Anteil der Segmente, die den Mindestanforderungen genügen (Tab. 3-1, Zeile *For4*). Tendenziell spiegeln die Ergebnisse der untersuchten realen Testsequenz die visuell erfassbaren Eigenschaften der Segmente in Abbildung 3-1-2 wieder. Demnach haben Segmente, die mit Hilfe des RHS-Algorithmus berechnet wurden, eine überdurchschnittlich ungenügende Kompaktheit. Diese Eigenschaft lässt sich durch die Funktionsweise des Segmentierungsverfahrens erklären. Der RHS-Algorithmus ist ein rein globales Top-Down Verfahren ohne jeden Merge-Anteil [Priese 98]. Ein solches Verfahren hat meist Probleme kleine Regionen zu finden. Weiterhin wurde eine Vielzahl von winzigen pixelförmigen Segmenten ermittelt, die nicht zu einer Region zusammengefasst wurden. Diese sind für die Merkmalsextraktion und somit für das Matching völlig ungeeignet.

Im Gegensatz zu dem RHS-Algorithmus lieferten der S&M- und CSC-Algorithmus bessere Ergebnisse. Somit weisen prinzipiell beide Verfahren eine Tauglichkeit zur Erzeugung von Matching-Primitiven bezüglich der untersuchten Segmente auf. Insbesondere hält sich der Anteil der ermittelten Segmente, die für das Matching ungeeignet sind, in Grenzen. Häufig auftretende Fehler des S&M-Verfahrens, wie zum Beispiel künstliche Kanten an langsamen Farbübergängen in großen Regionen, spielen bei der Anwendung zur Primitiverzeugung aus bewegten Regionen nur eine untergeordnete Rolle. Allerdings ist anzumerken, dass der Berechnungsaufwand des S&M-Algorithmus bedeutend höher ist, als der des CSC. Darüber hinaus entsteht ein Nachteil im Bezug auf die variable Größe der Bildregion, da das S&M-Verfahren im Allgemeinen Bilder der Größe  $2^n \times 2^m$  benötigt. Im Gegensatz dazu kann der CSC-Algorithmus sehr einfach auf beliebige Bildformate angewendet werden und ist echtzeitfähig.

Aus diesen Erwägungen heraus ist dem CSC-Verfahren der Vorzug bei der Zerlegung von bewegten Regionen in kleinen flächenhaften Primitiven gegeben worden. Diese Primitiven werden anschließend als CSC-Patches oder CSC-Segmente bezeichnet. Die Größe der Primitive wird durch den Segmentierungsgrad  $C_{SEG}$  beliebig gesteuert. Ein geringer CSC-Segmentierungsgrad (z.B.  $C_{SEG}=45$ ) führt beispielsweise dazu, dass wenige, dafür aber große CSC-Segmente berechnet werden. Aus der geringen Anzahl von Segmenten folgt, dass nur wenige Korrespondenzen ermittelt werden können. Besonders im Falle kleiner Regionen erhöht die geringe Informationsmenge die

### 3. MATCHING-PARADIGMA

Fehlerwahrscheinlichkeit bei der Bewegungsschätzung. Auch bei größeren Objektregionen führt ein geringer Segmentierungsgrad nur zu einem dünnen Verschiebungsvektorfeld. Die Reduzierung<sup>30</sup> der Segmentgröße bis zu einem gewissen Grad (ca.  $C_{SEG}=10$ ) führt zu einer dichten, dennoch genauen Korrespondenzherstellung. Dabei ist es unerheblich wie groß die zu segmentierende Region ist (Abb. 4-31 a/b,  $C_{SEG}=10$ ). Problematisch ist die Mehrdeutigkeit durch die erhöhte Anzahl von Segmenten.

#### 3.1.2 Korrespondenzanalyse mittels Multimatchs

Nach der Erfassung von Farbsegmenten, die als Grundlage für ein Matching angewendet werden, unterscheidet man zwei Schritte der Korrespondenzherstellung bei herkömmlichen Matchingverfahren. Zum einen beinhaltet es das 1:1 Matching, bei dem nur zwei Farbsegmente miteinander in Beziehung stehen können und zum anderen das so genannte m:n Matching, bei dem mehrere Segmente in Folgebildern einander zugeordnet werden (Abb. 3-1-3). D.h. es werden Korrespondenzen von n-Farbsegmenten des Bildes mit m-Farbsegmenten des Folgebilds ermittelt. Da es wegen des hohen Rechenaufwands nicht möglich ist, für ein Bild alle Kombinationen von Segmentcluster zu bilden, werden nur Cluster aus Segmenten gebildet, die beispielsweise farbähnlich sind.

In der Regel wird zuerst versucht, den 1:1 Matching vorab zu behandeln, um so die Anzahl der Farbsegmente, aus denen für m:n Matching Segmentcluster gebildet werden, zu reduzieren. Segmente mit höherer Matchinggüte beim 1:1 werden nicht mehr zu Segmentclustern eingezogen und bleiben beim weiteren n:m Matching unberücksichtigt. Die Entscheidung über die Zuordnung von m-Farbsegmenten eines Bildes zu n-Segmenten des nächsten Bildes wird durch eine Ähnlichkeitsfunktion realisiert, die auf den verschiedenen Merkmalen basiert.

In der Literatur wird eine Vielzahl solcher Merkmale, wie etwa Farbe, Fläche, Schwerpunkt, Proportion und Orientierung der äquivalenten Ellipse sowie kleinstes umschließendes Rechteck, beschrieben und angewendet [Rehmann 98, Melzer 02, Ross 03, Al-Hamadi 05c, Al-Hamadi 06b]. Daraus werden die Ähnlichkeitswerte für Segmentpaare beim Matching bestimmt. Diese ergeben sich aus der gewichteten Summe der Ähnlichkeiten der einzelnen Merkmale.

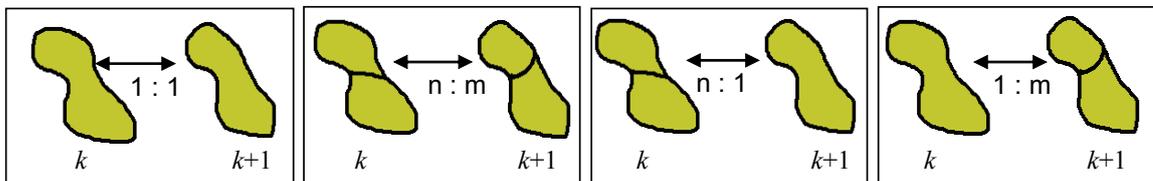


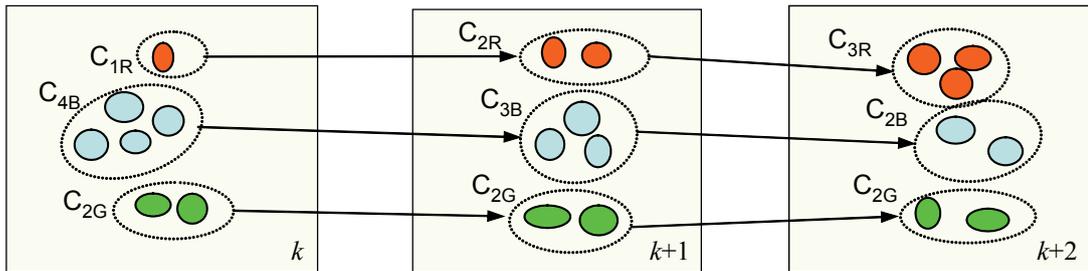
Abbildung 3-1-3: Fälle der Zuordnung beim Multi-Matching nach [Rehmann 98b]

Während des eigentlichen Matching-Prozesses, d.h. der Auswertung der Ähnlichkeitswerte aller Farbsegmente, erwächst bei der m:n Zuordnung schnell ein kombinatorisches Problem, welches nur durch eine geeignete Vorauswahl der zu korrelierenden Segmente in den Griff zu bekommen ist. Normalerweise versucht man das Problem zu lösen, indem die Auswahl der möglichen Matching-Partner durch Eigenschaften, wie

<sup>30</sup> Die Verkleinerung der Segmentgröße kann jedoch nicht weiter fortgesetzt werden, da diese eine Zuordnung im ikonischen Pixelbereich nahe kommen würde, für die ein angestrebtes Multi-Matching auf Merkmalebene nicht mehr zu gewährleisten ist.

### 3. MATCHING-PARADIGMA

topologische Nachbarschaft und Farbähnlichkeit, eingeschränkt wird. Dazu werden farbähnliche Segmente in Clustern zusammengefasst (Abb. 3-1-2). Die Segmente innerhalb der Cluster werden dann mit Hilfe der Merkmalsähnlichkeit weiter untersucht und entsprechend zugeordnet. Ein entscheidendes Problem bei diesem Ansatz ist die hohe Anzahl von Bildprimitiven, die in den Matching-Prozess eingehen, da dieser auf dem gesamten Bild operiert. Dieser Ansatz hat nicht nur einen hohen Rechenaufwand zur Folge, sondern dadurch wird auch die Wahrscheinlichkeit des Auftretens von Mehrdeutigkeiten und somit das Auftreten eines Korrespondenzproblems signifikant erhöht. Mit dem Auftreten vieler verschiedener und unabhängiger Objekte kommt es mitunter zur Interaktion zwischen diesen Objekten.



**Abbildung 3-1-4:** Clusterung benachbarter, farbähnlicher Segmente zur Reduzierung des Matching-Aufwandes (Index deutet auf die Anzahl der Farbkomponenten hin).

Unter diesen erwähnten Umständen sind die ermittelten Trajektorien kurz und es kann hier der zeitliche Kontext nicht optimal zum Erhalt der Objektrajektorien berücksichtigt werden. Dies ist wichtig, um Aussagen zu treffen, wann die Objekte verschmolzen oder zerfallen sind. Somit erfüllt dieses Verfahren die Forderung nach Robustheit und Flexibilität beim Tracking nur zum Teil. Dies motiviert die Entwicklung eines robusten Multi-Matching-Verfahrens, das nicht nur feste gewählte Regionen sondern auch beliebige veränderliche Objekte in Bildfolgen automatisch segmentiert, verfolgt und erkennt. Den Ausgangspunkt für dieses Verfahren bildet das Multi-Matching-Paradigma.

### 3.2 Ziele und Einordnung der eigenen Arbeiten

Mit Hinblick auf ein statisches Überwachungssystem von Verkehrs- und Fußgänger szenarien wird von einem statischen Aufnahmesystem<sup>31</sup> ausgegangen. Die Szenarien lassen demnach die Annahme zu, dass die Objekte einer realistischen Szene ein kontinuierliches Bewegungsverhalten aufweisen. Die Objekte können also nur begrenzte Distanzen zwischen zwei Aufnahmezeitpunkten zurücklegen bzw. keine ruckartigen Richtungs- und Geschwindigkeitsänderungen vornehmen.

Da das Ziel die Anwendung auf reale Szenarien ist, kann jedoch auf bestimmte Ereignisse kein Einfluss genommen werden. Dies bezieht sich sowohl auf die Umgebung als auch auf die Objekte der Szenarien, sodass auf bestimmtes Apriori-Wissen verzichtet werden muss. Weiterhin bringen diese realen Szenarien eine Vielzahl von verschiedenen Objekten mit sich. Die Anzahl der Objekte ist weder konstant noch

<sup>31</sup> Die Aufnahmegeschwindigkeit liegt bei 20-25 Bildern pro Sekunde, sodass die Bewegungen als flüssig wahrgenommen werden.

### 3. MATCHING-PARADIGMA

---

bekannt. Ebenso wenig bekannt ist ihr Eintreten und Verlassen der Szene. Besonders Außenszenen unterliegen Bedingungen wie Wetter oder Tageszeit, die nicht-kooperative Messsituationen herbeiführen können. So kann beispielsweise in verschiedenen Lichtsituationen das Farbauftreten der Objekte erheblich differieren.

Für die Auswertung solcher Szenarien ergeben sich somit vorhersehbare Problemsituationen, die in der Entwicklung eines Verfolgungssystems beachtet werden müssen. Mit dem Auftreten vieler verschiedener und unabhängiger Objekte kommt es mitunter zur Interaktion zwischen diesen. Objekte werden sich überholen, in ihrem Weg kreuzen, sich teilweise oder gar vollständig verdecken. Problematisch sind auch zeitweise unbewegte Objekte, die im Schritt der Bewegungsanalyse nicht erkannt und erfasst werden. Zur Verfolgung der Objekte werden geeignete Objektbeschreibungen benötigt. Diese Gegebenheiten der zu analysierenden Szenen müssen beachtet und verarbeitet werden, um eine stabile Multi-Objektverfolgung und die Ermittlung und den Erhalt der Objekttrajektorien zu erreichen.

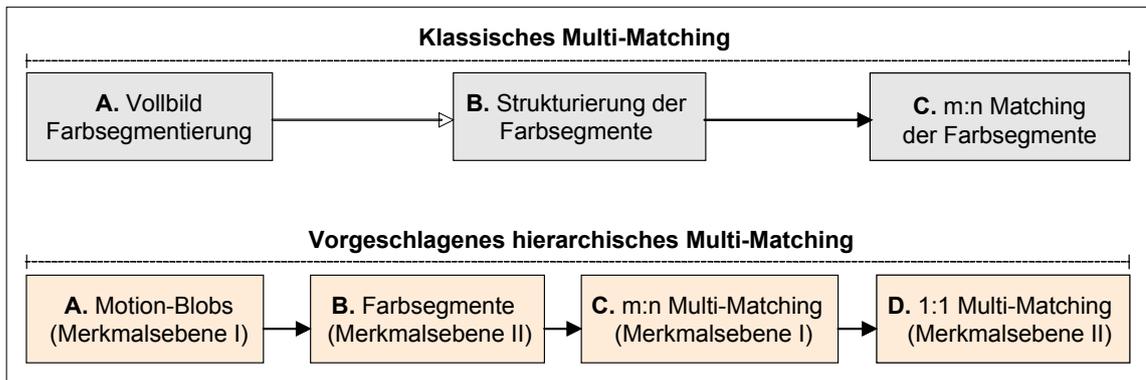
Das Ziel des hier entwickelten Verfahrens zum Multi-Matching ist die Anwendung auf Multi-Objektszenen, in denen mehrere Objekte Verfolgungsgegenstand sind und diese auch in mehrdeutige Situationen miteinander treten können. Hierfür lassen sich drei Aufgaben formulieren:

- 1). In diesem ersten Schritt sollen die Objekte unabhängig von ihrer Art und Form einfach durch ihre Bewegung erfasst werden.
- 2). Anschließend müssen Merkmale aus den erfassten Objektkandidaten extrahiert werden, die der Objektbeschreibung dienen. Dies muss losgelöst von der Bildebene geschehen, damit mögliche Formänderungen der Objekte nicht zu Irritationen führen.
- 3). Für die Anwendung auf Multi-Objektszenen muss im letzten Schritt eine geeignete Verarbeitung der umfangreichen Informationen aufgestellt werden, um vor allem die erfassten Merkmale zwischen den Objekten zeitlich zu zuordnen und somit ein stabiles zweistufiges Multi-Matching zuzulassen.

Diese Teilaufgaben stellen daher das Hauptaugenmerk in diesem Kapitel dar, in dem die Korrespondenzanalyse durch ein neues Paradigma für Multi-Matching erfolgt. Dafür wird in diesem Kapitel das Prinzip des herkömmlichen Multi-Matchings abgewandelt, in dem es erstens durch eine Bewegungssegmentierung ergänzt wird und zweitens durch eine hierarchische Merkmalsextrahierung und –zuordnung erweitert wird. Die Abbildung 3-2-1 zeigt ein vereinfachtes Schema, das den Ablauf der beiden Verfahren gegenüberstellt.

Wie in Abbildung 3-2-1 zu erkennen ist, wird bei der vorgeschlagenen Variante eine andere Strategie verfolgt. Es erfolgt eine Extrahierung der Bildmerkmale in zwei Schritten mit der Intention, eine möglichst frühzeitige Reduzierung des Datenmaterials zu erreichen. Diese Reduzierung geschieht durch die Segmentierung von ausgewählten Bildregionen, welche mit Hilfe einer modifizierten Differenzbildtechnik in Verbindung mit einem Nachverarbeitungsansatz, gewonnen werden (Schritt A). Die Differenzbildregionen stellen die erste Merkmalsebene (so genannte Motion-Blobs) dar, auf deren Grundlage eine weitere Zerlegung in kleinere Segmente durch eine Farbsegmentierung durchgeführt wird. Die so ermittelten Farbsegmente repräsentieren die zweite Merkmalsebene (Schritt B).

### 3. MATCHING-PARADIGMA



**Abbildung 3-2-1:** Struktur des entwickelten merkmalsbasierten Verfahrens zum Multi-Matching

Der Aufwand des anschließenden Matchings wird durch die Zweiteilung stark reduziert, ebenso die Gefahr des Auftretens von Mehrdeutigkeiten. Die Zuordnung der Motion-Blobs erfolgt hierbei durch ein m:n Multi-Matching (Schritt C). Auf dieser Grundlage werden schließlich die, durch einen hohen Segmentierungsgrad, sehr kleinen Farbsegmente der zweiten Merkmalsebene durch eine einfache eindeutige Zuordnung, d.h. 1:1 Matching, korreliert (Schritt D). Im Folgenden werden die Schritte A-D, welche das entwickelte zweistufige Multi-Matching-Paradigma realisieren, detailliert beschrieben.

### 3.3 Bewegungsbasierter Segmentierungsansatz

Merkmalsbasierte Verfahren der Bewegungsanalyse setzen eine robuste Segmentierung<sup>32</sup> von sich bewegenden Objekten in der Sequenz voraus, wenn nicht komplette Bewegungsvektorfelder bestimmt werden sollen. Im Vergleich zu den in Abschnitt 2.1 erwähnten Methoden erscheint der einfache differenzbildbasierte Ansatz (D-Ansatz) zur Extraktion von bewegten Objekten als besonders interessant, da hier nicht nur die Bewegung in einer Bildsequenz detektiert wird, sondern auch eine Reduktion der Datenmenge erfolgt. Daher stellt der D-Ansatz eine bewährte Grundlage dar und ist aufgrund ihrer Einfachheit und Schnelligkeit ein oft angewandtes Verfahren zur Bewegungssegmentierung, für das in dieser Arbeit eine spezielle Modifikation entwickelt wurde. Allerdings muss bei den differenzbildbasierten Ansätzen zwischen einer Subtraktion zweier Folgebilder innerhalb einer Sequenz und eines modellierten Hintergrundbildes unterschieden werden [Neri 98]. Ein komplettes Hintergrundbild steht jedoch oft nicht zur Verfügung, da es nicht immer generiert werden kann, wie zum Beispiel in kurzen Bildsequenzen mit vielen, dicht nebeneinander befindlichen Objekten. In solchen Situationen ist die Subtraktion der Folgebilder die einzig verbleibende Möglichkeit, um mit Hilfe der Differenzbildtechnik eine Veränderung zu detektieren.

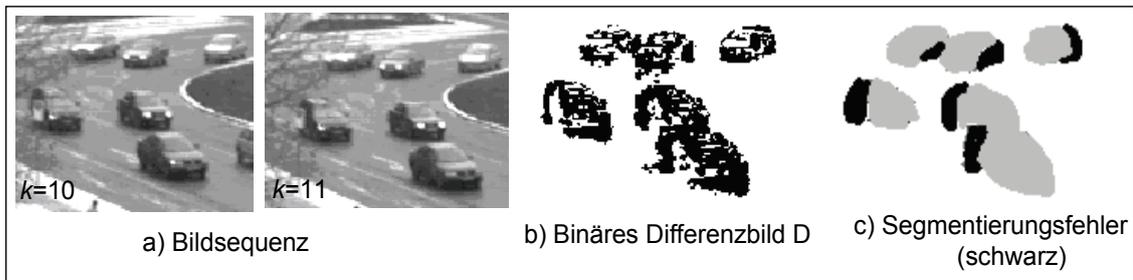
Die einfache pixelweise Subtraktion zweier aufeinander folgender Grauwertbilder  $I^k$  und  $I^{k-1}$  liefert bei Filterung mittels eines Schwellwertes  $H$  die Unterschiede zwischen beiden Bildern in Form eines binären Differenzbildes  $D$  (s. Abb. 3-3-1a). Der Schwellwert  $H$  ist generell von der verwendeten Bildsequenz abhängig und kann daher variieren.

<sup>32</sup> Generell kann dieser Vorgang der Segmentierung als Zuordnungsprozess verstanden werden, in dem ausgewählte Bilddaten zu Objekten zugeordnet werden.

### 3. MATCHING-PARADIGMA

$$D^k(x, y) = \begin{cases} 1; & \text{falls } |I^k(x, y) - I^{k-1}(x, y)| > H \\ 0; & \text{sonst} \end{cases} \quad (3-3-1a)$$

Problematisch an einem solchen Segmentierungsansatz ist jedoch, dass der Umkehrschluss, dass eine Veränderung des Bildes einer Objektbewegung entspricht, nicht unbedingt gilt. Auch nicht-kooperative Messsituationen, wie etwa Beleuchtungsänderungen, Schattenwurf oder Rauschen, können Veränderungen am Bildmaterial hervorrufen. Eine einfache Subtraktion hat auch den offensichtlichen Nachteil, dass nicht nur die Regionen erfasst werden, an denen sich Objekte bewegen, sondern auch solche, an denen der Hintergrund nach der Objektfortbewegung wieder frei liegt. Dies führt zu einem unnötigen Segmentierungsfehler (so genannte Ghosting-Effekt) (siehe Abb. 3-3-1c) und dadurch werden die Grenzen des Objektes wie auch zwischen den Objekten nicht sauber erfasst. Daher ist mit einem bewegungsbasierten Segmentierungsansatz ein gewisser Zusatzaufwand für eine eventuell notwendige nachträgliche Korrektur verbunden (Nachbearbeitung). Dies motiviert den Schritt zu einer modifizierten, erweiterten Differenzbildgewinnung (so genannten MD-Ansatz).



**Abbildung 3-3-1:** Einfache Differenzbildtechnik. Dadurch werden die Grenzen des Objektes wie auch zwischen den Objekten aufgrund der Segmentierungsfehler (Ghosting-Effekt) nicht sauber erfasst.

#### 3.3.1 MD-Ansatz mit Nachbearbeitung

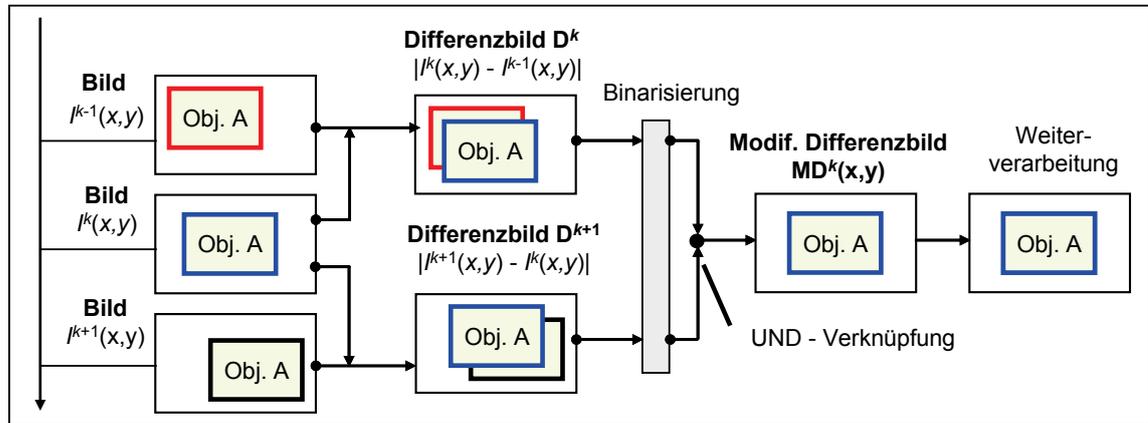
Die genaue Erfassung von Regionen, welche bewegte Objekte beinhalten, kann durch eine modifizierte Variante der Differenzbildtechnik erreicht werden. Statt der zeitlichen Ableitung wird die Berechnung von zwei zeitlich unmittelbar aufeinander folgenden Differenzbildern (DI) verwendet. Diese werden binarisiert. Danach erfolgt eine „*UND-Verknüpfung*“ dieser binarisierten Differenzbilder, um die bewegten Objekte für das mittlere Bild gemäß Abbildung 3-3-2 zu erhalten. Eine Verknüpfung kann auch durch eine Multiplikation<sup>33</sup> ersetzt werden [Al-Hamadi 03a, 04c]. Um eine glatte Steigung an den Objektkanten zu erhalten, wäre auch hier eine Glättung des Bildes zuerst vom Vorteil. Eine relativ sichere Erfassung der bewegten Objekte gegenüber dem Hintergrund ist so möglich.

$$MD^k(x, y) = (D^{k+1}(x, y)) \wedge (D^k(x, y)) \quad (3-3-1b)$$

<sup>33</sup>In diesem Fall ist darauf hinzuweisen, dass es von Vorteil ist, das Bild erst nach Durchführung der Multiplikation und nicht früher zu binarisieren. Dann werden dadurch Regionen mit schwachen Kanten und Helligkeitsübergänge, die sich schnell bewegen oder Objekten mit starken Kanten, die sich langsam bewegen, trotzdem berücksichtigt.

### 3. MATCHING-PARADIGMA

Dieser *modifizierte* differenzbildbasierte Ansatz (MD-Ansatz) unterscheidet sich von anderen Segmentierungsverfahren von bewegten Objekten (Differenzbildtechnik oder optischer Fluss), da er einerseits *einfach* und somit nicht *rechenaufwendig* ist und andererseits eine starke *Datenreduzierung* und *genaue* Objektbegrenzung bei ausreichenden Objekttextur liefert. Zudem wird mit diesem Ansatz der Ghosting-Effekt kompensiert.



**Abbildung 3-3-2:** Das Prinzip des modifizierten differenzbildbasierten Ansatzes zur Erfassung von bewegenden Bildregionen anhand einer realen Bildsequenz.

Allerdings ist ein *wesentliches Problem* mit diesem Ansatz noch nicht gelöst. Hat ein Punkt eines bewegten Objektes zufällig den gleichen Grauwert wie der stationäre Hintergrund oder weisen diese Punkte an dieser Stelle keinen texturierten Vordergrund auf, wird die Veränderung nicht detektiert. Diese Bildpunkte fehlen im Differenzbild, sodass bei einer *Binarisierung Löcher* entstehen (siehe Abb. 3-3-3). Infolgedessen wurde hier eine Weiterverarbeitungsstufe entwickelt, um verbleibende Störungen zu beseitigen und Regionen an denen eine Veränderung stattgefunden hat, hervorzuheben. Diese Nachbearbeitungsstufe basiert erstens auf morphologischen Operationen<sup>34</sup> und erreicht so eine Auffüllung der Löcher im resultierenden modifizierten Differenzbild (MD). Dafür wird hier ein Strukturelements  $M$  verwendet. Da die Qualität der Verarbeitung von der Größe des Strukturelements abhängt, stellt sich die Frage, wie groß das Strukturelement sein muss. Um dieses Problem zu lösen, wird eine Heuristik angewandt, mit der die getrennten Bereiche des Differenzbildes wieder zusammengefasst werden. In unserem Fall haben wir die optimale Einstellung der Größe des Strukturelements nach folgenden Schritten erreicht:

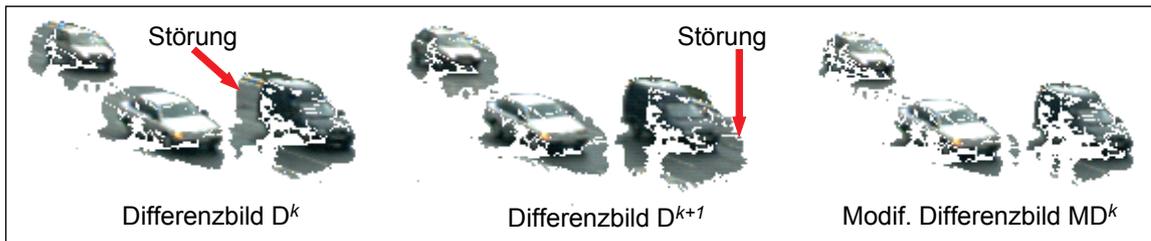
- Operation „Closing“ mit einem größeren quadratischen Strukturelement ( $W=7$ ), um Bereiche mit viel Bewegung miteinander zu verbinden
- Operation „Opening“ mit einem etwas kleinerem Strukturelement ( $W=3$ ), um nah liegende Objekte voneinander zu trennen

Das Ergebnis des Bearbeitungsprozesses ist eine binäre Maske  $M$ , welche Bildregionen beschreibt, die potenzielle Objektkandidaten enthalten. Durch die Anwendung von morphologischen Operatoren zur Unterdrückung von verbleibenden Störungen, kann es

<sup>34</sup> Die morphologischen Operationen werden auf ein modifiziertes Differenzbild (MD) angewandt [Al-Hamadi 03]. Generell muss für die Anwendung morphologischer Operationen aber ein geeignetes Strukturelement definiert werden.

### 3. MATCHING-PARADIGMA

vorkommen, dass auch die Kontur einer sich bewegenden Region auseinander gerissen wird. Dünne Verbindungen werden aus dem Differenzbild gelöscht.



**Abbildung 3-3-3:** Beispiel aus einer realen Bildsequenz für die modifizierte Differenzbildtechnik (Binärmasken in RGB).

Eine besondere Beachtung verdient die entwickelte Funktion zur konturbasierten Erosion, die sich als sinnvolle Ergänzung zur Closing-Operation anbietet. Diese Funktion hat zur Aufgabe, Konturen von Regionen in der binären Maske zu glätten, welche häufig Ausreißer enthalten. Als Nebeneffekt werden gleichzeitig solche Regionen getrennt, die bedingt durch die Closing-Operation an kleinen Übergängen oberflächlich verschmolzen sind. Der Algorithmus bearbeitet jede zusammenhängende binäre Region innerhalb der Maske  $M$ . Dazu werden in einem ersten Schritt alle Konturpunkte  $p_1, \dots, p_n$  einer Region  $R$  in Form von 2D-Koordinaten ermittelt. Im zweiten Schritt bearbeitet der Algorithmus sukzessiv jeden Konturpunkt  $p_i$  und prüft, ob ein anderer Punkt  $p_j$  existiert, der eine Euklidische Distanz zu  $p_i$  hat, die weniger als  $d_{\min}$  beträgt. In diesem Fall werden alle Pixel aus der Maske entfernt, die sich auf der Strecke zwischen  $p_i$  und  $p_j$  befinden.

$$\forall p_i, \exists p_j \wedge \|p_i - p_j\| < d_{\min} \rightarrow \text{entferne alle Maskenpunkte auf der Strecke } \overline{p_i p_j} \text{ und } p_i, p_j \in \{p_1, \dots, p_n\}. \quad (3-3-1c)$$

Die Erosion wird durch den Parameter  $d_{\min}$  gesteuert (Abb. 3-3-4c). Eine weitere Verbesserung der Bewegungsmaske kann durch die Filterung des Hintergrundes erreicht werden. Um den Hintergrund in einer Bildsequenz stückweise zu extrahieren, bietet es sich an, mit jedem neuen Bild all die Pixel in das potentielle Hintergrundbild zu übernehmen, die nicht zur aktuellen Bewegungsmaske gehören (Abb. 3-3-4).

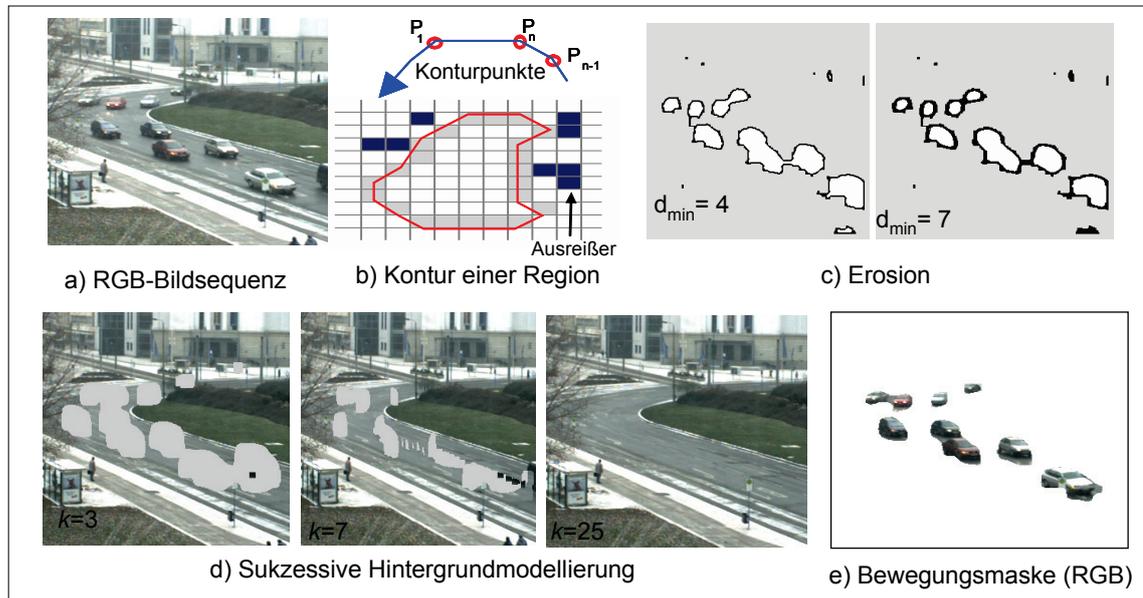
Die Bildpunkte, an denen der Hintergrund bereits ermittelt wurde, können genutzt werden, um die Bewegungsmaske zu optimieren. Dies lässt sich durch eine Differenzbildung mit anschließendem Schwellwertvergleich realisieren. Dazu wird das aktuelle Bild  $P$  an den Punkten der Bewegungsmaske  $M$  mit den Intensitäten des Hintergrundbildes  $HG$  verglichen.

$$\forall p(x, y), |P(x, y) - HG(x, y)| \leq H \rightarrow \text{lösche Punkt } p(x, y) \text{ aus der Bewegungsmaske } M \quad (3-3-1d)$$

Der Schwellwert  $H$  ist im Allgemeinen vom verwendeten Bildmaterial abhängig. Es hat sich jedoch gezeigt, dass dieser Wert in verschiedenen Bildsituationen nicht wesentlich variiert. Die beschriebene Differenzbildtechnik einschließlich der Nachbearbeitung ermöglichen es zuverlässig, Regionen, die bewegte Objekte enthalten, aus Bildfolgen zu

### 3. MATCHING-PARADIGMA

extrahieren (Abb. 3-3-4e). Nun bilden diese Regionen die Objektkandidaten für die weitere Bewegungsanalyse. In der Literatur werden Regionen, die auf der Grundlage einer Veränderung des Bildes segmentiert wurden, häufig als „Motion-Blobs“ bezeichnet. Dort wird die Bewegung der Bildregion über Merkmalszuordnung ermittelt.



**Abbildung 3-3-4:** Das Prinzip der konturbasierten Erosion dazu Konturpunktermittlung und konturbasierte Erosion (a,b, c). Sukzessive Modellierung des Hintergrundes (d) mit einem Beispiel einer Bewegungsmaske (e für RGB-Bildsequenz in a).

#### 3.3.2 Schattenfilterung

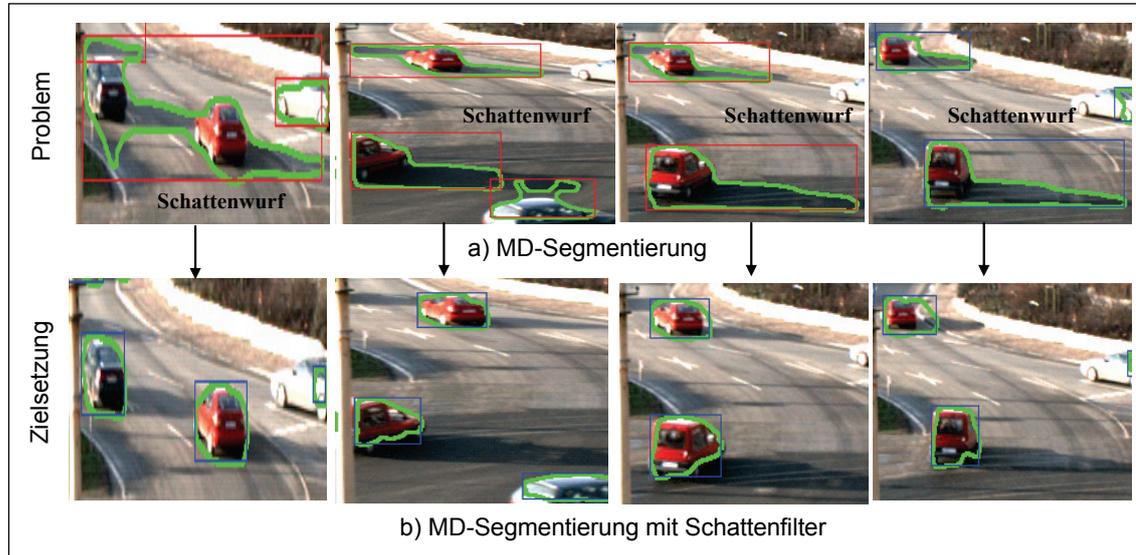
Der im vorhergehenden Abschnitt beschriebene MD-Ansatz arbeitet korrekt in ungestörten Bildaufnahmesituationen. Es ist jedoch anfällig für Fehlsegmentierungen, die durch direkte Schatten (so genannte Schattenwurf)<sup>35</sup> hervorgerufen werden (s. Abb. 3-3-5). Zur Gewährleistung der Robustheit eines Analysesystems ist ein Korrekturschritt zur Beseitigung der Schattenstörung unerlässlich (s. Abb. 3-3-5). Ein Überblick zum Thema Schattensegmentierung sowie ein effektives Filterverfahren für Schatten in Straßenverkehrsaufnahmen, das ebenfalls auf Blob-Regionen arbeitet ist in [Bevilacqua 03, Strauder 99, Prati 01] zu finden. In dem dort beschriebenen Ansatz [Bevilacqua 03] wird bspw. wie auch in den meisten Veröffentlichungen ein Hintergrundbild vorausgesetzt. Diese Einschränkung kann jedoch durch Nutzung von Farbinformationen überwunden werden [Al-Hamadi 05,06].

Hierzu müssen zunächst einige Annahmen über die Eigenschaften von Schatten getroffen werden. In Beobachtungen hat sich gezeigt, dass eine Schattenregion im Allgemeinen eine verminderte Helligkeit bezüglich ihrer Umgebung besitzt und zum überwiegenden Teil am Rand von Blob-Regionen [Al-Hamadi 05d, 06c] auftritt. Weiterhin ist eine Annahme über den Hintergrund des bewegten Objektes zu treffen. In Verkehrsaufnahmen kann hierzu festgestellt werden, dass die Fahrbahn normalerweise eine geringere Farbsättigung als die Fahrzeuge aufweist und bis auf Straßenmarkierun-

<sup>35</sup> Diese Art von Schatten, wird durch die bewegten Objekte selbst verursacht, da das Objekt zwischen der Lichtquelle und dem Bildaufnahmesensor sich befindet. Dieser Schattenwurf ist ein in realen Bildaufnahmesituationen von Verkehrsszenen häufig auftretendes Phänomen, welches die Ergebnisse einer auf der Grundlage der Differenzbildtechnik durchgeführten Bewegungsanalyse stark verfälschen kann.

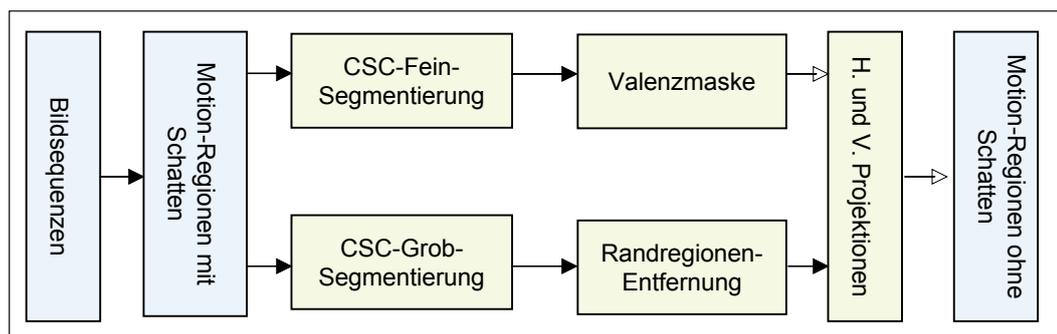
### 3. MATCHING-PARADIGMA

gen nur wenig ausgeprägte Gradienten enthält. In der Analyse von Verkehrsaufnahmen kann diese Information vorteilhaft zur Schattendetektion genutzt werden. Die Struktur des entwickelten Schattenfilters, welche die genannten Annahmen über Schattenregionen praktisch umsetzt, ist in Abb. 3-3-6 dargestellt. Der hier entwickelte Schattenfilter [Al-Hamadi 05d] wird durch eine Farbsegmentierung in zwei unterschiedlich hohen Segmentierungsstufen realisiert.



**Abbildung 3-3-5:** MD-Segmentierungsansatz mit und ohne Schattenfilterung.

Der vorgestellte CSC-Algorithmus eignet sich für diese Aufgabe besonders, da er zum einen die genaue Steuerung der Segmentgröße ermöglicht und zum anderen sehr schnell und echtzeitfähig ist. Blob-Regionen, die Schatten enthalten, die von einem bewegten Fahrzeug auf die Fahrbahn geworfen werden, besitzen an den betroffenen Stellen meist eine geringe Helligkeit und enthalten nur wenig Textur. Durch eine Farbsegmentierung werden in solchen Bereichen verhältnismäßig wenige, jedoch große Segmente erzeugt (Abb. 3-3-6, Abb. 3-3-7a).

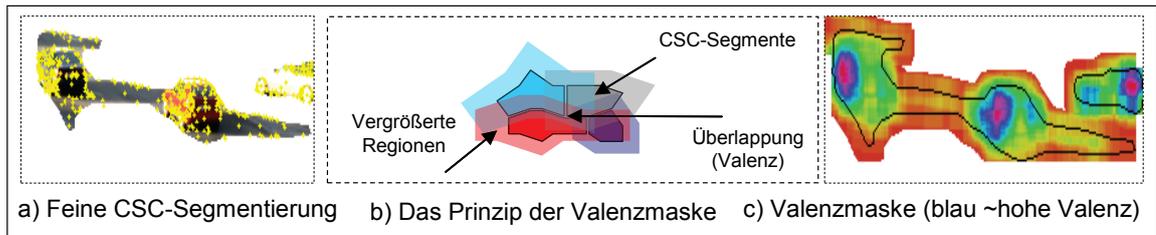


**Abbildung 3-3-6:** Schattentfernungsalgorithmus für Straßenverkehrsszenen auf der Grundlage eines CSC-Algorithmus.

Im Gegenteil dazu werden Blob-Regionen, die Objekte und damit stärkere Gradienten beinhalten, in viele kleine CSC-Segmente zerlegt. Diese Beobachtung kann durch die Erzeugung einer Maske ausgenutzt werden, welche Auskunft über die Verteilung von CSC-Segmenten innerhalb einer Blob-Region gibt. Eine derartige Maske wird als Valenzmaske bezeichnet (s. Abb. 3-3-6b).

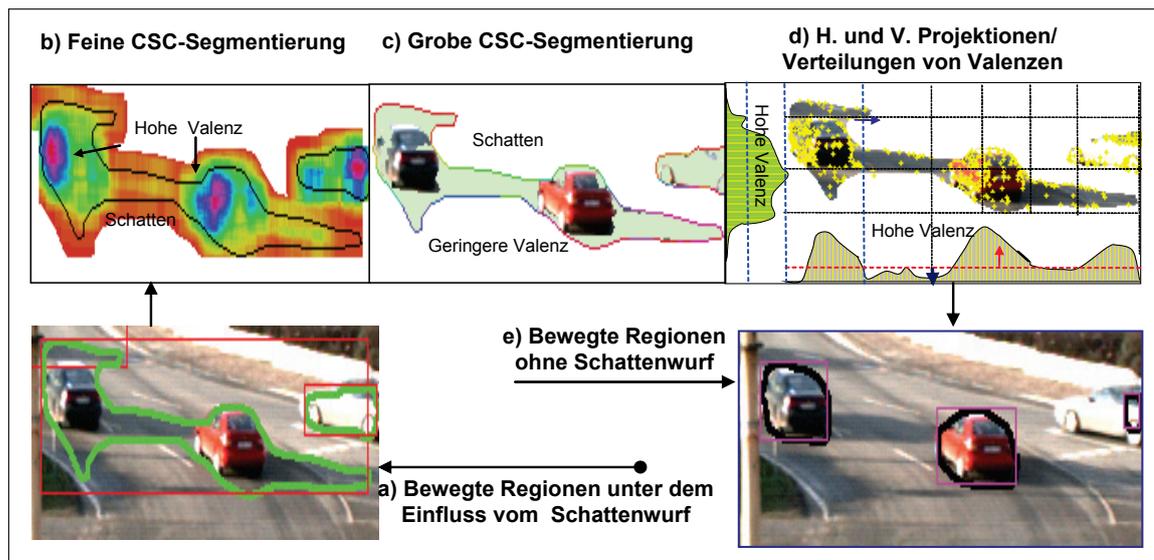
### 3. MATCHING-PARADIGMA

Die Valenzmaske<sup>36</sup> kann einfach erzeugt werden, indem alle CSC-Regionen um ein bestimmtes Maß  $t_{enl}$  vergrößert und ihre gemeinsamen Schnittpunkte aufsummiert werden. Anschließend können durch eine Schwellwertfilterung der Valenzmaskenwerte  $V_a(x, y)$  die Pixel aus der korrespondierenden Blob-Region entfernt werden, für die ein festgelegter Wert  $t_{min}$  unterschritten wurde (Abb. 3-3- c). Die Verwendung fester Schwellwerte ist unproblematisch, da die Werte der Valenzmaske völlig unabhängig von der Blob-Größe sind. Die Werte  $V_a(x, y)$  sind an solchen Stellen hoch, an denen sich Objekte befinden, jedoch gering, falls Schatten oder auch einfacher gradientenarmer Hintergrund, wie z.B. Strasse zugrunde liegt.



**Abbildung 3-3-7:** Ausschließung von Regionen mit schwachen Gradienten

Die Valenzmaske besitzt natürlich auch dort geringe Werte, wo sich Objekte befinden, die aufgrund homogener Bereiche nur in eine kleine Anzahl von CSC-Segmenten aufgeteilt wurden. Dieses Problem wird jedoch durch den zweiten Teil des Algorithmus wieder kompensiert (Abb. 3-3-6) [Al-Hamadi 05d]. Die zu Beginn getroffene Annahme, dass sich Schattenregionen hauptsächlich in der Nähe der Blob-Kontur befinden und damit solche Regionen, die sich im Innern befinden, automatisch zu Objekten gehören, wird hier praktisch umgesetzt. Auch an dieser Stelle bietet sich eine CSC-Segmentierung an, jedoch in einer gröberen Segmentierungsstufe, da dies den Vorteil hat, dass sich potentielle Schatten-Randregionen stärker ausprägen und damit leichter entfernt werden können (Abb. 3-4-8).

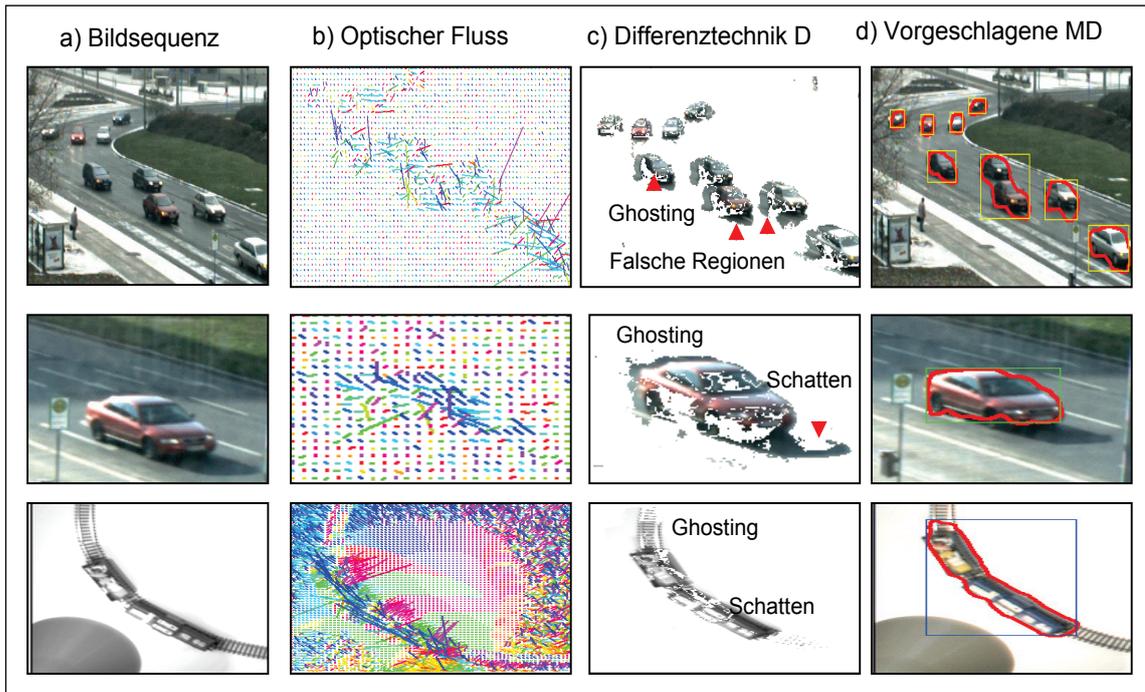


**Abbildung 3-3-8:** Ausschließung potentieller Schattenregionen. Die erfassten Regionen beschreiben die tatsächlichen bewegten Objekten (e).

<sup>36</sup> Wie viele Segmente überschneiden sich bei der Vergrößerung der Schwelle.

### 3. MATCHING-PARADIGMA

Die Extrahierung der inneren Blob-Segmente bzw. der Objektregionen kann leicht realisiert werden, indem aus der Menge aller CSC-Segmente diejenigen entfernt werden, die an die Blob-Kontur angrenzen. Es ist jedoch sinnvoll, Segmente, die einen bestimmten Sättigungswert  $s_{\min}$  überschreiten, beizubehalten, da solche Segmente mit hoher Wahrscheinlichkeit von Objekten und nicht vom Fahrbahnuntergrund stammen (Abb. 3-3-8a,b,c). In Untersuchungen hat sich gezeigt, dass die Kombination der beiden Teilschritte zur Schattenentfernung eine solide Grundlage zur Extrahierung von Objektregionen darstellt (Abb. 3-3-8e) und somit eine fundamentale Stütze für eine weitere Merkmalsextrahierung darstellt.



**Abbildung 3-3-9:** Leistungsfähigkeit des vorgeschlagenen MD-Ansatzes in Verbindung mit dem Nachverarbeitungsalgorithmus im Vergleich mit klassischen Verfahren zur Segmentierung von bewegenden Regionen.

**Diskussion:** Ein wesentlicher Vorteil des vorgeschlagenen MD-Ansatzes ist die relativ sichere und genaue Erfassung der bewegten Objekte gegenüber dem Hintergrund. Dazu unterscheidet sich dieser MD-Ansatz von anderen Segmentierungsverfahren (Abb. 3-3-9), da er einerseits *einfach* und nicht *rechenaufwendig* ist und andererseits eine starke *Datenreduzierung* und *genaue* Objektkonturen (*Begrenzung*) liefert. Ein Ghosting-Effekt wird dadurch kompensiert. Als weiterer Vorteil gilt auch hier die Unterdrückung der verbleibenden Fehler (wie Ausreißer, Homogenität oder Schattenwurf) durch die angewandten Nachverarbeitungsalgorithmen. Diese führen zur Verschmelzung von Objekten und schwierigen Merkmalsextraktionen.

Der Nachteil des MD-Ansatzes ist der Verlust der erfassten bewegten Regionen, wenn keine Bewegung im Bild wie bspw. beim Stau bei dichtem Straßenverkehr oder Ampeln registriert werden kann. Hier können mit Mitteln der Tiefeninformation oder modellbasierten Ansätzen (bspw. Active-Shape-Models) effektive Ergänzungen abgeleitet werden. Auf diese Weise können die zu verfolgenden Objekte unter günstigen Umständen- wenn sie frei vor der Kamera stehen und keine anderen Objekten sich auf der gleichen Tiefenebene befinden- vollständig vom Hintergrund getrennt werden.

Anders als die oben beschriebenen Verfahren ist die Auswertung der Tiefeninformationen für sich bewegende Kameras und dynamische Hintergründe geeignet, was insbesondere für den Einsatz in Fahrerassistenzsysteme notwendig ist. Zusätzlich kann auf Basis der Tiefenkarte die Erzeugung einer Segmentierungsmaske für die erhobenen Objekte erfolgen, die mit zwei Kameras aufgenommen werden. Zur Reduzierung des Rechenaufwandes kann in der Regel die Disparitätsberechnung nur für die erhobenen Kanten erfolgen. In diesem Disparitätshistogramm zeigen sich bei typischen Szenen mehrere Peaks. Durch Auswertung des Histogramms kann nun die Tiefenschwelle zwischen beide Peaks gelegt werden.

Die in der Literatur gefundenen Arbeiten berichten von guten Ergebnissen bei der Verwendung der Hintergrundsubtraktion. Nachteil ist die Notwendigkeit der Modellierung, wobei neben der Rechenintensität auch das Erfordernis der Initialisierung als Problem angesehen werden kann. Hier schafft die modifizierte Differenzbildtechnik Abhilfe. Hintergrundsubtraktion wird häufig verwendet und bringt gute Ergebnisse. Die Differenzbildtechnik findet aufgrund der Einfachheit und Schnelligkeit Anwendung im Bereich der einfachen Überwachungen. Allerdings liefert diese schlechtere Segmentierungsergebnisse (Ghosting-Effekt oder Homogenität).

#### 3.4 Hierarchische Merkmalsextraktion

Durch die Merkmalsextraktion sollen die in den zu erfassenden Bewegungsregionen vorhandenen relevanten Informationen auf einen niedrigdimensionaleren Merkmalsraum abgebildet werden. Dabei sollen redundante und unwichtige Primitiven unterdrückt werden, so dass die extrahierten Merkmale die gesamte relevante Information des Objektes besitzen, um eine genaue Zuordnung in der zweiten Ebene der Bewegungsanalyse zu ermöglichen. Dafür existieren einfache, skalare, geometrische Merkmale, die keine explizite Repräsentation der Regionskontur benötigen und sich direkt aus der Kontur berechnen lassen. Zu ihnen gehören u.a. Umfang, Seitenverhältnis des umschließenden Rechtecks, Kreisförmigkeit, Exzentrizität, usw.

Allerdings gibt es keine Theorie, die besagt, welche Merkmale bei einer bestimmten Anwendung extrahiert werden müssen. Somit ist die allgemeine Vorgehensweise erst viele Merkmale zu extrahieren und danach mittels statistischer Verfahren zu prüfen, wie stark die einzelnen Merkmale korreliert sind<sup>37</sup>. Allerdings sollen die Merkmale der Unterscheidung von Objekten dienen und deshalb folgenden Anforderungen im Bezug auf Korrespondenzanalyse genügen:

- Merkmale sind ähnlich für Objekte der gleichen Klasse in einer Bildsequenz
- Merkmale sind verschieden für Objekte unterschiedlicher Klassen
- Merkmale sollten untereinander weitgehend unabhängig und berechenbar sein

Der dritte Punkt ist deshalb zu beachten, weil unnötig viele Merkmale den Merkmalsraum ausdünnen, was - wie später beschrieben - zur schlechteren Korrespondenzanalyse/Zuordnung und Klassifikationsergebnissen führen können. Die hier beschriebenen Anforderungen an die Merkmale haben einen wesentlichen Einfluss auf die Auswahl der letztendlich für das Matching verwendeten Merkmale. Deshalb wird auf die Merkmalseignung für das Matching eingegangen. Diese fordert erst die Erfassung der

---

<sup>37</sup> Die Merkmale sollten also in ihren Mittelwerten verschieden sein, eine geringe Standardabweichung besitzen und zusätzlich über geringe Kovarianz verfügen.

### 3. MATCHING-PARADIGMA

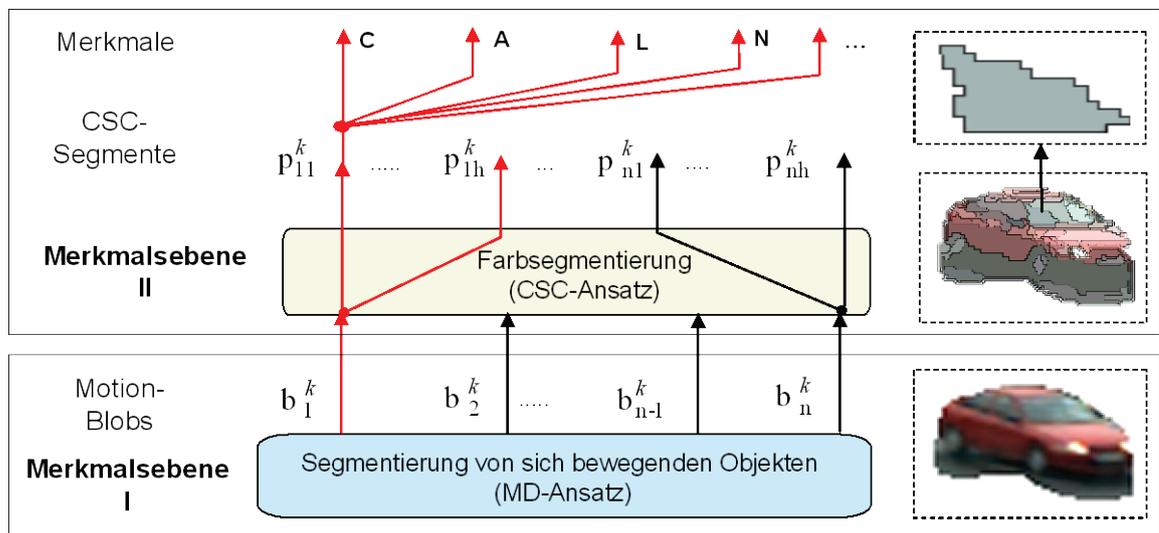
Bildmerkmale in zwei Ebenen (Abb. 3-4-1), die aus dem vorgestellten MD-Ansatz (Merkmalsebene I) und der Farbsegmentierung (Merkmalsebene II) hervorgehen.

#### 3.4.1 Merkmalsebene I - Motion-Blobs

Der im Abschnitt 3.3 erläuterte MD-Ansatz in Verbindung mit dem Schattenfilter ermöglicht die Erstellung einer Bewegungsmaske, welche genaue Bildregionen beschreibt, die einen sich im zeitlichen Verlauf verändernden, bewegten Inhalt besitzen und allgemein als Blobs bezeichnet werden. Im Folgenden wird für solche Regionen der Begriff Motion-Blob (MB) verwendet. Ein MB kann neben seinen deskriptiven statischen Merkmalen auch mehrere dynamische verhaltensabhängige Eigenschaften wie Erzeugung, Verschwinden, Verschmelzung, zeitweise Überlappung oder auch Zerfallen aufweisen.

- Erzeugung ereignet sich beispielsweise, wenn ein neuer MB von dem alten erzeugt wird und dadurch zusätzliche neue Merkmale entstehen.
- Verschwinden wird vorkommen, falls ein MB in der weiteren Analyse aufgrund der schwierigen Erfassung der Bewegungsinformation (z.B. Stau, Ampel oder Verlassen der Bildszene) nicht mehr berücksichtigt werden kann.
- Verschmelzung und Zerfallen<sup>38</sup> können auftreten, wenn eine Objektinteraktion stattfindet. Diese dynamischen Eigenschaften können nur in einem zeitlichen Kontext betrachtet werden.

Die dynamischen verhaltensabhängigen Eigenschaften führen zu Änderungen der deskriptiven statischen Eigenschaften eines MBs, die sich in spektrale, geometrische, statistische und texturbasierte Merkmale unterteilen lassen.



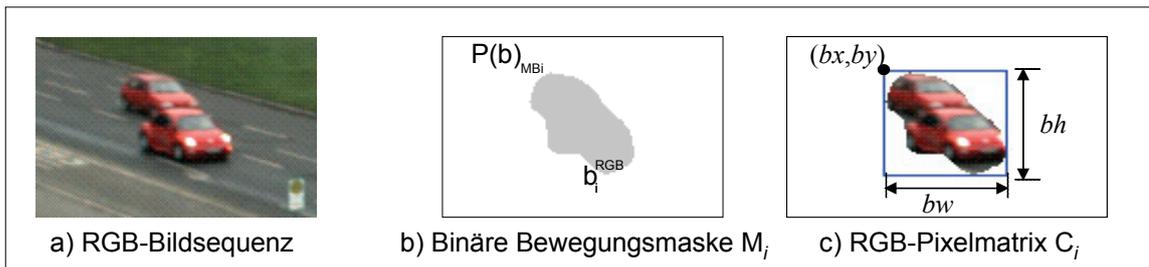
**Abbildung 3-4-1:** Das Prinzip der hierarchischen Merkmalsextraktion und die Zerlegung von MBs in kleine Farbsegmente bei der Verwendung des CSC-Farbsegmentierungsalgorithmus.

Während die spektralen Merkmale (bspw. Farbtöne) stark für die Verbesserung der Objekterkennung bei gestörten Situationen eingesetzt werden können, zeigen die restlichen formbasierten Merkmale prinzipiell eine gute Einteilungsfähigkeit für die

<sup>38</sup> Das Verschmelzen und das Zerfallen können in Verbindung mit der Kontextanalyse angewendet werden, um Anhaltspunkte über die dynamischen Änderungen und Mehrdeutigkeiten zu treffen.

### 3. MATCHING-PARADIGMA

definierten Objektklassen, wobei einzelne Merkmale besonders positiv in Erscheinung traten. Obwohl die Farbinformation durchweg eher entscheidende Ergebnisse bei der Erkennung bringen kann, konnten für die betrachteten Motion-Blobs (in der ersten Ebene) erst keine signifikanten Vorteile gegenüber anderen geometrischen Merkmalen nachgewiesen werden. Wesentlich dafür ist die Tatsache, dass die realen Bildsequenzen überwiegend Grauwerttöne beinhalten, die zur geringeren Varianzen der Farbkomponenten führen. Die geometrischen und statistischen Merkmale konnten insofern als hinreichend repräsentativ eingeschätzt werden, da sich mit einem realen Objekt tatsächlich viele charakteristische Daten gewinnen lassen (Blickwinkelwechsel). Bei einem Farbmerkmal ist dies nicht gültig (der Farbsättigungswert bleibt oft über die Szene gleich). Hier sind mehr Daten unterschiedlicher Objekte bei unterschiedlichen Witterungsbedingungen notwendig.



**Abbildung 3-4-2:** Beispiel für die RGB-Pixelmatrix eines Motion-Blobs. Die Punkte der Region  $M$  haben den Wert 1 und der Untergrundpunkte den Wert null b).

Ausgehend davon, wurden in der ersten Ebene der Merkmalsextraktion nur die geometrische und spektrale Charakteristik des MBs berücksichtigt. Dabei verwenden wir  $C$  für Farbe,  $A$  die Fläche,  $N$  den Umfang und  $J_i^k$  beschreibt hier die Menge von den Farbsegmente im MB. Natürlich können noch weitere Merkmale berücksichtigt werden, die aber von der Anwendung abhängig sind. Eine Entscheidung über die Anzahl der Merkmale wird in den Optimierungsschritten in Abschnitt 3.6 dargestellt. Dies führt dazu, dass sich der Motion-Blob  $b_i$  in der ersten Merkmalsebene nun durch den *Viertupel*  $b_i^k$  beschreiben lässt:

$$B^k = \{b_1^k, b_2^k, \dots, b_m^k\} \quad (3-4-1a)$$

$$b_i^k = (C_i^k, A_i^k, N_i^k, J_i^k) \text{ mit } b_i^k \in B_i^k \text{ und } 1 \leq i \leq m \quad (3-4-1b)$$

Hierbei ist  $b_i$  der  $i$ te Motion-Blob und  $m$  ist die Gesamtanzahl von MBs in Bild  $k$ .  $B^k$  in Gleichung 3-4-1a stellt eine Menge der erkannten MBs in Bild  $k$  dar. Das Bezugsmerkmal  $C$  ist hier eine Menge von RGB Farbvektoren, welche die Topologie eines MBs mitdefinieren. Die Eigenschaften eines Motion-Blobs  $b_i$  werden im Detail wie folgt definiert:

- Sei  $M^k$  eine binäre Region, die einen MB innerhalb eines Bildes  $I$  maskiert. Die Punkte der Region innerhalb der Bewegungsmaske haben den Wert 1 und der Untergrundpunkte den Wert null.
  - Sei  $(bx, by)$  die linke obere Koordinate von  $b_i$  in der Maske  $M^k$  zum Zeitpunkt  $k$ .
  - Sei  $bw$  und  $bh$  die Breite bzw. Höhe von  $b_i$  innerhalb des RGB-Bildes  $I$  zur Zeit  $k$
- a) RGB-Farbmatrix

### 3. MATCHING-PARADIGMA

$$C_i^k = \sum_{y=by}^{by+bh-1} \sum_{x=bx}^{bx+bw-1} M_i^k(x,y) \bullet I^k(x,y) \text{ für } M_i^k(x,y) = \{0,1\} \quad (3-4-2a)$$

Die Matrix  $C_i^k$  entspricht einem MB und beinhaltet ein Farbvektorfeld, das sich aus der binären Bewegungsmaske  $M_i$  und dem aktuellen Bild  $I$  durch ein elementweises Produkt berechnen lässt. Aus der Matrix  $C_i^k$  lassen sich alle weiteren Komponenten des Merkmalsvektors (*Viertupel*  $b_i^k$ ) ableiten.

b) Fläche

$$A_i^k(b_i^k) = \sum_{y=by}^{by+bh-1} \sum_{x=bx}^{bx+bw-1} M_i^k(x,y) \text{ für } M_i^k(x,y) = \{0,1\} \quad (3-4-2b)$$

Die Fläche  $A_i^k$  entspricht der Anzahl aller Pixel von  $b_i$  in der Maske  $M_i$ . Dieses Flächenmerkmal ist gemäß der Gleichung 3-4-2b einfach zu berechnen, auch beim Zusammenfügen zweier MB-Regionen, für die das Merkmal bekannt ist.

c) Farbsegmentmenge

$$J_i^k(b_i^k) = \{p_{i1}^k, p_{i2}^k, \dots, p_{ih}^k\} = \bigcup_h p_{ih}^k \quad (3-4-2c)$$

Durch die Anwendung eines Farbsegmentierungsalgorithmus, der auf den RGB-Pixeldaten der Matrix  $C_i^k$  operiert, wird für jeden MB  $b_i$  eine Menge von Farbsegmenten erzeugt. Die Farbsegmentmenge  $J_i^k$  repräsentieren die Eingangsbildmerkmale der zweiten Ebene (siehe Abschnitt 3.4.2). Die Berechnung der einzelnen Merkmale sowie die für das Matchingverfahren notwendige Bestimmung eines Ähnlichkeitsmaßes werden in den folgenden Abschnitten beschrieben.

#### 3.4.2 Merkmalsebene II - CSC-Patches

Die effiziente Erzeugung von qualitativ hochwertigen Farbsegmentprimitiven ist für die Verbesserung der Güte des späteren behandelten Matching von grundlegender Bedeutung. In dieser Arbeit lassen sich die Eigenschaften eines CSC-Segments  $p_v$  durch die Tupel  $p_v^k$  beschreiben, die zusätzlich zu den erwähnten Merkmalen in der ersten Ebene weitere Merkmale wie Kompaktheit ( $Ko$ ), Schwerpunkte ( $L$ ) und Gültigkeitsstatus ( $G$ ) beinhalten. Die Beschränkung auf diese Merkmale ist aus Effizienzgründen<sup>39</sup> fundiert und hat folgende Vorteile: gute Segmentbeschreibung und Datenreduzierung bei der Zuordnung.

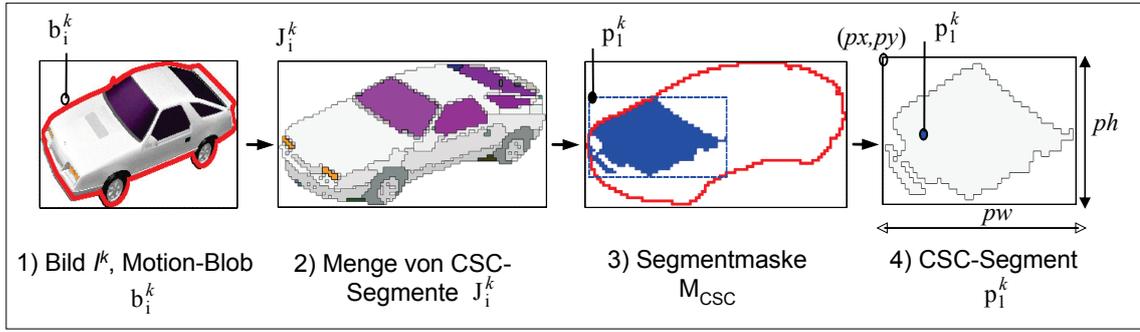
$$J_i^k = \{p_1^k, p_2^k, \dots, p_h^k\} \quad (3-4-3a)$$

$$p_v^k = (\mathbf{F}_v^k, A_v^k, L_v^k, N_v^k, Ko_v^k, G_v^k) \text{ mit } p_i^k \in J_i^k, 1 \leq i \leq m \text{ und } 1 \leq v \leq h \quad (3-4-3b)$$

Hierbei ist  $v$  die Farbsegmentnummer im Motion-Blob  $b_i$  in Bild  $k$  und  $h$  die Anzahl aller Segmente  $p_v$  innerhalb des Motion-Blobs  $b_i$ .  $\mathbf{F}$  ist eine Menge von RGB Farbvektoren, welche das Segment  $v$  charakterisiert und  $L$  gibt die Position vom Segment  $v$  an.

<sup>39</sup> Optimierungsschritte in Abschnitt 3.6

### 3. MATCHING-PARADIGMA



**Abbildung 3-4-3:** Berechnung eines CSC-Patch Segments  $p_i$ . Die Punkte der Maske  $M_{CSC}$  haben den Wert 1 und die Untergrundpunkte den Wert null.

In Abhängigkeit von der Anwendung könnten auch zusätzliche Merkmale definiert werden. Die Eigenschaften eines CSC-Segments  $p_v$  bzgl. MB  $b_i$  werden im Detail wie folgt definiert:

- $M_{CSC}^k$  sei eine binäre Maske, die das CSC-Segment  $p_j$  innerhalb  $b_i$  maskiert. Die Punkte der Maske haben den Wert 1 und die Untergrundpunkte den Wert null.
- $(px, py)$  sei die linke obere Koordinate von  $p_v$  in der Maske  $M_{CSC}^k$  zur Zeit  $k$ .
- $pw$  und  $ph$  sei die Breite bzw. Höhe von  $p_h$  innerhalb des Motion-Blob ( $b_i$ ) zur Zeit  $k$

a) RGB-Pixelmatrix

$$p_v^k = \sum_{y=py}^{py+ph-1} \sum_{x=px}^{px+pw-1} M_{CSC}^k(x,y) \cdot b_i^k(x,y) \text{ für } M_{CSC}^k(x,y) = \{0,1\} \quad (3-4-4a)$$

Die RGB-Pixelmatrix  $p_v^k$  entspricht einem CSC-Segment und berechnet sich aus der binären Maske  $M_{CSC}^k$  und der RGB-Pixelmatrix des MB  $b_i$  (Abb. 3-4-8). Aus  $p_v^k$  lassen sich alle weiteren Komponenten des Merkmalsvektors (*Viertupel*  $\Gamma_i^k$ ) ableiten.

b) Fläche

Die Fläche  $A_v^k$  eines Segmentes entspricht der Anzahl aller Pixel des Segments  $p_v$  und kann aus dem Moment  $m_{00}$  gerechnet werden, wie sich unmittelbar aus der Definition des Moments ergibt<sup>40</sup>.

c) Umfang

$$N_v^k(p_v^k) = \sum_{y=py}^{py+ph-1} \sum_{x=px}^{px+pw-1} C_{CSC}(x,y) \text{ und } C_{CSC}(p_v^k) = \bigcup_n P_n(x,y) \quad (3-4-4b)$$

$N_v^k$  ist der Umfang des Segmentes, der sich aus den Bildpunkten des zum Umfang transformierten Binärbildes ergibt, entspricht in der Regel der Anzahl aller Konturpunkte.  $C_{CSC}$  repräsentiert die Menge aller  $n$  2D-Konturpunkte ( $P$ ) eines Segments  $p_j$ . Im Allgemeinen gilt für alle Beschreibungsvarianten von Konturen, dass sie sich durch hinzufügen einzelner Bildpunkte am Rand des Segmentes oder durch zusammenfassen

<sup>40</sup> In diesem Fall sind  $x^p$  und  $y^q$  immer 1 und können daher weggelassen werden.

### 3. MATCHING-PARADIGMA

---

von Segmenten verändern. Diese Veränderung der Merkmale ist nicht einfach bestimmbar und erfordert in der Regel eine Neuberechnung. Somit sind sie als Beschreibungsmerkmal einer Region wenig geeignet. Trotzdem können sie als Grundlage für die Berechnung weiterer Merkmale dienen.

e) Schwerpunkt

$L_v^k$  ist der 2D-Flächenschwerpunkt eines Segments  $p_v$  und errechnet sich aus dem arithmetischen Mittel der Koordinaten aller dem Segment zugehörigen Flächenpunkte. Um  $L_v^k$  bzgl. der Bildkoordinaten zu betrachten, müssen die Koordinaten des Segments  $p_j$  und des zugrunde liegenden MB  $b_i$  hinzuaddiert werden (Abb. 3-18).

f) Farbe

$$\mathbf{F}_v^k(p_v^k) = \frac{1}{A(p_v)} \sum_{y=y_p}^{yp+ph-1} \sum_{x=px}^{px+pw-1} p_v^k(x, y) \quad (3-4-4c)$$

Die Farbe  $\mathbf{F}_v^k$  eines Segments  $p_j$  wird durch den durchschnittlichen Farbvektor  $\mathbf{F}_v^k$  aller zu der Region gehörender Bildpunkte beschrieben. Dazu wird für jede RGB Farbkomponente die Summe aller Farbwerte der Bildpunkte bestimmt und durch die Anzahl der Bildpunkte  $A$  geteilt.

g) Kompaktheit  $Ko_v$

$$Ko_v^k(p_v^k) = \frac{N_v(p_v)^2}{4\pi A_v(p_v)} \quad (3-4-4d)$$

Die Kompaktheit  $Ko_v^k$  eines Segments  $p_v$  stellt ein größenunabhängiges Merkmal dar. Der Kreisumfang bezieht sich auf den kleinsten vollständig umschließenden Kreis. Die Kompaktheit nimmt im Falle eines Kreises mit  $Ko=1$  den Minimalwert an. Für alle anderen geometrischen Objekte ist die Kompaktheit größer. Auf reale Objekte angewendet fällt auf, dass die Kompaktheit für Autos signifikant kleinere Werte liefert, als für Personen. Anschaulich betrachtet entspricht ein größer werdender Kompaktheitswert  $Ko$  einer zunehmenden Zerklüftung, d.h. einer wachsenden Zahl von Ausreißern an der Segmentgrenze.

h) Gültigkeitsstatus

$$G_v(p_v^k) = \begin{cases} \text{gültig,} & \text{falls } \langle A_{CSC} > A_{\min} \rangle \wedge \langle Ko_v < Ko_{\max} \rangle \\ \text{ungültig,} & \text{sonst} \end{cases} \quad (3-4-4e)$$

Der Gültigkeitsstatus  $G_v^k$  eines CSC-Segments  $p_v$  bestimmt, ob eine Korrelation mit anderen Segmenten möglich ist. Diese Entscheidung wird in Abhängigkeit von der Segmentgröße und Kompaktheit getroffen. Hat ein Segment eine sehr kleine Fläche oder einen hohen Kompaktheitswert, was auf eine zerklüftete, entartete Form hindeutet, so ist es sinnvoll dieses als ungültig zu klassifizieren und zur präventiven Fehlervermeidung von der Korrelation auszuschließen. Außerdem haben solche entarteten Regionen weitgehend unzuverlässige und instabile Merkmale in zeitlichem Verhalten, um bei dem CSC-Patch Korrelationsverfahren zu einem sinnvollen Ergebnis zu kommen.

Somit wurden im Prinzip die unterschiedlichen Merkmale extrahiert, die eine Grundlage für ein Multi-Matching darstellen können. Einige Merkmale lassen sich durch vergleichsweise einfache Rechenoperationen aus einer Menge von Hilfsmerkmalen (z.B. Farbwerte, Gültigkeitsstatus, Momente nullter bis zweiter Ordnung, Koordinaten usw.) errechnen, die wiederum bereits während des Regiongrowing, zur Bestimmung der CSC-Segmentpixelmaske leicht bestimmt werden können. Mittels der definierten Hilfsmerkmale insbesondere der Gültigkeitsstatus kann ferner auch das n:m Matching unterstützt werden (s. nächste Abschnitt).

#### 3.5 Hierarchisches merkmalsbasiertes Multi-Matching

Dieser Abschnitt beschreibt die zeitliche Zuordnung von Merkmalen, die erstens durch die Bewegungssegmentierung (Merkmalsebene I) und zweitens durch Farbsegmentierung (Merkmalsebene II) erfasst worden sind. Hier gelten Regionen als korrespondierend, wenn ein Ähnlichkeitskriterium, das zwischen beschreibenden Merkmalen berechnet wird, einen eindeutigen Extremwert liefert. Idealerweise sollten diese korrespondierenden Bereiche im Bild mit den durch die Projektion tatsächlicher Objekte im Bild entstehenden Bildbereichen übereinstimmen.

Allerdings kann auch eine robustere Bewegungs- oder Farbsegmentierung in den realen Bildsequenzen nicht gewährleistet werden, sodass die Regionen von Bild zu Bild nicht gleichmäßig stabil erfasst werden können. Kleine Änderungen im Bild, die nicht zwangsweise nur durch die Bewegung von Objekten verursacht werden müssen, sondern beispielsweise auch die Folge von nicht-kooperativen Messsituationen sind, können dazu führen, dass Regionen in einem Bild im nächsten Bild in mehrere Regionen zerfallen oder zu einer zusammenschmelzen und somit ein Zuordnungsproblem im Sinne eines Korrespondenzproblems in aufeinander folgenden Bildern entsteht. Diese Tatsache wird vom vorgeschlagenen Zuordnungsverfahren berücksichtigt, indem eine Zuordnung von  $n$  Regionen in einem Bild auf  $m$  Regionen in einem anderen Bild unter dem Einfluss von nicht-kooperativen Messsituationen vorgesehen ist.

Die Merkmalszuordnung ist hierarchisch aufgebaut, so dass das MB-Matching (Ebene I) als Basis für das CSC-Matching (Ebene II) dient. Dies ist in Abbildung (3-5-1) dargestellt. Dementsprechend erfolgt die Korrelation der Merkmale zuerst in der Ebene der MBs.

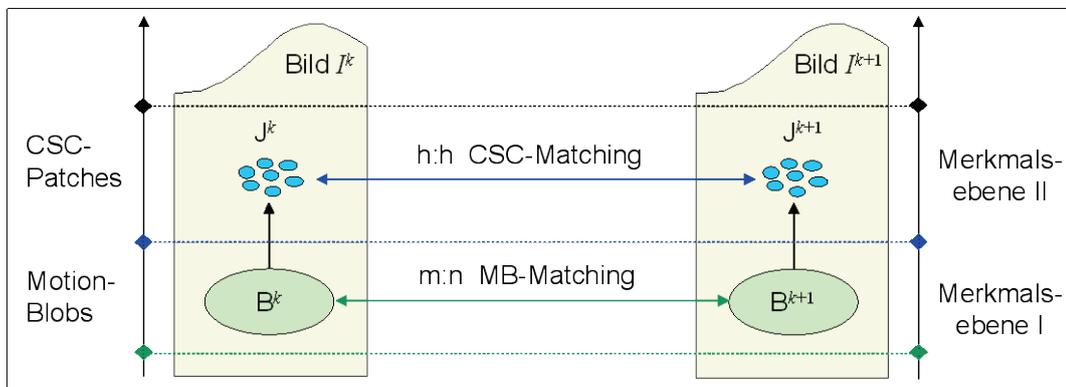


Abbildung 3-5-1: Multi-Matching mit zwei Merkmalsebenen

### 3. MATCHING-PARADIGMA

Auf der Grundlage des dort durchgeführten m:n MB-Matchings wird in einem zweiten Schritt die eineindeutige Zuordnung der CSC-Segmentprimitiven der gefundenen MB-Paare durchgeführt (Abb. 3-5-1). Die Stabilität des Multi-Matchings beruht somit auf der Trennung des Korrespondenzproblems in zwei Ebenen. In dem folgenden Abschnitt soll nun darauf eingegangen werden, wie die m:n Zuordnung der Ebene  $I$  realisiert werden kann und welche Probleme dabei auftreten können.

#### 3.5.1 m:n Motion-Blob-Matching

Das Matching kann im Prinzip für jede Blobregion entweder eine beliebige Anzahl von Korrespondenzen liefern (wobei die maximale Anzahl durch Festlegung einer Schwelle begrenzt werden kann) oder auch gar keine Korrespondenz (im Fall von Stausituationen) herstellen. Dadurch, dass die Blobs zumindest mit Teilen der Projektion von Objekten einer realen Szene korrespondieren, gelten in den meisten Fällen die folgenden Annahmen:

- Motion-Blobs verändern weder ihre Bewegungsrichtung noch die Bewegungsgeschwindigkeit plötzlich, so lange keine bildspezifischen Störungen vorkommen.
- Motion-Blobs können ihre Größe und Form signifikant verändern, wenn gestörte Bildsituationen auftreten. Eine (signifikante) Veränderung der Regionenform und -größe lässt darauf schließen, dass der zugehörige MB sich verschmilzt oder zerfällt, teilweise verdeckt wird oder sich in einer Richtung bewegt, die nicht parallel zur Bildebene liegt. In jedem Fall ist davon auszugehen, dass sich die geometrischen und statistischen Merkmale sich bewegender Regionen plötzlich verändern können. Diese Tatsache führt zu einem Korrespondenzproblem aufgrund der physikalischen wie auch visuellen Erscheinungen.

Die Aufgabe des MB-Matchings ist die Zuordnung einer Menge von  $m$  MB in Bild  $I^k$  zu einer Menge von  $n$  MB in Bild  $I^{k+1}$ . Dieses MB-Matching basiert auf der Annahme, dass reale Objekte, nur eine begrenzte Entfernung zwischen zwei Folgebildern zurücklegen können. Daraus folgt, dass ein Objekt  $O$ , welches in Bild  $I^k$  auf einen MB  $b_1$  abgebildet wurde, in Bild  $I^{k+1}$  auf einen MB  $b_2$  abgebildet wird, der sich in der Nähe von  $b_1$  befindet (Abb. 3-5-2).

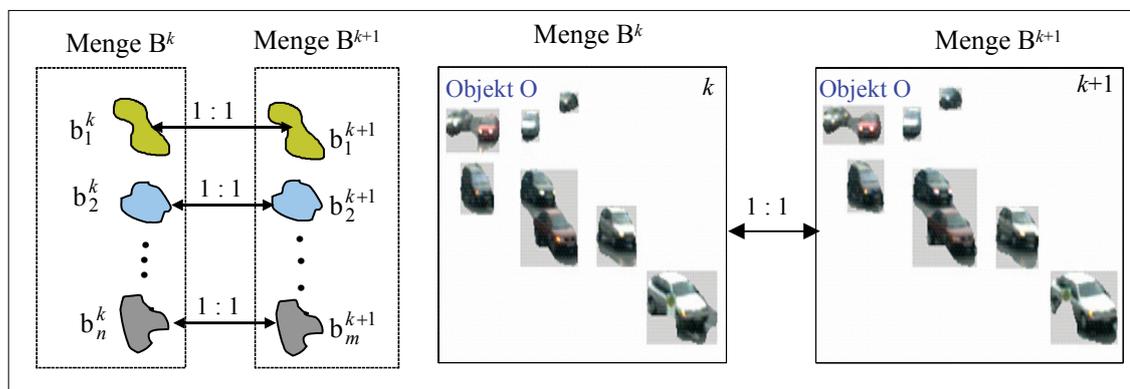
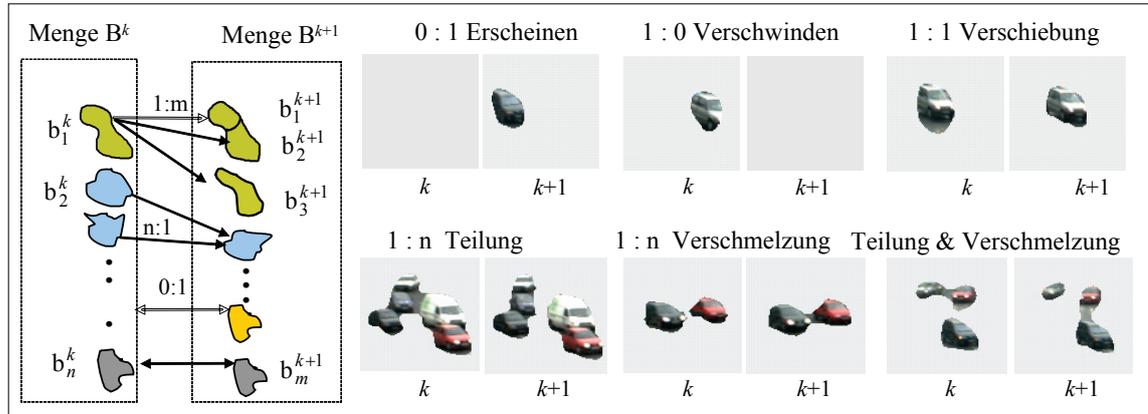


Abbildung 3-5-2: MBs in Folgebildern bei einfacher 1:1 Zuordnung

Allerdings ist bei der Zuordnung zu berücksichtigen, dass ein MB sich nicht nur bewegen, sondern auch verschmelzen, zerfallen oder verschwinden kann. Die Zuordnungsmöglichkeiten sind daher vielfältig. Der einfachste und häufigste Fall ist

### 3. MATCHING-PARADIGMA

eine 1:1 Zuordnung, bei der ein MB aus Bild  $I^k$  genau einem in Bild  $I^{k+1}$  zugeordnet wird (Abb. 3-5-2). Die allgemeine m:n Zuordnung entsteht, wenn aufgrund räumlicher Nähe, mehrere MB simultan verschmolzen oder geteilt werden (Abb. 3-6-3).



**Abbildung 3-5-3:** Das Prinzip von n:m Zuordnung und die häufig auftretende Zuordnungsmöglichkeiten beim Motion-Blob-Matching.

Formal betrachtet ist die Ausgangssituation des Matchings zwischen zwei Bildern  $I^{k+1}$  und  $I^k$  folgende:

$$B^k = \{b_1^k, b_2^k, \dots, b_n^k\} = \bigcup_i b_i^k ; B^k \in I^k \text{ und} \quad (3-5-1)$$

$$B^{k+1} = \{b_1^{k+1}, b_2^{k+1}, \dots, b_m^{k+1}\} = \bigcup_i b_i^{k+1} ; B^{k+1} \in I^{k+1}$$

Die Menge  $B^k$  repräsentiert die Menge alle MBs in Bild  $I^k$  und die Menge  $B^{k+1}$  die Menge der MB im Bild  $I^{k+1}$ . Durch das m:n Multi-Matching sind die Mengen vollständig zu partitionieren, dabei ist jeweils einer Untermenge  $U^k \subseteq B^k$  eine Untermenge  $U^{k+1} \subseteq B^{k+1}$  zuzuordnen ( $n = |U^k|$  und  $m = |U^{k+1}|$ ). Das entwickelte n:m-Matching basiert auf folgenden zwei Annahmen:

- korrespondierende MB-Regionen befinden sich in unmittelbarer Nähe
- die von den MB-Regionen bedeckte Fläche ist ähnlich

Der formale Algorithmus, der diese Annahmen umsetzt, ist in Abbildung 3-5-4 beschrieben. Die aus dem Matching resultierenden Ergebnismengen  $U^k$  und  $U^{k+1}$  werden zusammen als MB-Matching-Paar  $M_{MB}^k = \{U^k, U^{k+1}\}$  bezeichnet. Die häufigsten der in praktischen Untersuchungen aufgetretenen Zuordnungsfälle von Matching-Paaren  $M_{MB}^k$  sind in Abbildung 3-5-4 dargestellt. Die Gesamtheit aller Paare  $M_{MB}^k$  stellt die Grundlage für die Korrelation der Farbsegmente in der zweiten Merkmalsebene dar.

Um die Schwachstellen dieses MB-Matchings<sup>41</sup> zu verdeutlichen, sei zunächst angenommen, dass bei der Bewegungssegmentierung n MBs gefunden worden seien. Sei n die Anzahl der in einem Bild gefundenen MBs und m die Anzahl der im zweiten Bild gefundenen MBs, dann gibt es  $2^n$  mögliche Untermengen ( $U^0$ ) der MBs des ersten

<sup>41</sup> Es kann hier nicht für sämtliche möglichen Untermengen der MBs aus zwei Bildern eine endgültige Zuordnung gefunden werden, da erstens der Aufwand hierfür schon bei einer relativ kleinen Anzahl von MBs sehr hoch ist und zweitens das Korrespondenzproblem durch diese Zuordnung nicht vollständig lösbar ist. Hier geht es in erster Linie darum, eine grobe Zuordnung der MBs zu erzielen, die zu einer signifikanten Reduzierung der Mehrdeutigkeiten in der zweiten Segmentebene führt.

### 3. MATCHING-PARADIGMA

---

Bildes und  $2^m$  mögliche Untermengen ( $U^1$ ) von MBs des zweiten Bildes. Um einen Matchversuch von jeder möglichen MB-Teilmenge des ersten Bildes auf jede mögliche MB-Teilmenge des zweiten Bildes vorzunehmen, müssten nun  $2^{n+m}$  Vergleiche vorgenommen werden. Die Anzahl der möglichen Teilmengen steigt mit der Anzahl der Regionen in den Teilmengen extrem schnell, falls nur Teilmengen einer relativ geringen Anzahl von Regionen gebildet werden.

---

```
MB-Matching ( $k=0$ )
{
  while ( $B_0 \neq \emptyset$ ){
     $U^0, U^1 = \emptyset$  // neue Untermengen  $U^0$  und  $U^1$ 
     $\forall b \in B_0 \{ U^0 = U^0 \cup b, B_0 = B_0 \setminus b \}$  // füge b als Element von  $S_0$  zu  $U^0$  hinzu
    //und entferne b aus der Menge  $S_0$ 
     $\forall (b' \in B_1, b' \neq b) \{ U_1 = U_1 \cup b', B_1 = B_1 \setminus b' \}$  //füge für alle MB  $b'$  aus  $S_1$ , die MB b hinzu
     $A_0 = \text{Fläche}(b)$  //schneide  $b'$  zu  $U^1$  hinzu und entferne  $b'$ 
  aus  $B_1$ 
     $A_1 = \sum \text{Fläche}(b')$ 
    //für alle  $b_0$  aus  $B_0$ , die  $b_1$  aus  $U^1$  schneiden und deren Fläche die Differenz zwischen  $A_0$  und  $A_1$ 
    //minimieren, füge  $b_0$  zu  $U^0$ , entferne  $b_0$  aus  $B_0$  und addiere die Fläche von  $b_0$  zu  $A_0$ .
     $\forall ((b_0 \in B_0) \# (b_1 \in U^1) \wedge |A_0 + \text{Fläche}(b_0) - A_1| \rightarrow \min) \{$ 
       $U^0 = U^0 \cup b, B_0 = B_0 \setminus b, A_0 = A_0 + \text{Fläche}(b_0)$ 
    }
  }
   $M_{MB}^k = \{U^0, U^1\}$ 
   $k = k + 1$ 
}
```

---

**Abbildung 3-5-4:** Algorithmus m: n MB-Matching [Niese 03, Senst 09]

Auch bei realen Bildern können zusätzlich bildspezifische Störungen auftreten, die dazu beitragen können, dass Untermengen nicht eindeutig erfassbar sind. Sie können zerfallen, verschmelzen oder verschwinden.

Allgemeiner betrachtet muss man davon ausgehen, dass mehrere zerfallene oder verschmolzene Regionen, die teilweise zu einem MB gehören und teilweise zum Bildhintergrund oder zu einem anderen MB, eine uneindeutige n:m Zuordnung verursachen. Eine Folge davon sind schlechtere Ähnlichkeitswerte, als bei einer 1:1 Zuordnung der Blobregionen allein. Wesentlich dafür ist die Tatsache, dass beispielsweise die Fläche oder der Schwerpunkt, als eines der für das Matching herangezogenen Merkmale sich natürlich bei bewegten Regionen verändert, was die Ähnlichkeit zwischen beiden Flächen verringert. Die Veränderungen der resultierenden Flächen fällt aber in der Regel stärker aus, wenn sich bewegende Regionen mit anderen Regionen verschmelzen. Die Ähnlichkeitsberechnung dieser Merkmale ist dann schlechter. Das hier beschriebene Problem hängt grundsätzlich mit dem n:m-Matching zusammen.

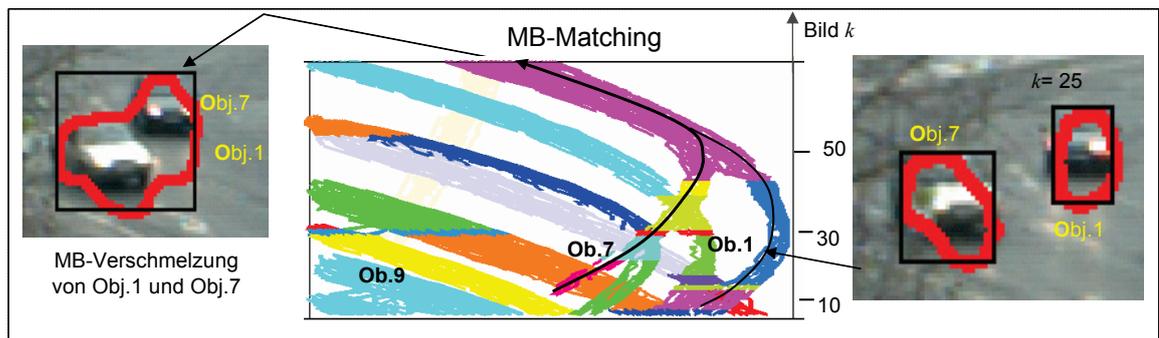
**Fazit:** Durch den in [Rehrmann 98a] angegebenen Vorschlag, für die Bildung der Regionenteilmengen für das Matching die Farbähnlichkeit der Regionen als ein Kriterium zu betrachten, tritt das Problem nur noch dann auf, wenn die sich unterschiedlich bewegenden Regionen ähnliche Farbwerte haben. Allein die Beschränkung auf die *Farbmerkmale* in dieser Ebene wäre nicht ausreichend. Die Farbsättigung kann durchweg eher mittelmäßige Zuordnungsergebnisse bringen.

Für das hier betrachtete Matching konnten keine signifikanten Vorteile gegenüber anderen Merkmalen nachgewiesen werden, solange man die MBs nicht in kleine

### 3. MATCHING-PARADIGMA

Farbcluster zerlegt hat. Leider würden dadurch aber bessere Übereinstimmungen mit größerem  $n$  bzw.  $m$  möglicherweise nicht mehr untersucht und damit nicht erkannt werden. Der dadurch entstehende Nachteil wäre vermutlich größer als die so vermeidbaren Fehler. Deshalb wird diese Herangehensweise hier nicht weiter betrachtet und das Problem im Rahmen der Implementierung in dieser Arbeit reduziert, indem ein Multi-Matching der beteiligten MBs auf der Basis von Farbsegmenten<sup>42</sup> gefordert wird.

Das Ziel des Multi-Matchings ist eine eindeutige Zuordnung eines CSC-Segmentes  $p_v^0 \in J_0$  mit  $J_0$  der Menge der CSC-Segmente aus der Menge der MB  $U_0$  zu einem CSC-Segment  $p_\mu^1 \in J_1$  der Menge der CSC-Segmente aus der MB-Menge  $U_1$ , für eine MB-Paarung, basierend auf der Berechnung von Ähnlichkeitsmaßen. Aus diesem Grund wird dieses Matching in dem folgenden Abschnitt als 1:1 CSC-Multi-Matching bezeichnet.



**Abbildung 3-5-5:** Aufgrund von dichtem Straßenverkehr oder Reflektionen treten hier häufig Probleme auf, die zur Verschmelzung von Objekten führen. Das MB Matching wird hier nicht ausreichen, da Teilverdeckungen, Verschmelzungen und Teilungen auftreten. Durch die sichtbaren Merkmale (Segmente) kann man eine Aussage über die Zuordnung treffen

#### 3.5.2 1:1 CSC- Multi-Matching

Dieser Abschnitt beschreibt das so genannte CSC-Multi-Matching<sup>43</sup> von Segmenten, die aus der Farbsegmentierung von MBs resultieren. Die verwendeten Merkmale dafür wurden im Abschnitt 3.5.1 beschrieben. Dabei werden nur die Segmente in dem Matchingsprozess einbezogen, die eine minimale Ähnlichkeit aufweisen. Änderungen der MBs, die nicht zwangsweise durch die Bewegung von Objekten hervorgerufen werden, sondern auch die Folge von nicht-kooperativen Situationen sind, können dazu

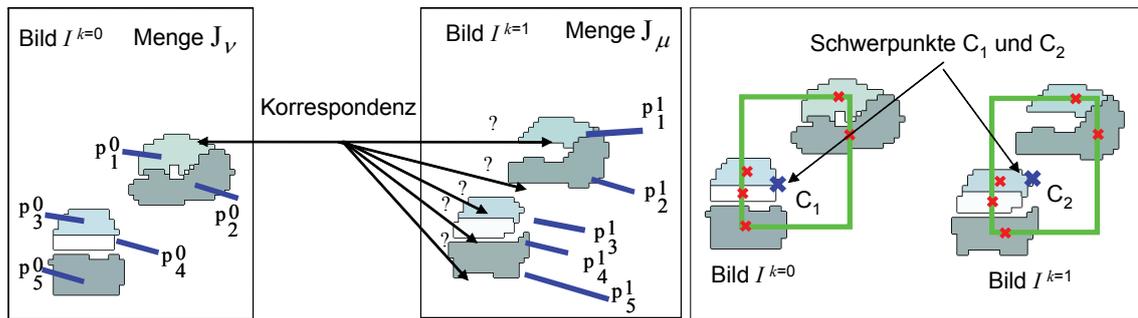
<sup>42</sup> Hier wird die Farbinformation im doppelten Sinne ausgenutzt: Zum einen zur Zerlegung von MBs in kleine Segmente und zum anderen zu Korrespondenzherstellung zwischen Segmenten.

<sup>43</sup> Das CSC-Multi-Matching wird in diesem Dokument auch als CSC-Matching bezeichnet.

### 3. MATCHING-PARADIGMA

führen, dass die Segmente nicht konstant bleiben. Eine Glättung der Segmente über die Bildfolge konnte zwar auf der Ebene von Bildpunkten realisiert werden, nicht aber auf der Ebene von Segmenten.

Die Kenntnis, dass Segmente, die stets dasselbe Szenenobjekt repräsentieren, in der Regel nur geringfügigen Bild-zu-Bild Änderungen unterliegen, konnte in der Segmentierung nicht unmittelbar genutzt werden. Diese Tatsachen werden vom CSC-Multi-Matching berücksichtigt, indem ein Matching von  $\nu$ -Segmente in einem Bild auf  $\mu$ -Segmentregionen in einem anderen Bild vorgesehen ist. Die Verknüpfung dieser Segmente miteinander erfolgt durch ein Ähnlichkeitskriterium. Dieses liefert einen Extremwert (Maximum oder Minimum) je nachdem, welches Ähnlichkeitskriterium man verwendet. Es hängt von der ausgeprägten lokalen Struktur des Segmentes ab, ob eine eindeutige Korrespondenzbeziehung ermittelt werden kann.



**Abbildung 3-5-6:** Zwei CSC-Patch-Mengen eines MB-Matching-Paares und (rechts) Schwerpunkte der CSC-Patch-Mengen.

Die mathematische Formulierung zur Berechnung dieses Ähnlichkeitskriteriums für alle verwendeten Merkmale wird wie folgt beschrieben.  $M_{MB}^k$  sei ein MB-Matching-Paar, das durch die  $m:n$  MB-Zuordnung in zwei Bildern  $I^0$  und  $I^l$  bestimmt wurde (Abschnitt 3.5.1).

$$M_{MB}^k = \{U^k, U^{k+1}\} \text{ und } U^k, U^{k+1} \cong \text{korrelierte MB-Menge} \in \text{Bild } I^k \text{ und } I^{k+1} \quad (3-5-2a)$$

Ferner seien  $J_\nu^k$  und  $J_\mu^{k+1}$  zwei Mengen von CSC-Segmenten, die den MBs  $b_i \in U^k$ , bzw.  $b_j \in U^{k+1}$  zugehörig sind (Abb. 3-5-6).

$$J_\nu^k = \{p_1^k, p_2^k, \dots, p_h^k\} = \bigcup_\nu p_\nu^k(b_i), \quad b_i \in U^k \text{ und} \quad (3-5-2b)$$

$$J_\mu^{k+1} = \{p_1^k, p_2^k, \dots, p_h^k\} = \bigcup_\mu p_\mu^k(b_j), \quad b_j \in U^{k+1} \quad (3-5-2b)$$

$$1 \leq \nu \leq h \text{ und } 1 \leq \mu \leq h$$

Das 1:1 Matching, welches eine eindeutige Zuordnung eines CSC-Patches  $p_\nu^k \in J_\nu^k$  zu einem CSC-Patch  $p_\mu^{k+1} \in J_\mu^{k+1}$  realisiert, basiert auf der Berechnung von Ähnlichkeitsmaßen. Durch eine Kombination dieser Maße, die sich aus den Eigenschaften der zu korrelierenden CSC-Patches errechnen, kann eine hohe Genauigkeit bei verhältnismäßig geringem Aufwand erreicht werden (siehe Abschnitt 3.6). Geeignete Ähnlichkeitsmaße für das Matching von CSC-Segmenten sind:

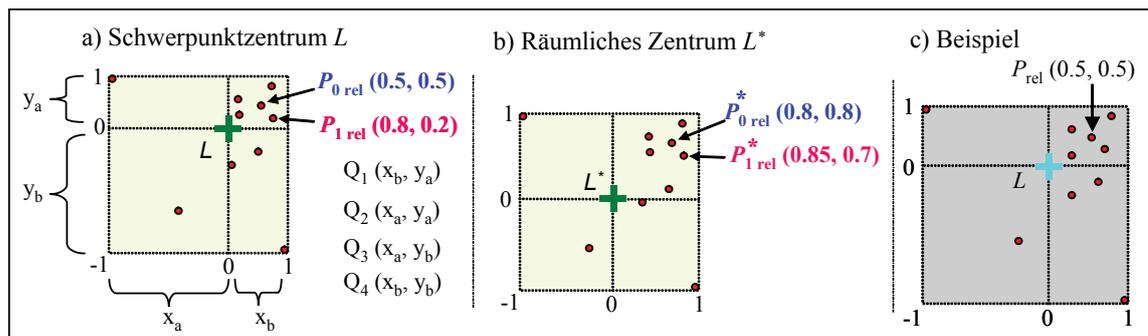
- a). SL - Relative Position innerhalb einer MB-Menge

- b). SI - Inter-Frame-Distanz
- c). SC - Farbwert
- d). SA - Größe

Die Berechnung der einzelnen Ähnlichkeitsmaße erfolgt zwischen allen CSC-Segmenten der Mengen  $J_v^k$  und  $J_\mu^{k+1}$ . Daraus folgt für das 1:1 CSC-Matching eine Komplexität  $O(n^2)$ . Der Aufwand lässt sich jedoch etwas reduzieren<sup>44</sup>. Im folgenden werden vier Ähnlichkeitsmaße im Detail vorgestellt:

**(a) SL - Relative Position innerhalb einer MB-Menge**

Die relative Positionsinformation einzelner CSC-Patch Segmente ist sehr hilfreich, um bei der Korrelation solche Segmente schnell auszuschließen, die falsch lokalisiert sind. Es kann zum Beispiel über ein Segment die Aussage getroffen werden, dass es sich im linken unteren Bereich einer CSC-Patch-Menge  $J_v^k$  befindet und damit sicher nicht mit einem Segment einer anderen Menge  $J_\mu^{k+1}$  in Beziehung steht, welches sich dort im rechten oberen Bereich befindet. Zur Umsetzung dieser Annahme hat es sich als vorteilhaft erwiesen, für die CSC-Patch-Mengen  $J_v^k$  und  $J_\mu^{k+1}$  jeweils den Schwerpunkt  $L_1$  bzw.  $L_2$  zu ermitteln (Abb. 3-5-7). Das Schwerpunktzentrum  $L$  einer CSC-Patch-Menge  $M$  wird durch das arithmetische Mittel der Positionen  $P_v(x_v, y_v)$  aller in  $J_v^k$  enthaltenen CSC-Patches  $p_v^k$  berechnet. Das räumliche Zentrum  $L^*$  entspricht hingegen dem Mittelpunkt der Gesamtausdehnung  $D$  der CSC-Patch-Menge  $J_v^k$  (Abb. 3-5-7b). Das räumliche Zentrum  $L^*$  ist sehr störanfällig, da ein einzelner Ausreißer dazu führen kann, dass sich  $L^*$  stark verändert. Aus diesem Grund ist  $L^*$  zur relativen Positionsangabe nicht praktikabel. Hingegen haben vereinzelte Ausreißer (im Falle des Schwerpunktzentrums  $L$  meist nur eine geringe Auswirkung.



**Abbildung 3-5-7:** Relative CSC-Patch-Position, Beispiel a) Schwerpunkt, b) räumlich.

Das Gebiet mit der Gesamtausdehnung  $D$ , welches durch die Segmente einer Menge  $J_v^k$  aufgespannt wird, wird durch das Zentrum  $L$  bzw.  $L^*$  in vier Quadranten  $Q_1$ - $Q_4$  geteilt (Abb. 3-5-7). Die relative Position  $P_{vrel}$  eines CSC-Segments  $p_v$  errechnet sich aus der Abbildung der Segment Position  $P_v$  auf einen der vier Quadranten. Die Abbildung entspricht einer einfachen Verhältnisgleichung mit Fallunterscheidung für jeden der vier Quadranten. Das Ergebnis  $P_{vrel}$  wird somit im Intervall  $[-1,1]$  für die  $x$ - und  $y$ -Richtung

<sup>44</sup> Es ist beispielsweise nicht sinnvoll, wenn zwei unabhängige Ähnlichkeitsmaße nur eine minimale Übereinstimmung anzeigen, noch ein drittes oder viertes zu berechnen.

### 3. MATCHING-PARADIGMA

bestimmt. Das Ähnlichkeitsmaß SL errechnet sich aus dem Grad der Übereinstimmung der relativen Positionen  $P_{vrel}^k(\tilde{x}_{relv}, \tilde{y}_{relv})$  und  $P_{\mu rel}^{k+1}(\tilde{x}_{rel\mu}, \tilde{y}_{rel\mu})$  zweier CSC-Patches  $p_v^k \in J_v^k$  und  $p_\mu^{k+1} \in J_\mu^{k+1}$ .

$$SL_{v,\mu}^k = 1 - \frac{\sqrt{(\tilde{x}_{relv} - \tilde{x}_{rel\mu})^2 + (\tilde{y}_{relv} - \tilde{y}_{rel\mu})^2}}{\sqrt{2}} \quad (3-5-3)$$

$$1 \leq v \leq h \quad \text{und} \quad 1 \leq \mu \leq h$$

Dabei ist  $(\tilde{x}_{relv}, \tilde{y}_{relv})$  die relative Position eines Segmentes  $p_v^k$  im Bezug auf dem Schwerpunkt  $L$  aller CSC-Segmente  $J_v^k$  innerhalb eines Motion-Blobs zur Zeit  $k$ . Zur Verbesserung der Robustheit des Matchings sind für dieses Maß einige Einschränkungen bzgl. der Merkmalszuverlässigkeit zu machen. Es ist nicht nützlich eine relative Position anzugeben, falls eine CSC-Patch-Menge  $J$  nicht über eine bestimmte Mindestausdehnung  $d_{\min}$  in  $x$ - oder  $y$ - Richtung aufweist oder  $J$  nicht über eine Mindestanzahl von  $N_{\min}$  Segmenten verfügt. Liegen zwei Segmente dicht beieinander, wird eine Korrelation lediglich zugelassen. Es sind jedoch weitere Kriterien nötig, um die Übereinstimmung zu sichern.

**Tabelle 3-5-1:** Beispiel einer Korrespondenzermittlung mittels SL-Maßes für zwei Mengen von Segmenten in aufeinander folgenden Bildern

SL	$p_1^1$	$p_2^1$	$p_3^1$	$p_4^1$	...
$p_1^0$	1	0	0	0	
$p_2^0$	0	1	0.794	0.554	
$p_3^0$	0	0.794	1	0.74	
$p_4^0$	0	0.554	0.74	1	
...					...

#### (b) SI- Minimale Inter-Frame-Distanz

Dieses Maß resultiert aus der Annahme, dass Segmente und deren Abbildungen lediglich eine begrenzte Entfernung zwischen zwei Bildern zurücklegen können. Die Umsetzung dieser Annahme kann durch die Berechnung des Euklidischen Abstandes zwischen den Mittelpunkten von zwei Segmenten  $p_v^k \in J_v^k$  und  $p_\mu^{k+1} \in J_\mu^{k+1}$  erfolgen. Der Abstand beider Segmente ist indirekt proportional zum Korrelationswert.

$$SI_{v,\mu}^k = 1 - \frac{\sqrt{(\tilde{x}_v - \tilde{x}_\mu)^2 + (\tilde{y}_v - \tilde{y}_\mu)^2}}{\text{dist}_{\max}}, \quad 1 \leq v \leq h \quad \text{und} \quad 1 \leq \mu \leq h \quad (3-5-4)$$

Dabei ist  $\text{dist}_{\max}$  die maximale Distanz zwischen zwei Segmenten in  $J_v^k$  und  $J_\mu^{k+1}$ .  $(\tilde{x}_v, \tilde{y}_v)$  und  $(\tilde{x}_\mu, \tilde{y}_\mu)$  sind die entsprechende Schwerpunktskoordinaten der Segmente  $v$  und  $\mu$  innerhalb zwei MBs. Das Segmentabstandskriterium kann dennoch sinnvoll genutzt werden, um eine Korrelation von Segmenten, die weit voneinander entfernt sind, zu unterbinden. Das Ähnlichkeitsmaß  $SI^k$  hat eine vergleichbare Wirkung<sup>45</sup>

<sup>45</sup> Sie unterscheiden sich in den Situationen, in der MBs verschmelzen oder sich teilen.

wie  $SL^k$ , die Kombination beider Maße führt jedoch zu einer höheren Robustheit beim Ausschluss ungeeigneter Korrelationspartner.

#### (c) SC- Ähnlichkeit des Farbwertes

Die Farbinformation ist ein zentrales Kriterium bei der Lösung des Korrespondenzproblems. Für die Ermittlung des Farbkorrelationswertes wird die Durchschnittsfarbe eines jeden CSC-Segmentes ausgewertet. Um Farbregionen vergleichen zu können, benötigt man ein Maß für die Ähnlichkeit von Farben. Eine verbreitete Methode zum Farbvergleich ist die Messung der Euklidischen Distanz zweier Farben in einem bestimmten Farbraum [Smith 96]. Hier entspricht eine minimale Distanz einer maximalen Übereinstimmung. Entscheidend ist die Wahl eines geeigneten Farbraumes, da die Ergebnisse mit den Farbräumen variieren.

In der Literatur werden RGB- und FSH-Farbraum (Farbton, Sättigung, Helligkeit) häufig genannt, um eine Farbähnlichkeitsbestimmung durchzuführen. Die Ähnlichkeit  $SC^k$  zweier normierte RGB-Farbvektoren  $\mathbf{F}_v^k = (r_v, g_v, b_v)$  und  $\mathbf{F}_\mu^{k+1} = (r_\mu, g_\mu, b_\mu)$  ist folgendermaßen definiert:

$$SC_{v,\mu}^k = 1 - \frac{1}{\sqrt{3}} \sqrt{(r_v^k - r_\mu^{k+1})^2 + (g_v^k - g_\mu^{k+1})^2 + (b_v^k - b_\mu^{k+1})^2} \quad (3-5-6)$$

$$1 \leq v \leq h \quad \text{und} \quad 1 \leq \mu \leq h$$

Der Farbabstand kann auf das Intervall  $[0,1]$  normalisiert werden, indem der Abstand durch den maximal möglichen Farbabstand im verwendeten Farbraum geteilt wird. Für die würfelförmigen Farbräume ergibt sich dieser maximale Abstand durch die Länge einer Diagonalen zwischen zwei Eckpunkten durch den Mittelpunkt des Würfels. Das Ergebnis des Ähnlichkeitswertes SC zweier Farben kann zwischen den beiden Farbräumen beträchtlich variieren. Bei der Untersuchung der Farbähnlichkeit einer Vielzahl von Farbsegmenten, welche aus realen Bildsequenzen gewonnen wurden, hat sich gezeigt, dass die Farbähnlichkeit im RGB-Farbraum tendenziell günstigere Eigenschaften für eine Verwendung als Ähnlichkeitsmaß aufweist. Insbesondere wurde hierzu die durchschnittliche Farbe von korrelierten CSC-Segmenten verglichen. Dabei wurde festgestellt, dass das Ähnlichkeitsmaß im FSH-Farbraum im Falle schwach gesättigter, dunkler Farben wie sie in realen Bildern häufig auftreten, meist nur schwach ausgeprägte Ergebnisse liefert. Insbesondere fallen Vergleiche zwischen Farbpaaren im FSH-Raum häufiger sehr ähnlich aus. Diese Beobachtung lässt sich durch die schwach ausgeprägte Verteilung der wenig gesättigten, dunklen Farben im FSH-Farbraum erklären, die bei der Euklidischen Distanzmessung, nur zu geringeren Unterschieden führt. Die Ergebnisse im RGB-Farbraum sind hingegen gleichmäßiger verteilt und eignen sich daher besser für die Aufgabe des Farbvergleichs von CSC-Segmenten.

#### (d) SA - Ähnlichkeit der Größe korrelierter CSC-Segmente

Der Flächeninhalt A eines Segments ist unabhängig von Translation parallel zur Bildebene und Rotation um die optische Achse. Bei den restlichen 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. Das

### 3. MATCHING-PARADIGMA

Ähnlichkeitsmaß  $SA_{\text{SIZE}}$  von den zwei Segmenten  $p_v^k$  und  $p_\mu^{k+1}$  berechnet sich aus dem Größenverhältnis der Segmentflächen  $A_v^k$  und  $A_\mu^{k+1}$ .

$$SA_{v,\mu}^k = 1 - \frac{|A_v^k - A_\mu^{k+1}|}{A_v^k + A_\mu^{k+1}}, \quad 1 \leq v \leq h \quad \text{und} \quad 1 \leq \mu \leq h \quad (3-5-7)$$

Die Ähnlichkeit der Segmentgröße ergänzt die Maße (a), (b) und (c) in sinnvoller Weise, da bei diesen das Merkmal der Größe nicht berücksichtigt wird. Der Unterschied zwischen der Fläche zweier Regionen oder Regionengruppen ergibt sich durch die absolute Differenz der beiden Flächen  $A_v^k$  und  $A_\mu^{k+1}$  normiert auf die maximale gemeinsame Fläche  $A_v^k + A_\mu^{k+1}$  und damit liegt das Ergebnis der Division im Bereich  $[0,1]$ . Um ein Ähnlichkeitsmaß zu erhalten, das mit Null keine und mit Eins eine volle Übereinstimmung wiedergibt, wird der Term von 1 abgezogen. Die Flächentopologie ist einzig von den Ergebnissen des Segmentierungsalgorithmus abhängig und weist gelegentlich leichte Variationen auf, die verschiedene Ursachen haben können, wie beispielsweise spontane Beleuchtungsänderungen oder Teilverdeckungen.

Berechnet man nun diese Ähnlichkeitsmaße für alle möglichen Segmente, so erhält man eine Zuordnungsmatrix der Ergebnisse. Da an dieser Stelle die Ähnlichkeitsfunktion von einem einzelnen Merkmal beispielsweise dem Flächeninhalt nicht unbedingt eine ausreichende zuverlässige Zuordnung aufweist, sind auch die anderen Ähnlichkeitsmaße zu berücksichtigen, wenn nämlich eine Mehrdeutigkeit aufgrund von einer Störung auftritt.

Formal erweitert sich dadurch die Matrixdimension des SM-Ähnlichkeitsmaßes. Der Vorteil dieser Erweiterung liegt darin, dass das unterschiedliche Verhalten der einzelnen Ähnlichkeiten optimal ausgenutzt werden kann, wenn Effekte aufgrund von nichtkooperativen Messsituationen auftreten. Diese kann verdeutlicht werden, indem man hier die Ähnlichkeitsergebnisse bei Deformation oder Verschmelzung oder Zerfallen genau betrachtet. Während die SA-Ähnlichkeitsfunktion sich signifikant ändert, bleiben hier die SC und SL nahezu konstant. Ausgehend von diesen Überlegungen erscheint es sinnvoll, aus den merkmalspezifischen Ähnlichkeitsmaßen ein kombiniertes Gesamtkriterium abzuleiten (gemäß Gleichung (3-5-8)).

$$SM_{v,\mu}^k = \left( \sum_{i=1}^4 w_i \right)^{-1} \cdot (w_1 \cdot SL_{v,\mu}^k + w_2 \cdot SI_{v,\mu}^k + w_3 \cdot SC_{v,\mu}^k + w_4 \cdot SA_{v,\mu}^k) \quad (3-5-8)$$

Traditionelle Möglichkeiten der Bestimmung der Wichtungsfaktoren, basieren entweder auf dem Vorschlag in [Al-Hamadi 01a], in dem das Gesamtkriterium zu gleichen Anteilen aus einzelnen Ähnlichkeitsmassen gebildet wird oder durch die Verwendung eines künstlichen neuronalen Netzes [Al-Hamadi 02c, 03f]<sup>46</sup> eingeschätzt wird. Die richtige Wahl der Gewichte  $\mathbf{w} = (w_1 \ w_2 \ w_3 \ w_4)^T$  ist grundlegend für die Stabilität Qualität des Matchings. Daher soll die Wichtung<sup>47</sup> der einzelnen vorgestellten

<sup>46</sup> Diese beiden erwähnten Vorschläge sind für intensitätsbasierte Ähnlichkeitsmaße (Betragsdifferenz) realisiert.

<sup>47</sup> Für die Bestimmung der Gewichte wären segmentierte Bildsequenzen ideal, in denen die korrekten Regionpaarbildungen 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 Farbsegmente befinden und bei Änderungen in der Farbsegmentierung die Lernsequenzen jeweils erneut erstellt werden müssten.

### 3. MATCHING-PARADIGMA

merkmalspezifischen Ähnlichkeitsmaße entsprechend ihrer Zuverlässigkeit erfolgen (s. Abschnitt 3.6).

Für die merkmalspezifischen Ähnlichkeitsmaße SL, SI ist aus der Untersuchung eine durchschnittliche Gewichtungsbewertung ( $w_1=1$  und  $w_2=1$ ) vorerst angemessen<sup>48</sup>. Das merkmalspezifische SC-Ähnlichkeitsmaß ist aufgrund der Farbinformation ein überaus robustes Zuordnungsmerkmal, besonders bei der Verwendung vieler kleiner Farbsegmente. Daher ist es angemessen, SC mit ( $w_3=4$ ), ein höheres Gewicht zuzuordnen<sup>49</sup>. Damit die hieraus resultierenden Variationen in der Segmentgröße nicht überbewertet werden, ist es sinnvoll für das Ähnlichkeitsmaß  $SA_{SIZE}$  eine durchschnittliche Gewichtung ( $w_4=1$ ) zu verwenden. Eine genaue Betrachtung und Optimierung der Gewichtungsfaktoren  $w$  erfolgen im Abschnitt 3.6.

Das Gesamtähnlichkeitsmaß SM wird in das Intervall  $[0,1]$  abgebildet, dabei entspricht die 0 keiner bzw. die 1 einer perfekten Übereinstimmung. Die Gewichte sind zum einen aufgrund der unterschiedlichen Größenordnungen der Einzelähnlichkeitskriterien erforderlich, zum anderen ermöglichen sie eine Bewertung der Merkmale bezüglich ihrer Relevanz für die Zuordnung. Die Menge aller Ähnlichkeitswerte lässt sich vorteilhaft anhand einer zweidimensionalen Korrelationstabelle beschreiben (Tabelle 3-5-2). Die Spalten der Korrelationstabelle beziehen sich auf Segmente aus  $J_1$ , die Zeilen auf Segmente aus  $J_0$ . Der Tabellenwert an der Position  $(v,\mu)$  repräsentiert den Korrelationswert zwischen den CSC-Segmenten  $p_v^k$  und  $p_\mu^{k+1}$ .

**Tabelle 3-5-2:** CSC-Patch-Korrelationstabelle, entspricht Beispiel aus Abb. 3-5-6

SM	$p_1^1$	$p_2^1$	$p_3^1$	$p_4^1$	...
$p_1^0$	0.997	0.546	0.529	0.473	
$p_2^0$	0.544	0.999	0.929	0.783	
$p_3^0$	0.523	0.929	0.998	0.818	
$p_4^0$	0.463	0.782	0.819	0.998	
...					...

Zur Auswertung der Korrelationsergebnisse muss jedem Segment  $p_v^k$  genau ein Segment  $p_\mu^{k+1}$  zugeordnet werden, falls der Korrelationswert ( $SM_{v,\mu}$ ) maximal für das betreffende Paar ist und einen unteren Schwellwert  $SM_{min}$  überschreitet. Ein Problem, das bei dieser sequentiellen Zuordnung auftreten kann, ist die Mehrfachzuweisung eines Segments. Diese Mehrdeutigkeit kann durch eine simple und schnelle Strategie, die zu einer optimierten Neuordnung führt, aufgelöst werden. Dies wird erreicht, indem die mehrfach zugewiesenen Segmente neu verteilt werden. Die Neuordnung erfolgt zu solchen Segmenten, die den nächst höchsten Korrelationswert aufweisen. Die Auflösung der Mehrdeutigkeit des Beispiels ist in Tabelle 3-5-3 dargestellt.

Eine formale Beschreibung der Neuordnung von CSC-Segmenten auf der Grundlage einer Korrelationstabelle ist in Abbildung 3-3-8 zu finden. Ist die Anzahl der Segmente

<sup>48</sup> Eine höhere Wertung ist nicht sinnvoll, da die Korrelation zwischen zwei Segmenten nicht zwingend optimal sein muss, wenn die Distanz zwischen ihnen am geringsten ist.

<sup>49</sup> Auch in anderen Arbeiten findet das Farbähnlichkeitsmaß eine höhere Gewichtung als andere Maße, beispielsweise die Formähnlichkeit [Melzer 02].

### 3. MATCHING-PARADIGMA

der Menge  $J_\nu$  kleiner als die Anzahl der Segmente der Menge  $J_\mu$ , sucht der Algorithmus nach mehrfach zugeordneten Segmenten und ordnet diese, wenn sie nicht der maximale Korrelationspartner sind, einem neuen Segment zu. Durch diesen Verarbeitungsschritt wird die eindeutige Zuordnung der Elemente der Menge  $J_\nu$  zu den Elementen der Menge  $J_\mu$  erreicht. Dieser Algorithmus erreicht die eindeutige Zuordnung von CSC-Segmenten zweier Mengen  $J_\nu$  und  $J_\mu$ . Betrachtet man einen größeren Kontext, d.h. die Menge aller MB-Matching-Paare  $M_{MB}^k$  eines Bildpaares  $I^k/I^{k+1}$ , so entspricht das Matching der zugehörigen CSC-Segmente der Zuordnung aller Segmentprimitive.

Algorithmus zur Auflösung mehrdeutiger Zuordnungen ()	
Wenn Anzahl der Spalten $\geq$ Anzahl der Zeilen tue	
für jede Zeile i tue	
Zuordnung der CSC-Patches $p_i^0$ und $p_i^1$	// siehe Tab. 3-5-3
mit maximalem Korrelationswert $SM_{ij}$	
für jede Spalte j tue	
n = Anzahl der Korrelationspartner von CSC-Patch $p_j^1$	
wenn $n > 1$	
für jede Zeile i tue	
Neuzuordnung des CSC-Patches $p_i^0$	// siehe Tab. 3-5-3
falls dieser mit $p_j^1$ korreliert ist und der	
Korrelationswert $SM_{ij}$ nicht maximal ist	
sonst	
umgekehrt	

**Abbildung 3-5-8:** Algorithmus zur Auflösung mehrdeutiger Zuordnungen

**Tabelle 3-5-3:** Beispielhafter Fall einer Mehrdeutigkeit in der Korrelationstabelle (links).  
Beispielhafte Neuordnung der Segmente (rechts)

SM	$p_0^1$	$p_1^1$	$p_2^1$	$p_3^1$	SM	$p_0^1$	$p_1^1$	$p_2^1$	$p_3^1$
$p_0^0$	0.9	0.4	0.1	0.4	$p_0^0$	0.9	0.4	0.1	0.4
$p_1^0$	0.4	0.9	0.5	0.4	$p_1^0$	0.4	0.9	0.5	0.4
$p_2^0$	0.2	0.8	0.7	0.2	$p_2^0$	0.2	0.8	0.7	0.2

Aus den somit ermittelten Korrespondenzen lassen sich nun Verschiebungsvektoren (VV) berechnen, welche die Grundlage für eine anschließend durchgeführte Bewegungsschätzung darstellen. Der Verschiebungsvektor  $\mathbf{v} = (u, v)$  errechnet sich aus der Differenz der Flächenschwerpunkte zweier korrespondierender CSC-Segmente  $p_\nu^k$  und  $p_\mu^{k+1}$  und schließt daher indirekt die Verschiebung eines dem CSC-Patch zugrunde liegenden Motion-Blobs  $b_i$  ein, da die Angabe des Schwerpunktes die Position von  $b_i$  berücksichtigt (s. Abschnitt 3.5.2). Eine separate Berechnung der Verschiebungsvektoren von MB-Regionen ist hingegen nicht sinnvoll, da diese nur eine grobe Verschiebungsrichtung angeben, jedoch keine Berechnung eines feingliedrigen Verschiebungsvektorfeld (VVF) erlauben.

#### 3.5.3 Gruppierung der ermittelten Verschiebungsvektoren

Durch Nutzung des vorgeschlagenen bewegungsbasierten MD-Ansatzes werden die Bildbereiche, die potenziell bewegte Objekte enthalten auf die MBs reduziert. Problematisch bleibt dennoch, dass erstens durch die CSC-Segmente eine Gruppierung der Verschiebungsvektoren erforderlich ist und zweitens ohne eine Modellannahme keine Aussage darüber getroffen werden kann, wie viele Objekte tatsächlich innerhalb eines segmentierten MBs enthalten sind. Infolgedessen besteht das Problem, das zunächst einmal definiert werden muss, wodurch ein Objekt charakterisiert wird. Dazu ist besonders bei der Untersuchung von Verkehrsaufnahmen mit dicht fahrenden PKWs damit zu rechnen, dass verschiedene Objekte zu einem MB verschmolzen werden und dadurch Anhaltspunkte über die Regionentrennung durch das zeitliche Verhalten oder Objektbeschreibung schwierig herauszufinden sind. Dies motiviert erstens die Gruppierung von Verschiebungsvektoren der CSC-Segmente und zweitens die Umsetzung einer weiteren Ebene der Objektbeschreibung, die die Korrelationsinformation der Matchingebenen I und II im zeitlichen Kontext ausnutzt.

Die Gruppierung von Verschiebungsvektoren der CSC-Segmente kann prinzipielle durch ein Homogenitätskriterium erfolgt werden, das die Länge, Orientierung und räumliche Lage der Verschiebungsvektoren in Betracht zieht. Die Eingangsdaten der Gruppierung bestehen aus optischen Flussvektoren zwischen aufeinander folgenden Bildern, die die aktiven Trajektorien bilden. Das heißt, dass diese Trajektorien sich aus einem Satz aneinander geketteter Verschiebungsvektoren der einzelnen aktiven CSC-Segmente in Motion-Blobs zusammensetzen. Die aktiven Trajektorien deuten auch darauf hin, dass das jeweilige korrespondierende CSC-Segment in der Zeitspanne nicht verschwunden ist. Während der optische Fluss die momentane Segmentbewegung wiedergibt, beinhalten Trajektorien Bewegungsinformation über einen gewissen Zeitraum (aktiver Zeitraum) und ermöglichen somit eine differenzierte Objektrennung durch eine weitere Meta-Objekt-Korrelation. In dieser Arbeit müssen die Segmente folgende Voraussetzungen erfüllen, damit sie gruppiert und anschließend zu einem Objekt zugeordnet werden können:

- Ein Objekt muss aus mehreren Segmentregionen (>6 Segmente) bestehen.
- Ein Objekt muss sich aus Segmentregionen zusammensetzen, deren umschreibende Rechtecke sich überlappen.
- Die Objekte müssen eine signifikante Bewegung in der Bildebene aufweisen und die *aktiven* Trajektorien der Segmente eines Objektes müssen sich in ihren Verläufen ähneln.

Durch die erste Forderung verhindert man, dass einzelne CSC-Segmentregionen, deren Trajektorien fehlerhaft bestimmt wurden, zu falschen Objekthypothesen führen. In den realen Sequenzen hat sich gezeigt, dass Objekte, die zur Gruppe der Verkehrsteilnehmer gehören, zumeist in mehr als 6 Segmentregionen aufgeteilt werden. Die Tatsache, dass die Oberfläche eines kompakten Szenenobjekts auf *eine* Fläche aus zusammenhängenden Segmentpunkten projiziert wird, führt dazu, dass Segmentregionen, 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 somit Lücken zwischen den verbleibenden Farbsegmenten eines Objektes entstehen. Um in solchen Fällen trotzdem eine relativ robuste Gruppierung der

### 3. MATCHING-PARADIGMA

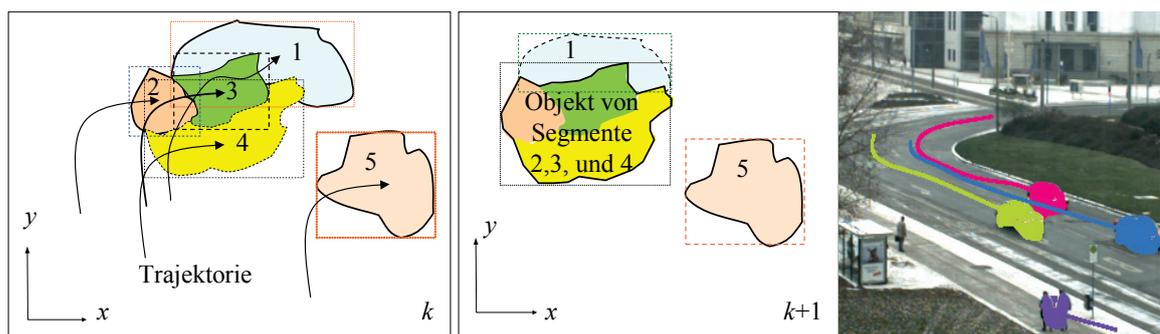
Segmentregionen zu ermöglichen, wird die Forderung nach gemeinsamen Regionengrenzen abgeschwächt und lediglich eine Überlappung der umschreibenden Rechtecke der Regionen gefordert (zweite Forderung). Zusätzlich wird durch die dritte Voraussetzung verlangt, dass die aktiven Trajektorien der Farbregionen eine gegebene Mindestlänge überschreiten.

Des Weiteren wurde in allen Untersuchungen<sup>50</sup> eine minimale Schwerpunktverschiebung von 9 Bildpunkten innerhalb der letzten drei Bilder gefordert. Dies entspricht einer mittleren minimalen Bild-zu-Bild Verschiebung von 3 Bildpunkten. Die Ähnlichkeit der Bewegungen zweier aktiver Segmente kann auch durch die Korrelation ihrer Trajektorien bestimmt werden [Al-Hamadi 06b]. Die Attribute der Korrelationen können die Länge, Orientierung und Lage beinhalten. Der Übereinstimmungswert zweier Trajektorien  $T_1$  und  $T_2$  wird hierbei prinzipiell durch den Vergleich der Winkel  $\varphi_1$  und  $\varphi_2$  bzw. der Vektorlänge  $l_1$  und  $l_2$  errechnet. Die aus dem Vergleich berechnete Ähnlichkeit des Winkels ergibt sich aus Gleichung (3-5-9b) und die der Länge aus Gleichung (3-5-9a).

$$\psi(T_1, T_2) = \left(1 - \frac{|l_1 - l_2|}{l_1 + l_2}\right) \cdot \frac{\sum_{k=k_A}^{k_E} (\varphi_1(k) - \bar{\varphi}_1)(\varphi_2(k) - \bar{\varphi}_2)}{\sqrt{\left(\sum_{k=k_A}^{k_E} (\varphi_1(k) - \bar{\varphi}_1)^2\right) \cdot \left(\sum_{k=k_A}^{k_E} (\varphi_2(k) - \bar{\varphi}_2)^2\right)}} \quad (3-5-9a)$$

$$\varphi_i = \arctan\left(\frac{v_y}{v_x}\right), \quad l = \sqrt{v_x^2 + v_y^2} \quad \text{und} \quad l_i = \sum_{k=k_A}^{k_E} (l(k)) \quad (3-5-9b)$$

Das Intervall  $[k_A, k_E]$  gibt den Ausschnitt aus der Bildsequenz vor, der zum Vergleich der Trajektorien verwendet wird, indem beide Trajektorien existieren. Der erste Faktor in Gleichung 3-5-9a berücksichtigt Abweichungen in der Länge der Trajektorien. Der zweite Faktor ist die lineare normierte Kreuzkorrelationsfunktion der Trajektorien-Richtung ( $\varphi_i$ ). Sie wird 1, wenn die Trajektorien bis auf eine Verschiebung oder einen Skalierungsfaktor gleich sind.



**Abbildung 3-5-9:** Das Prinzip des Gruppierungsprozesses. Segmente werden zu einem Objekt geordnet. Dies basiert auf ihren Attributen, die von den Trajektorien abgeleitet sind. Die Segmente 2, 3 und 4 sind aufgrund der Ähnlichkeit ihrer Trajektorien zu einer Objekthypothese zusammengefasst. Region 5 hat auch ähnliche Eigenschaften allerdings stimmt die räumliche Lage nicht. Auf der rechten Seite ist eine endgültige Gruppierung mit der Gesamttrajektorie zu sehen.

<sup>50</sup> Bei stehender oder langsam bewegter Kamera können in der Regel damit die Regionen bewegter Szenenobjekte von denen der unbewegten Szenenteile getrennt werden. Bei Eigenbewegung der Kamera kann nur die Forderung von Objekten erfüllt werden, die eine signifikante Relativbewegung zur Kamera besitzen.

### 3. MATCHING-PARADIGMA

Bei Gegenläufigkeit der Trajektorien wird sie in der Regel zu -1. Eine Korrelation von 0 ergibt sich, wenn die Trajektorien der Regionen zu jedem Zeitpunkt orthogonal zueinander verlaufen. Um zu entscheiden, ob zwei Segmente  $P_1$  und  $P_2$  zum selben Objekt gehören, wird geprüft, ob die  $\psi$ -Ähnlichkeitsfunktion (Gl. 3-5-9a) einen gegebenen Schwellwert überschreitet. Dadurch wird eine fundamentale Annahme zur Gruppierung von aktiven Segmenten auf der Basis der Verschiebungsvektordinformationen abgeleitet (s. Abb. 3-5-9). Diese, für die nächste Meta-Objekt-Korrelation (MOK) fundamentale Annahme, leitet sich aus der Tatsache ab, dass die Punkte eines Szenenobjektes physikalisch gekoppelt sind und somit ihre Segmente und Bewegungen voneinander abhängen.

#### 3.5.4 Erweiterungsmöglichkeiten auf eine Meta-Korrelation

Angestrebtes Ziel der Erweiterung einer Meta-Objekt-Korrelation (MOK) ist es je einem MB der Szene genau ein Meta-Objekt (MO) zuzuordnen, welches das reale Objekt über die gesamte Sequenz hin denotiert, d.h. es ermöglicht das reale Objekt zu verfolgen und zu beschreiben auch bei Verschmelzung oder Teilverdeckung. Hier können die Meta-Objekte analytisch erfasste Abbildungen von Objekten der realen Szene sein, die sich während der Sequenz fortbewegen und zum größeren Teil von den Segmenten abhängig sind. Konzeptionell kann deshalb eine Dreiebeneinteilung stattfinden, bei der unterschieden wird zwischen a) realen Objekten, die eine Bewegung ausführen, b) den Abbildungen dieser Objekte, aus welchen die Elemente der Merkmalsebene I und II hervorgehen und c) einer abstrakten Objektbeschreibung, welche als Meta-Objekt bezeichnet werden soll [Al-Hamadi 07b]. Infolgedessen stellen die Meta-Objekte einen abstrakten Bedeutungsträger dar und werden aus den Bewegungskomponenten der erfassten Segmente generiert.

Allerdings enthält ein MB nicht notwendigerweise nur ein MO, sondern im Allgemeinen n-viele. Außerdem kann es durch Einschränkungen in der Vorverarbeitung passieren, dass ein Element der realen Szene durch mehrere MOs denotiert wird, was ein Problem darstellt, das ohne entsprechendes Vorwissen nicht behoben werden kann. Nun besteht die Aufgabe der MB- und CSC-Zuordnung darin, MOs zu generieren und sie durch den Szenenverlauf zu verfolgen, indem sie aus den bildinhaltsändernden Regionen, also den MBs einzelner Bildpaare extrahiert bzw. zugeordnet werden.

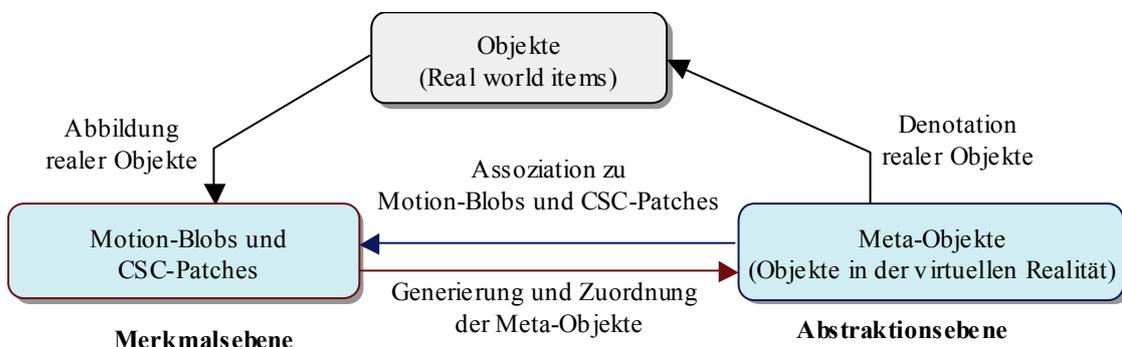


Abbildung 3-5-10: Konzeptionelle Teilung des Tracking Paradigmas in drei Komponenten

Um die MO-Zuordnung zu realisieren, besitzen MBs und MOs verschiedene Attribute. Dazu gehören u.a. Konturpunkte, Größenangaben und die durch Bewegungstrajektorien zusammengefassten CSC-Patches, welche Unterregionen in MBs darstellen. Die initiale

### 3. MATCHING-PARADIGMA

---

Annahme dieser MO-Zuordnung ist, dass jeder MB ein *Meta-Objekt* darstellt. Diese unwahrscheinliche Annahme der Situation wird im Verlauf der Sequenz dem realen Sachverhalt möglichst genau angeglichen. So werden Meta-Objekte die initial zu einem MB zugeordnet sind auch nach der Verschmelzung des MBs weiterverfolgt, indem die unabhängig korrelierten und mit den Meta-Objekten assoziierten Farb-Patches, betrachtet werden. Zerfällt hingegen ein MB der mit nur einem Meta-Objekt verknüpft war, so wird nun die Existenz von zwei Meta-Objekten angenommen. Da dies im ungünstigsten Fall aber auch durch eine Fehlsegmentierung passieren kann, muss im weiteren Verlauf der Sequenz getestet werden, ob ein derart geteiltes MO wieder verschmilzt. Der gesamte Prozess der Objektverfolgung ist also dynamisch gestaltet.

Aus dieser Situation resultiert die Aufgabe ein spezielles MO ( $o_i$ ) in einem MB  $b_j$  zu finden. Somit basiert die MO-Zuordnung auf der Grundlage von MO- und MB-Attributen, d.h. auf dem Vergleich von Eigenschaften der CSC-Farbsegmente, wie der Kontur, Position, Größe und Bewegungsinformation. Die Eigenschaften der MOs werden mit jedem Frame aktualisiert, um sie mit den folgenden MBs vergleichen zu können.

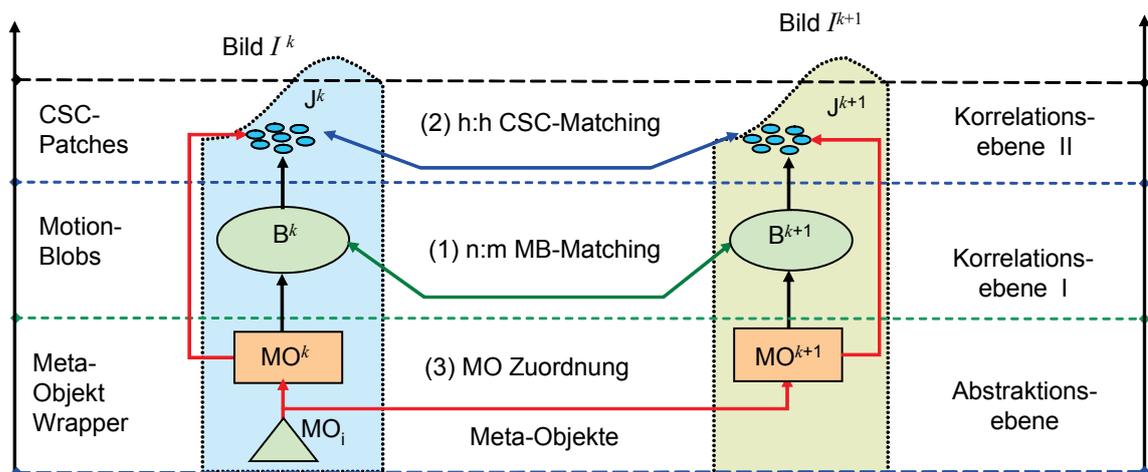
Im Allgemeinen bedeutet das, es werden  $k$  von  $n$  vielen Blobs  $b_{1i}$ , die aus einem zerfallenden Blob  $b_0$  entstehen, je einem neuen *Objekt* zugeordnet, wenn  $b_0$  mit weniger *Objekten* assoziiert wurde als es  $n-k$  viele Blobs  $b_{1i}$  gibt. Im Anschluss wird die Information über die Art der Beziehung korrelierter MBs, d.h. den entsprechenden Zuordnungsfall genutzt, um alle MOs  $o_i$  den jeweils relevanten MBs  $b_j$  zuzuordnen und zu verfolgen. Kurz betrachtet entstehen die folgenden Situationen:

- 1). *Ausgangssituation*: Wie zuvor beschrieben, wird in der Initialisierungsphase der MO-Korrelation angenommen, dass jeder MB ( $b_i$ ) genau ein MO ( $o_i$ ) repräsentiert, d.h. jedem MB wird ein neues MO zugeordnet.
- 2). *Auftreten eines neuen MO*: Wird bei der Blob-Korrelation ein neuer MB ohne auffindbaren Vorgänger entdeckt (0:1 Zuordnung), wird diesem genau ein neues MO zugeordnet.
- 3). *Auftreten eines neuen MO durch Zerfall eines MB*: Allen MBs  $b_{1i}$ , die aus einem zerfallenden MB ( $b_0$ ) entstehen, wird je ein neues MO ( $o_j$ ) zugeordnet, wenn  $b_0$  mit nur einem MO assoziiert wird. Im Allgemeinen gilt, dass zu jedem MB ( $b_{1i}$ ), der aus einem zerfallenden MB ( $b_0$ ) entsteht, ein neues MO zugeordnet wird, falls bereits alle in  $b_0$  enthaltenen MOs, anderen MBs ( $b_{1j}$ ) zugeordnet sind. Das Prinzip der Zuordnung wird im Anschluss genauer beschrieben.
- 4). *Zuordnung vorhandener MOs bei Zerfall eines MB*: Zerfällt ein MB  $b_0$ , welcher mit  $n$  MOs assoziiert wird in  $m$  viele MBs  $b_{1i}$  ( $m \leq n$ ), so wird eine dem Zuordnungsprinzip entsprechende Verknüpfung von MBs zu den MOs durchgeführt.
- 5). *Zuordnung vorhandener MOs bei Verschmelzung eines MB*: Verschmelzen  $n$  MBs ( $b_{0i}$ ) zu einem MB ( $b_1$ ), so werden sämtliche MOs, die mit  $b_{0i}$  assoziiert sind,  $b_1$  zugeordnet.
- 6). *Zuordnung vorhandener MOs bei simultaner Teilung und Verschmelzung mehrere MBs*: Wird dieser Fall bei der Blob-Korrelation entdeckt, werden sämtliche mit  $b_{00}$  und  $b_{01}$  assoziierten MOs nach dem Zuordnungsprinzip an ( $b_{10}$ ) und ( $b_{11}$ ) weitergeleitet.

### 3. MATCHING-PARADIGMA

7). Zuordnung eines vorhandenen MO bei einfacher Fortbewegung eines MB: Wird im 1:1 Fall ein MB einfach weiterbewegt, so wird auch das dem MB zugehörige MO einfach weiterbewegt. Gehören einem MB mehrere MOs an, wird nach dem Zuordnungsprinzip jeweils die individuelle Verschiebung berechnet.

Abbildung (3-5-11) zeigt die Beziehung zwischen hierarchischen Merkmalskorrelationen in Merkmalsebene I und II. Die Schnittstelle zur zweiten Ebene dieser Technik und damit zur Korrespondenzermittlung, soll durch sog. Meta-Objekte MO erfolgen, die Objekte der realen Welt in der gesamten Sequenz bezeichnen. Dazu wird möglichst einem realen Objekt ein MO zugeordnet. Die zweite Ebene des Systems (Abstraktions-ebene) soll nun greifen, Meta-Objekte werden korreliert, indem ihnen jeweils aktuelle Farb-Patches zugeordnet werden.



**Abbildung 3-5-11:** Die Beziehung zwischen hierarchischen Merkmalskorrelationen in den unterschiedlichen Merkmalsebenen.

Eine Besonderheit der Meta-Objekt-Korrelation ist die feine Gruppierung von CSC-Regionen durch die definierten Attribute. Der Vorteil dieser Vorgehensweise ist eine deutliche Vereinfachung und damit Beschleunigung des Gruppierungsvorgangs. Durch die Trennung von Bildmerkmalen (MBs und CSC-Patches) und damit assoziierten Bedeutungsträgern (Meta-Objekte) werden Regionen erst im zeitlichen Verlauf sicher erkannt und verfolgt. Generell könnten zur Verbesserung der Fehlerrobustheit zusätzliche Parameter in den Prozess der Meta-Objekt Detektierung mit einfließen, wie z.B. Annahme einer bestimmten Objektform bzw. Kontur in Kombination mit Homogenitätseinschränkungen bzgl. der Farbe. Diese Erweiterung führt zu einer Verbesserung der Objektbeschreibung und Erkennung.

**Diskussion:** Die Bewegung eines Objektsegmentes beeinflusst die Regionen der umgebenden Motion-Blobs, indem das Objektsegment eines Motion-Blobs Bereichsegmente, die zum einen anderen MB gehören, überdeckt. Dadurch werden Segmente, die sich bezogen auf die Bewegungsrichtung, vor einer Region befinden, kleiner und verschwinden. Dahinter werden sie größer oder sie entstehen neu. Dadurch kann es sogar sein, dass der Schwerpunkt des Segmentes sich scheinbar durch die Bewegung einer relativ großen Region in die entgegengesetzte Richtung bewegt. Genauso wie Regionen, die (teilweise) von anderen überdeckt werden, stellen auch Regionen, die sich am Bildrand befinden ein Problem dar. Da davon auszugehen ist, dass die Regionen in der Regel außerhalb des im Bild sichtbaren Bereich noch fortgesetzt sind, entsprechen die für den im Bild sichtbaren Teil der Regionen bestimmten Bewegungs-

vektoren nicht dem tatsächlichen Bewegungsvektoren, die für diese Regionen ermittelt würden, wenn sie vollständig im Bild zu sehen wären. Beide Probleme zeigen sich auch im Rahmen der Experimente und sind dadurch bedingt, dass jeweils ein Teil der betrachteten Objekte im Bild nicht sichtbar ist, was letztendlich bereits durch das Blendenproblem beschrieben ist.

Die *bisher* verwendeten Gewichte sind einerseits aufgrund der Funktionsweise der einzelnen Ähnlichkeitsmaße erklärbar und wurden andererseits durch die Ergebnisse, die in einer Vielzahl von Experimenten erzielt wurden, bestätigt. Die Gewichte können mit Hilfe eines zweistufigen Verfahrens aus dem nächsten Abschnitt optimiert werden, die auf ausgewählte Referenzsequenzen angewandt wurden. Im ersten Schritt werden die Bilder der Referenzsequenz unabhängig voneinander untersucht. Um festzustellen, in welchen Größenordnungen die Einzelähnlichkeitskriterien bei einer falschen Zuordnung liegen, wird jedes Segment mit den sie umgebenden Segmenten im selben Bild verglichen. Für jeden Vergleich können die Einzelfehler eingeschätzt und festgehalten werden. Im zweiten Schritt wurde nun untersucht, wie sich die Merkmale in aufeinander folgenden Bildern verhalten. Im Unterschied zum ersten Schritt, wird ein Segment mit Segmenten des Folgebilds verglichen. Die Menge der zum Vergleich herangezogenen Segmente wird wieder durch einen kreisförmigen Umgebungsbereich, jetzt aber im darauf folgenden Bild, eingeschränkt. Die Einzelfehler werden mit den im ersten Schritt bestimmten Gewichten optimiert. Diese Vorgehensweise erfordert keine Kenntnisse über die korrekten Segmentzuordnungen und lässt sich auf beliebige Segmentierungsverfahren anwenden. Mehr Details über die Optimierung kann im nächsten Abschnitt 3.6 gefunden werden.

Das Verschiebungsvektorfeld (VVF), welches durch das vorgestellte Merkmalsmatching in zwei Ebenen erzielt wird, weist robuste und nützliche Eigenschaften auf, die im folgenden Kapitel anhand verschiedener Situationen beschrieben werden. Allerdings haben leichte Unterschiede in den Segmentierungsergebnissen einen relativen Einfluss auf die Schwerpunktverschiebung der Farbregionen, wodurch kleine Fehler in der Bestimmung der Bewegungsvektoren hervorgerufen werden. Aus diesem Grund sind gewisse Schwankungen der für korrespondierende Regionengruppen ermittelten Bewegungsvektoren zu erwarten. Eine Glättung der Verschiebungsvektoren über die Bildfolge konnte zwar auf der Ebene von Verschiebungsvektoren realisiert werden, nicht aber auf der Ebene von Segmenten. Die Kenntnis, dass Verschiebungsvektoren, die stets dasselbe Szenenobjekt repräsentieren, in der Regel nur geringfügigen Bild-zu-Bild Änderungen unterliegen, konnte in der Zusammenfassung dieser Verschiebungsvektoren unmittelbar genutzt werden.

#### 3.6 Verbesserte Störsicherheit durch Optimierungsschritte

In dem vorherigen Abschnitt wurde beschrieben, wie die Zuordnung der CSC-Segmente von den vier verschiedenen merkmalspezifischen Ähnlichkeitsmaßen abhängig ist. Dabei wird die Wirkungsintensität eines merkmalspezifischen Ähnlichkeitsmaßes auf das Multi-Matching durch den Gewichtsvektor  $w=(w_1 w_2 w_3 w_4)^T$  reguliert. Bei gleichen Ähnlichkeitswerten und verschiedenen Gewichten kann also ein unterschiedliches CSC-Matchingergebnis auftreten.

### 3. MATCHING-PARADIGMA

Die Aufgabe besteht nun darin, denjenigen Gewichtsvektor zu finden, der die Ähnlichkeitsmaße so gewichtet, dass das CSC-Matching<sup>51</sup> eine optimale Lösung liefert. Um dies zu erreichen, wird eine Gütefunktion in Abhängigkeit von den Gewichtsvektoren  $\boldsymbol{w}$  aufgestellt, die ein optimales CSC-Matching ausdrückt, wenn sie ihr Minimum<sup>52</sup> erreicht hat. Dafür wird vor allem ein Datensatz<sup>53</sup> mit entsprechenden Grundwahrheiten benötigt, der die korrekten CSC-Paare beinhaltet. Die Schwierigkeit dabei ist es, dass die korrekten CSC-Segment-Zuordnungen nicht ermittelt werden können, weil einerseits die Anzahl der CSC-Segmente viel zu groß ist, um sie per Hand zu markieren und andererseits selbst per Hand die korrekten CSC-Paare nicht eindeutig zu ermitteln sind, da einige CSC-Segmente sich über die Zeit instabil verhalten. Aus diesem Grund werden die korrekten CSC-Paare indirekt über die bekannten Bewegungsinformationen der CSC-Schwerpunkte eingeschätzt. Dazu stehen künstlich erzeugte Bildsequenzen, in denen die Verschiebungsvektoren aller CSC-Segmente bekannt sind und verschiedene reale Bildsequenzen, in denen die Segmentverschiebungen mit dem PLK-Verfahren (s. Kapitel 2) bestimmt werden, zur Verfügung.

Die Bewegungsinformation eines CSC-Segmentes wird durch den Verschiebungsvektor  $\mathbf{v}_c(\mathbf{p}_v^k)$  beschrieben. Die Verknüpfung des Schwerpunktes eines Segmentes  $\mathbf{p}_v^k$  mit allen möglichen Segmenten  $\mathbf{p}_\mu^{k+1}$ , liefert den Bewegungsvektor  $\mathbf{v}_e(\mathbf{p}_v^k, \mathbf{p}_\mu^{k+1})$ , für das CSC-Paar  $\{\mathbf{p}_v^k, \mathbf{p}_\mu^{k+1}\}$ . Der Fehler dieses CSC-Paares lässt sich durch das *SRE*-Fehlermaß in Anlehnung an Kapitel 2 aus der Gleichung (3-6-1a) ermitteln.

$$SRE(\mathbf{p}_v^k, \mathbf{p}_\mu^{k+1}) = \frac{\|\mathbf{v}_e(\mathbf{p}_v^k, \mathbf{p}_\mu^{k+1}) - \mathbf{v}_c(\mathbf{p}_v^k)\|}{\|\mathbf{v}_c(\mathbf{p}_v^k)\|} \quad (3-6-1a)$$

Das *SRE*-Fehlermaß bildet somit aus der bekannten Verschiebungsinformation aus der Grundwahrheit ein Maß für die Bewertung eines CSC-Paares. Je besser die Verschiebungsvektoren übereinstimmen, desto kleiner wird das *SRE*-Fehlermaß. Bei einer kompletten Übereinstimmung, die sich als richtig zugeordnetes CSC-Paar interpretieren lässt, ist der *SRE* Null. Die daraus resultierenden Daten liegen nun in Form einer Zuordnungsmatrix  $\mathbf{S}_c$  vor, die die korrekten CSC-Zuordnungen enthält. Die Matrix  $\mathbf{S}_c$ , auch Korrelationstabelle der Grundwahrheit genannt, wird aus den *SRE*-Maßen und damit aus den bekannten Bewegungsvektoren  $\mathbf{v}_c(\mathbf{p}_v^{k=0})$  gebildet, die zunächst in den Vektoren  $\mathbf{EC}_i$  zusammengefasst werden (Gl. (3-6-1b)).

$$\mathbf{EC}_v = (SRE(\mathbf{p}_v^0, \mathbf{p}_0^1), \dots, SRE(\mathbf{p}_v^0, \mathbf{p}_{h-1}^1))^T \quad (3-6-1b)$$

$$\mathbf{S}_c = \begin{pmatrix} 0 & \mu & \forall \mu = 0 \dots M-1, & SRE(\mathbf{p}_0, \mathbf{p}_\mu) = \min(\mathbf{EC}_0) \\ \vdots & \vdots & \vdots & \\ h-1 & \mu & \forall \mu = 0 \dots M-1, & SRE(\mathbf{p}_{h-1}, \mathbf{p}_\mu) = \min(\mathbf{EC}_{h-1}) \end{pmatrix} \quad (3-6-1c)$$

In der Praxis stellt sich allerdings heraus, dass durch Zerfall oder Verschmelzung der Segmente, einem Element der CSC-Menge  $J_0$ , mehreren Elementen der nachfolgenden

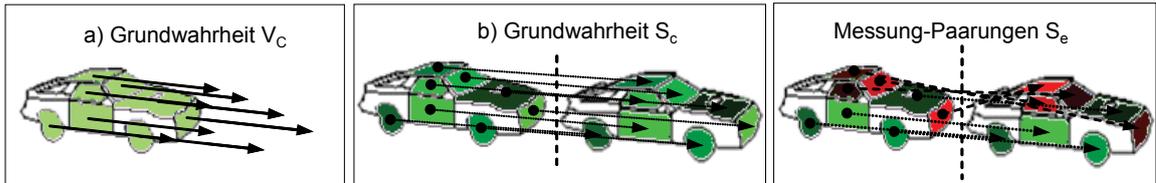
<sup>51</sup> Das CSC-Matching wird durch das definierte SM-Ähnlichkeitsmaß im Abschnitt 3.5 realisiert.

<sup>52</sup> Die Optimierung der Gewichte stellt sich durch die Korrelation als ein Minimierungsproblem dar, das sich im einfachsten Fall lokal um einen Punkt in eindeutiger Weise linear approximieren lässt.

<sup>53</sup> In der Optimierungsdatenbasis sind Daten der Grundwahrheit und die Messdaten enthalten.

### 3. MATCHING-PARADIGMA

CSC-Menge  $J_1$  zugeordnet werden können. Diese mehrdeutigen Zuordnungen werden bis auf das CSC-Paar mit dem minimalen  $SRE$  entfernt. Dadurch ist eine eindeutige CSC-Zuordnung gewährleistet. Im Weiteren Schritt erfolgt die Bestimmung der Zuordnungsmatrix  $S_e(\mathbf{w})$  für jedes Segmentpaar  $\{p_v^k, p_\mu^{k+1}\}$  durch die aktuellen Gewichte nach Gleichung 3-5-8. Aus der Variation des Gewichtsvektors ergibt sich ein variierendes SM-Gesamtähnlichkeitsmaß, mit diesem die Zuordnungsmatrix des jeweiligen Gewichtsvektors  $S_e(\mathbf{w})$  bestimmt wird. Eine Messung bezeichnet also die Zuordnungsmatrix  $S_e(\mathbf{w})$  als CSC-Matchingresultat für einen bestimmten Gewichtsvektor  $\mathbf{w}$  und feste Ähnlichkeitsmaße  $S_x$ <sup>54</sup>.



**Abbildung 3-6-1:** Beispiel für die Optimierungsdatenbasis. Mit (●) den CSC-Segment zum Zeitpunkt  $k$  und (○) den CSC-Segmenten zum Zeitpunkt  $k+1$ . Die Grundwahrheit  $S_c$  (b) wird durch ein Matching über die Bewegungsvektoren der CSC-Schwerpunkte (a) gewonnen. Der Vergleich der Grundwahrheit  $S_c$  und der Messung für einen bestimmten Gewichtsvektor mit  $S_e$  werden richtige (grün) und falsche (rot) Korrespondenzen erkannt und können zu einem Gütekriterium zusammengefasst werden. Die Matrizen  $S_e$  und  $S_c$  bilden so die Datenbasis für das Optimierungsverfahren.

#### 3.6.1 Gütefunktion zur Gewichtsoptimierung

Eine Gütefunktion  $E(\mathbf{w})$  soll nun durch den Vergleich der Zuordnungsmatrizen  $S_c$  und  $S_e$  definiert werden. Diese besteht aus einem Zuordnungsfehler und einem Restfehler. Der Zuordnungsfehler (erster Term in Gleichung 3-6-2d) setzt sich aus der Anzahl der richtig zugeordneten Paarungen<sup>55</sup>  $|Pa_r|$  und der Anzahl der falsch zugeordneten Paarungen  $|Pa_f|$  zusammen.

Der Restfehler  $fSRE$  gibt die falsch zugeordneten Paarungen normiert auf den maximalen SRE-Fehler (siehe Gleichung 3-6-2d) an. Je geringer der Restfehler also ist, desto näher befinden sich die falschen Matchingpartner bei den korrekten. Auf diese Weise gibt die formulierte Gütefunktion  $E(\mathbf{w})$  in Gleichung (3-6-2d) nicht nur an, wie viele Segmente richtig zugeordnet werden, sie beschreibt auch wie sehr die falsch zugeordneten Paarungen von dem Idealmatching entfernt liegen. Durch die Erzeugung der korrekten und gemessenen Zuordnungsmatrizen  $S_c$  und  $S_e$  über den Minimum-Operator (im Fall  $S_c$ ) und Maximum Operator (im Fall  $S_e$ ) und der damit verbundenen Berechnung der Kardinalitäten der Mengen  $Pa_r$  und  $Pa_f$  ist die Gütefunktion  $E(\mathbf{w})$  eine nichtstetige Funktion.

$$Pa_r = \bigcup_{v,\mu} \{p_v^k, p_\mu^{k+1}\}, \forall (v, \mu) \in (S_e \cap S_c) \quad (3-6-2a)$$

$$Pa_f = \bigcup_{v,\mu} \{p_v^k, p_\mu^{k+1}\}, \forall (v, \mu) \in (S_e \setminus S_c) \quad (3-6-2b)$$

<sup>54</sup>  $S_x$  bezeichnet die Kurzform stellvertretend für jedes einzelne merkmalsbasierte Ähnlichkeitsmaß  $SL_{POS}$ ,  $SI_{DIST}$ ,  $SC_{RGB}$  und  $SL_{SIZE}$ .

<sup>55</sup> Die Anzahl der Elemente einer Menge wird als Kardinalzahl bezeichnet und für eine Menge  $P$  mit  $|P|$  angegeben. Dabei ist die Menge  $Pa_r$  die Menge der korrekt zugeordneten Paarungen und die Menge  $Pa_f$  die Menge der falsch zugeordneten Paarungen.

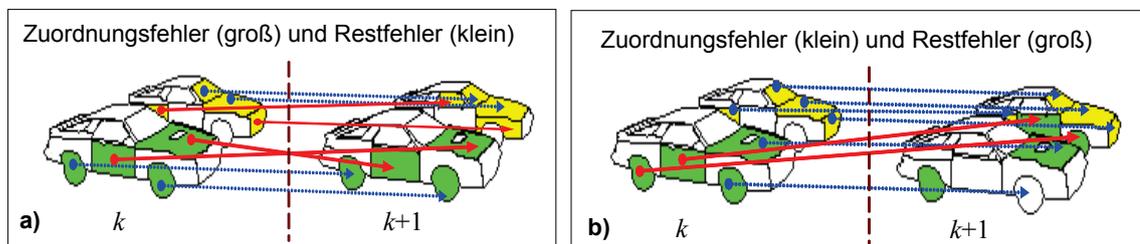
### 3. MATCHING-PARADIGMA

$$fSRE = \frac{\sum_{v \in Pa_f} \sum_{\mu \in Pa_f} SRE(p_v^k, p_\mu^{k+1})}{|Pa_f|} \quad (3-6-2c)$$

$$E(\mathbf{w}) = \frac{1}{2} \left( 1 - \frac{|Pa_r| - |Pa_f|}{|Pa_r| + |Pa_f|} \right) + \frac{fSRE}{\max_{v, \mu}(SRE(p_v^k, p_\mu^{k+1}))}, \text{ mit } E(\mathbf{w}) \in [0, 2] \quad (3-6-2d)$$

Der Wertebereich, der beiden Summanden ist jeweils auf das Intervall  $[0, 1]$  normiert, so dass sich der Wertebereich für das Gesamtgütemaß im Intervall  $[0, 2]$  befindet, wobei bei einem Wert von Null das ideale Matching erreicht wird, d.h. alle Segmente sind korrekt und bei einem Wert von zwei, dementsprechend alle Zuordnungen fehlerhaft zugeordnet sind.

Die Abbildung 3-6-2 verdeutlicht warum es wichtig ist auch den Restfehler (zweiter Term in Gleichung 3-6-2d) zu betrachten. Einerseits soll mit dem Zuordnungsfehler überwacht werden, wie viele CSC-Paare richtig zugeordnet werden. Da bei einer Erhöhung der falsch zugeordneten Paare die Wahrscheinlichkeit steigt, dass sich die CSC-Segmente verschiedener Objekte vermischen. Es ist außerdem wichtig zu überwachen, wie falsch die Segmente zugeordnet sind. Das bedeutet, wenn sie sich in der Nähe des korrekten Matchingpartners befinden, wie im oberen Bild zu sehen, ist die Wahrscheinlichkeit weniger gering, dass sich die CSC-Segmente zweier Objekte vermischen, als wenn sie sich weiter von ihrem korrekten Matchingpartnern entfernt befinden (Abb. 3-6-2b). Dies wird durch den Restfehler in Form des *SRE*-Fehlermaßes abgebildet. Je geringer also der Restfehler ist, desto näher befinden sich die falschen Matchingpartner bei den Korrekten.



**Abbildung 3-6-2:** Die Darstellung eines CSC-Matchings mit Fehlerbetrachtung. Der Zuordnungsfehler ergibt sich aus den richtig zugeordneten CSC-Paaren (blau) und den falsch zugeordneten Paaren (rot). Die Abbildung b) zeigt, dass bei einem kleinen Zuordnungsfehler, die Bewegungsverfolgung durch die Vermischung der Objekte gestört werden kann. Deshalb ist es zusätzlich notwendig den Restfehler der falsch zugeordneten CSC-Paare zu betrachten.

Um die optimalen Gewichtskonfigurationen  $\mathbf{w}$  nun einschätzen zu können, muss die definierte Gütefunktion  $E(\mathbf{w})$  in Gleichung (3-6-2d) minimiert werden. Ordnet man die verschiedenen Gewichtskonfigurationen nach definierte Nachbarschaftsstruktur, so ergibt sich ein Suchraum. Durch diesen Suchraum bewegt man sich in der Regel schrittweise während der Minimierung der definierten Gütefunktion. Anschaulich bewegt man sich während der Optimierung von Punkt zu Punkt dieses Suchraumes. Ordnet man jedem dieser Punkte die entsprechende Gütefunktion zu, erhält man die sog. Hügel-Täler-

Landschaft. Dafür gibt es eine Vielzahl von Lösungen<sup>56</sup>, so dass ein einfaches Durchprobieren keinen Erfolg verspricht. Um den Optimierungssuchraum mit guten Resultaten in kurzer Zeit zu durchwandern, bedient man sich verschiedener Heuristiken. Diese unterscheiden sich hauptsächlich in der Wahl der Akzeptanzregel für die Annahme einer neu erzeugten Konfiguration. Hier wird die Minimierung mit einem stochastischen Optimierungsverfahren den *Threshold-Accepting-Algorithmus* (Toleranzschwellenalgorithmus) durchgeführt.

#### 3.6.1.1 *Threshold-Accepting-Algorithmus*

Der Threshold-Accepting-Algorithmus (TAAs) stellt eine Abwandlung des *Simulated Annealing Algorithmus* (SAA) dar [Pham 00, Arts 91, Dueck 90], der für die Minimierung der Funktion  $E(\mathbf{w})$  gut geeignet ist. Die Grundidee des Simulated Annealing Algorithmus stammt aus der Physik und bildet den Prozess des langsamen Abkühlens von Metallen nach, bei dem sich die Metallatome so anordnen, dass sie nach dem Abkühlen einen Zustand minimaler Energie einnehmen. Der Prozess des Abkühlens wird durch die Boltzmannverteilung modelliert, die durch einen Schwellwert  $H$  vereinfacht wird.

Von einem *initialen* Anfangswert werden, wie bei der lokalen Zufallsuche, Konfigurationen in der Umgebung generiert. Eine Konfiguration dabei ist eine mögliche Lösung der Optimierungsfunktion, die aber nicht notwendigerweise aller Nebenbedingungen einhalten muss. Die Menge aller Konfigurationen bildet den Konfigurationsraum. Aufgrund der vielen Konfigurationen spricht man von einem hochdimensionalen Raum. Dort wird die Qualität der lokalen Lösung gemessen. Ist diese besser als die Qualität der Referenzlösung, nimmt sie den Platz der Referenzlösung ein. Ist die Qualität dieser Konfiguration schlechter als die der vorhergehenden, kann mit Hilfe des Schwellwertparameters  $H$  bestimmt werden, ob die jetzige Konfiguration übernommen wird. Mit zunehmender Iteration wird der Schwellwert  $H$  bis auf Null verkleinert.

Der Zweck des Schwellwertes ist es, dem Algorithmus die Möglichkeit zu geben, auch Verschlechterungen in der Güte (Zielfunktionswert) zu akzeptieren, um aus lokalen Minimastellen ausbrechen zu können. Die Güte hängt allerdings von der Ausgangslösung, der Nachbarschaftssuche und den Akzeptanzbedingungen ab. Die Akzeptanzbedingungen legen die Schwellwert  $H$ , die Sinkrate und Stopp-Schwellwert fest. Die Nachbarschaft beinhalten hier die Menge aller Lösungen, die von der Ausgangslösung mit genau einem Übergang erreicht werden können. Neuer Belegungsplan wird akzeptiert, wenn die Differenz zwischen alter und neuer Konfiguration den Schwellwert  $H$  nicht überschreitet. Mit abnehmendem Schwellwert wird diese Fähigkeit ausgenutzt, sodass der Algorithmus konvergiert<sup>57</sup>. Der Abbruch des Algorithmus findet statt, wenn keine Ergebnisverbesserung oder Stopp-Schwellwert erreicht. Die Besonderheit dieses Algorithmus ist die relative einfache Struktur und die gute Lösungsqualität. Dazu

---

<sup>56</sup> Die Optimierung der Gewichte stellt sich als ein Minimierungsproblem dar, dass sich im einfachsten Fall lokal um einen Punkt in eindeutiger Weise linear approximieren lässt. Durch deterministische Optimierungsverfahren ist es möglich eine Optimierung der Fehlerfunktion  $E(\mathbf{w})$  aufgrund der Differenzierbarkeit durchzuführen. Hingegen können nichtstetige Optimierungsprobleme durch stochastische Verfahren gelöst werden, welche rechenaufwendiger und ungenauer sind. Dafür kann ein Ansatz beispielsweise auf das Gradientenabstiegsverfahren zurückgreifen, das das Ziel hat, sich auf der Fehleroberfläche talabwärts auf das Minimum zu bewegen. Allerdings findet das Gradientenabstiegsverfahren in der Regel nur *lokale* Minima, d.h. dass das Verfahren unter bestimmten Konstellationen das Endergebnis der Iteration nicht das gesuchte globale Minimum liefern wird, sondern nur das nächstliegende lokale Minimum der Fehleroberfläche.

<sup>57</sup> Obwohl der Datenraum des Gütekriteriums sich durch seine Unstetigkeit auszeichnet, ist der Algorithmus in der Lage damit umzugehen und eine Minimierung vorzunehmen.

### 3. MATCHING-PARADIGMA

---

garantiert dieser Algorithmus, dass eine neue Konfiguration niemals angenommen wird, wenn sich die vorhergehende Lösung stark verschlechtern würde. Allerdings ist der Rechenaufwand hoch. Für die Optimierung stehen *erst* zwei Szenen<sup>58</sup> mit künstlich generierten Objekten und bekannten Pixelverschiebungen zur Verfügung. Für die Minimierung der Funktion  $E(\boldsymbol{w})$  wird hier der TAA mit der initialen Gewichtungskonfigurationen  $\boldsymbol{w}=(1\ 3\ 4\ 1)^T$  initialisiert.

Der Schwellwert  $H$  wird initial auf 0.21 gesetzt. Das bedeutet, dass Konfigurationen mit einem Fehler von 5%-10%, akzeptiert werden. Sobald lange keine Verbesserung stattfindet, aber frühestens nach 50 Iterationen, wird der Schwellwert  $H$  um  $dH = 0.01$  erniedrigt. Dies verhindert das zu schnelle Abkühlen der Suche, um lokale Minima zu vermeiden. Die Abbruchbedingung ist dann erfüllt, wenn der Schwellwert  $H$  Null ist und mindestens 20 Iterationen vergangen sind. Eine Mindestanzahl an Iterationen wurde eingeführt, weil es sich herausgestellt hat, dass die Fehleroberfläche große Plateaus besitzt. Um diese zu überwinden, bedarf es einer gewissen Anzahl an Iterationen.

---

```
Threshold_Accepting ()
{
  initiere (Schwellwert  $H > 0$ , anfängliche Konfiguration  $C =$  Referenzlösung  $C_{ref}$ );
  while (Abbruchbedingung){
    while (bis lange keine Verbesserung){
      generiere ( $C'$  von  $C$ );
       $dQ =$  Qualität( $C'$ ) - Qualität( $C$ );
      if (Qualität( $C'$ ) > Qualität( $C_{ref}$ ))  $C_{ref} = C'$ ;
      if ( $dQ > -H$ )  $C = C'$ ;
    }
    verringere  $H$  { $H=H-dH$ };
  }
};
```

---

**Abbildung 3-6-3:** Threshold-Accepting-Algorithmus (TAA)

Für die Bewertung der ermittelten optimierten Gewichte mittels TAA werden die Teilszenen mit den errechneten Parametern durchlaufen. Aus dem Vergleich mit der Zuordnungsmatrix der Grundwahrheit  $S_c$  werden die *falsch* und *richtig* zugeordneten Paarungen, sowie der Restfehler *fSRE* erfasst. Diese sind in der Tabelle 3-6-2a dargestellt. Die nicht erkannten Paarungen ergeben sich aus der Differenz der Anzahl der Segmente  $p_v^0$  und der Anzahl der Zeilen von  $S_c$ . Als Referenzwerte zu den Ergebnissen der optimierten Gewichte können der gleichförmige Gewichtsvektor  $\boldsymbol{w}=(1\ 1\ 1\ 1)^T$  (*Simple*) und der im Abschnitt 3.5 und in der Arbeit [Al-Hamadi 06b] vorgeschlagene Vektor  $\boldsymbol{w}=(1\ 3\ 4\ 1)^T$  (*Initial*) dienen.

Zur Vereinfachung der Optimierungsschritte ergeben sich in der Regel bei der Betrachtung von der realen Szene *drei* perspektivisch-abhängige Abschnitte.

- In dem *ersten* Abschnitt (TA 1) bewegen sich die Objekte auf den Betrachter/ Kamare zu. Hier entsteht eine perspektivische Änderung in Form einer Objektvergrößerung. Die dominierenden CSC-Segmente werden sich über den Teilabschnitt in abgeleiteten Farb-, Form- und Positionmerkmal relativ stabil verhalten. Es hat zur Folge, dass sich

---

<sup>58</sup> Die erste Szene „*ein Auto*“ zeigt das Modell eines roten PKWs, der sich in einem Kreis auf den Betrachter zu und wieder weg bewegt. Die zweite Szene „*zwei Autos*“ zeigt die Modelle eines roten und eines grünen PKWs, die nebeneinander dieselbe Bewegung durchführen. In dieser Szene findet eine Verdeckung des roten durch den grünen Wagen (gestörte Situationen) statt und im letzten Abschnitt der Szene entfernen sich die Autos und der Halbkreis ist vollendet.

### 3. MATCHING-PARADIGMA

die Korrespondenz der vorhandenen nichtgestörten Segmente über die stabilen Farbwerte und Positionen der Schwerpunkte herstellen lassen kann. Durch die Drehung des Objektes werden sich die kleinen Segmente im Laufe der Drehung mit den größeren Segmenten verschmelzen. Wesentlich dafür ist die Tatsache, dass durch die Drehung eine perspektivische Verzerrung und Farbtonänderungen entstehen.

- Der *zweite* Abschnitt (TA 2) zeigt die Objekte von der Längsseite und wird von großen positionsstabilen Segmenten dominiert. Bei gestörten Bildsituationen verschwinden und entstehen neue CSC-Segmente, z.B. durch Verdeckung. Somit stellt sich heraus, dass die Interframedistanz stabilisierend auf diese Ereignisse wirkt.
- Der *letzte* Abschnitt (TA 3) zeigt die Rückansichten der Objekte. Die Szenen bestehen aus kleinen Segmenten, die nach und nach verschwinden. Verschwindet ein CSC-Segment, so wirkt sich dies auf seine Nachbarn aus, indem es für Störungen der Positionen der Nachbarschwerpunkte sorgt. Dementsprechend wird die Farbinformation eine Bedeutung haben.

Es ist deutlich zu erkennen, dass in allen drei definierten Abschnitten der Teilsequenzen erhebliche Verbesserungen eingetreten sind. Die Rate der falsch und nicht erkannten Paarungen (FzS und NeS) ist zurückgegangen und die der richtig zugeordneten (RzS) gestiegen. Der Wert der Gütefunktion  $E(w)$ , bei dem das Abbruchkriterium erreicht wurde, ist in der Tabelle 3-6-2a angegeben. Je geringer der Wert der Gütefunktion ist, desto besser ist das erreichte Ergebnis. Tabelle 3-6-2a und Abbildung 3-6-4b stellen die errechneten Gewichte dar. Auch nach der Optimierung bleibt ein Restfehler. Allerdings kann die Optimierung durch die auftretenden Verbesserungen als aussagekräftig ausgewertet werden. Die folgende Darstellung der Ergebnisse (Tab. 3-6-2b, 3-6-3a und 3b) beinhaltet:

FzS/ $|Pa_f|$ ) Anzahl der falsch zugeordneten Segmente/ Paarungen

RzS/ $|Pa_r|$ ) Anzahl der richtig zugeordneten Segmente/Paarungen

NeS) Anzahl der nicht erkannten Segmente/Paarungen

**Tabelle 3-6-2a:** Vergleich der optimierten Gewichte mit den Initialwerten über die Grundwahrheit.

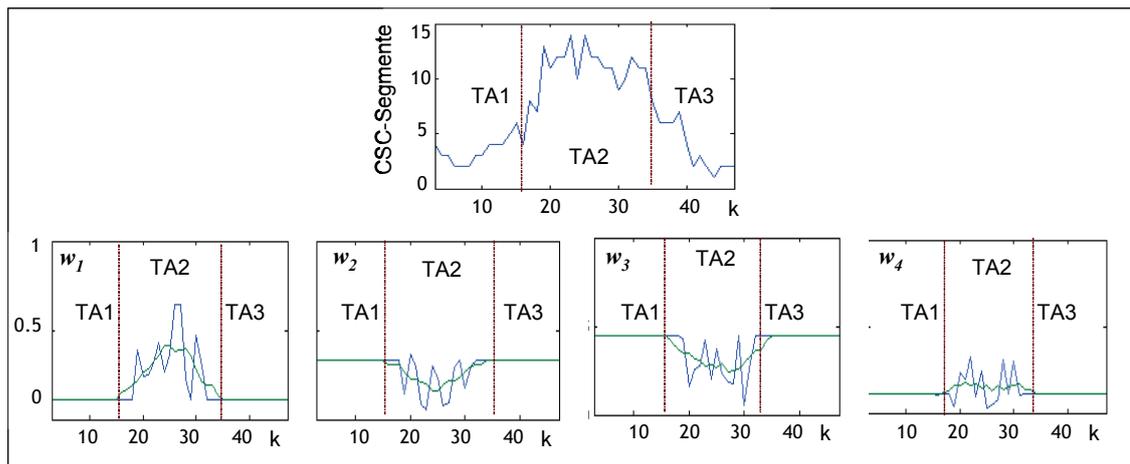
Szene „ein Auto“									
	Abschnitt 1			Abschnitt 2			Abschnitt 3		
	Simple	Initial	Optimiert	Simple	Initial	Optimiert	Simple	Initial	Optimiert
FzS	73	100	52	245	295	173	101	123	65
RzS	241	207	258	253	175	323	206	167	240
NeS	24	31	28	54	82	56	41	58	43
fSRE	0,827	0,974	0,742	0,717	0,878	0,466	0,8567	1,007	0,660
Szene „zwei Autos“									
FzS	135	156	87	425	452	290	219	244	145
RzS	242	209	282	316	242	454	245	210	319
NeS	17	29	25	82	129	79	93	103	93
fSRE	0,867	0,960	1,152	0,748	0,845	0,492	0,828	0,862	0,953

### 3. MATCHING-PARADIGMA

Tabelle 3-6-2b:  $E(\mathbf{w})$ -Fehler der optimierten Gewichte

Szene	„ein Auto“			„zwei Autos“		
	Abschnitt 1	Abschnitt 2	Abschnitt 3	Abschnitt 1	Abschnitt 2	Abschnitt 3
$E(\mathbf{w})$	0,248	0,156	0,159	0,216	0,289	0,296
$w_1$	0,510	0,726	0,663	0,490	0,649	0,520
$w_2$	0,160	0,028	0,021	0,103	0,046	0,098
$w_3$	0,274	0,206	0,275	0,325	0,257	0,331
$w_4$	0,057	0,039	0,042	0,081	0,049	0,051

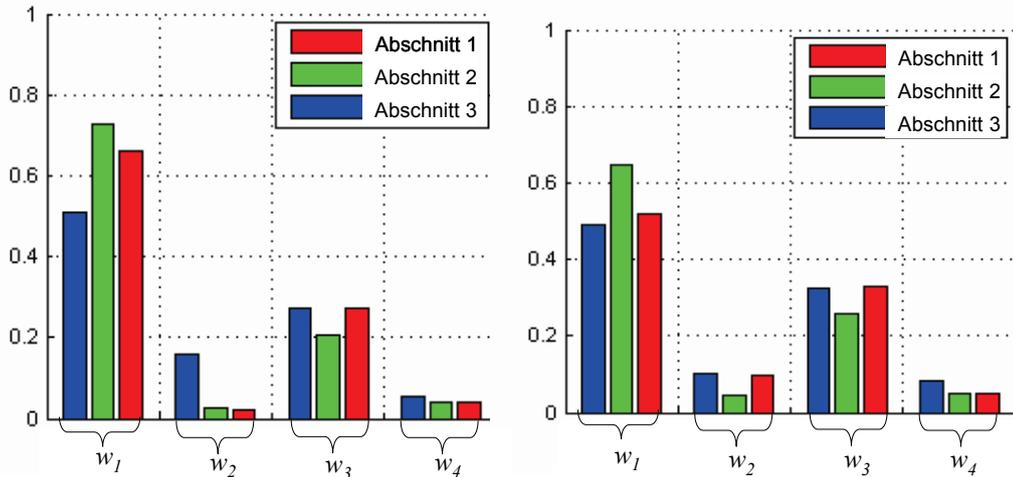
**Fakt:** Die Optimierung der Gewichte wird Bild für Bild durchgeführt. Dazu stellt sich heraus, dass die optimierten Gewichte der Szenen zeitlich sehr starken Schwankungen (Abb. 3-6-4) unterliegen. Wesentlich dafür ist die Tatsache, dass durch den sehr kleinen Optimierungsdatensatz und damit verbunden die geringere Anzahl der Segmente der Matrix  $S_c$  das Gütekriterium  $E(\mathbf{w})$  eine sehr geringe Varianz aufweist. Die Folge ist, dass sich die Minima in Plateaus umwandeln und es in dem Sinne nicht mehr möglich wird, eine Optimierung durchzuführen. Das heißt, dass die Menge der minimalen Lösungen derart steigt, dass eine Optimierung aufgrund fehlender Vergleichsbeispiele unzuverlässig wird. Bei einer zu geringen Anzahl an Segmenten verringert sich somit die Variabilität der CSC-Zuordnungen, wodurch die Fehleroberfläche der Gütefunktion  $E(\mathbf{w})$  aus Gleichung (3-6-2d) homogen wird.



**Abbildung 3-6-4a:** Anzahl der CSC-Segmente der Szene „ein Auto“ für die Optimierung. Unten ist das Verhalten der optimierten Gewichte mittels TAA für die Szene „ein Auto“ (blau) mit einer Glättung durch zeitliche Mittelung der Gewichte (grün).

Die entsprechenden Ähnlichkeitsmaße „relative Position“ und „Farbe“ nehmen die größten optimalen Gewichte im Abschnitt TA 1 (Abb. 3-6-4a) ein. Da die Position eines Schwerpunktes eines Segmentes durch eine Verschmelzung erheblich verschoben werden kann, erweist sich das Merkmal der Interframedistanz und der Farbe als bestimmend für die korrekte Zuordnung der fusionierenden kleineren Segmente.

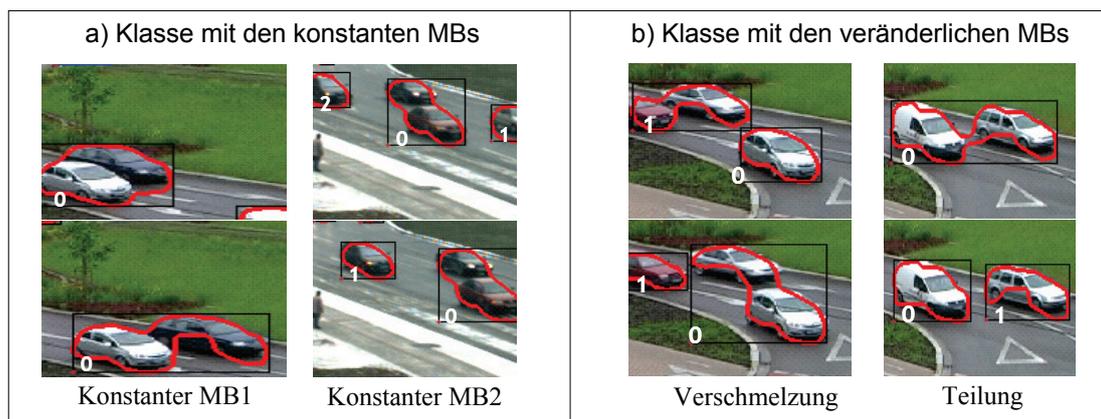
### 3. MATCHING-PARADIGMA



**Abbildung 3-6-4b:** Durch das Zusammenfassen der Szenen in drei Abschnitte (TA1, TA2 und TA3), ergeben sich die optimierten Gewichte für „ein Auto“ links und „zwei Autos“ rechts.

Im Weiteren werden für die Überprüfung und Verallgemeinerung der Tendenzen der errechneten Gewichte aus den synthetischen Szenen nun *reale* Szenen betrachtet (s. Anhang 8.4). Die Pixelbewegungen der CSC-Schwerpunkte werden mit dem PLK-Verfahren approximiert. Durch die spezielle Eigenschaft des Ähnlichkeitsmaßes der relativen Position, dass es von der Größe des MBs abhängig ist, wird das CSC-Matching, bei einer hohen Gewichtung der relativen Position, während der Verschmelzung oder der Teilung, eines MBs gestört. Es ist deshalb notwendig die Optimierung der Gewichte für Szenen mit Verschmelzung und Teilung und Szenen mit konstantem MB getrennt zu betrachten.

Die kritische Stelle bei der Verschmelzung, an der ein genaues Matching notwendig ist, befindet sich zwischen den beiden MBs (s. Abb. (3-6-5a)). Es können falsche Paarungen entstehen, die den PKW des MB 0 mit dem des MB 1 verbinden. Diese falsch zugeordneten Segmente (FzS) bilden Zentren, durch die sich im späteren Verlauf die Fahrzeuge des MB 0 nicht mehr von denen des MB 1 trennen lassen.



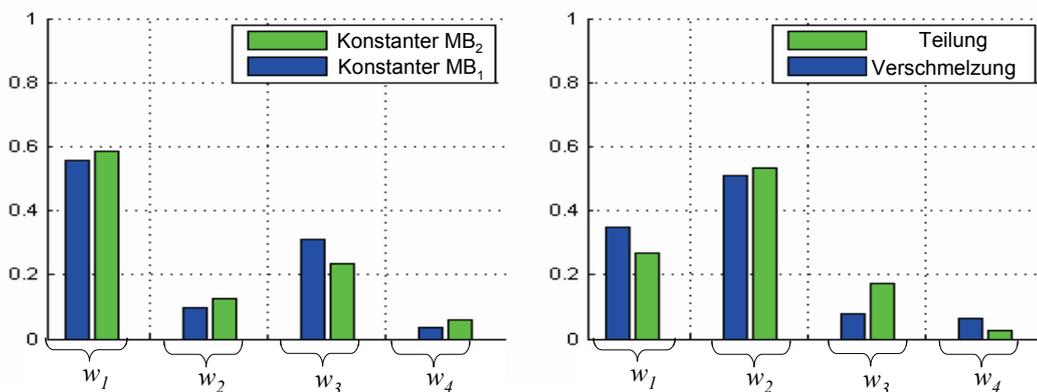
**Abbildung 3-6-5a:** Trennung der Optimierung der realen Testszene in zwei Klassen

Der gleichgewichtete Fall weist falsche Paarungen auf, die den PKW des MB 0 mit dem des MB 1 verbindet (Abbildung 3-6-5a). Diese falsch zugeordneten Segmente werden durch eine verbesserte Gewichtung, den initialen und optimierten Gewichten beseitigt. Auch bei Teilung treten die gleichen Probleme auf und somit ist die kritische Stelle bei

### 3. MATCHING-PARADIGMA

der Teilung eines MB, dass das Matching der CSC-Segmente nicht über die Fahrzeuggrenzen hinaus fehlerhaft verläuft.

Durch die Optimierung der Gewichte beschränken sich die Bereiche der falschen Paarungen ausschließlich auf die der PKWs. Nach der Verschmelzung zweier MBs ist es wichtig, die Positionsinformation der enthaltenen Objekte bis zur Teilung korrekt beizubehalten. Aus diesem Grund wurden Verkehrsszenen analysiert, in denen sich die MBs konstant verhalten. Es zeigt sich, dass durch die Optimierung nicht mehr korrekte CSC-Paarungen, als in den beiden anderen Fällen, erkannt werden. Das *SRE*-Fehlermaß wird allerdings deutlich gesenkt. Daraus lässt sich schlussfolgern, dass Paarungen mit großem *SRE*-Fehlermaß, die also weit entfernt von ihrer wirklichen Lage geschätzt wurden, eliminiert werden konnten. Das Verfahren wird so robuster, da die falsche Zuordnung in der näheren Umgebung der idealen Zuordnung liegt, und der Effekt, dass ein CSC-Paar von einem zum anderen Objekt überspringt, weniger wahrscheinlich wird.



**Abbildung 3-6-5b:** Durch eine getrennte Betrachtung der Szene mit veränderlichen und konstanten MBs, resultieren unterschiedliche Optimierungen für den Gewichtsvektor  $w$ .

Anhand der Tabellen (3-6-3b) zeigt sich auch, dass in der Szene „konstanter MB1“ die initialen Gewichte eine Verschlechterung zu den gleichverteilten Gewichten darstellen. Der Vergleich der Szenen mit veränderlichen und konstanten MBs zeigt, dass bei konstanten MBs die relative Position das dominante Merkmal ist, während sich bei einer Verschmelzung oder Teilung die Interframedistanz als ausschlaggebend erweist. Es zeigt sich auch, dass die Trennung der Optimierung anhand veränderlicher und konstanter MBs notwendig ist. (Weitere Abbildungen sind im Anhang 8.4 zu finden).

**Tabelle 3-6-3a:** Vergleich der optimierten Gewichte mit den initialen und gleichverteilten Werten bei Teilung und Verschmelzung von MBs.

Szene	„Verschmelzung“			„Teilung“		
	Simple	Initial	Optimiert	Simple	Initial	Optimiert
$w_1$	1	1	0,348	1	1	0,268
$w_2$	1	3	0,510	1	3	0,535
$w_3$	1	4	0,079	1	4	0,171
$w_4$	1	1	0,069	1	1	0,025
FzS	51	54	62	33	36	28
RzS	145	139	126	88	84	87
NeS	6	6	14	5	6	11
fSRE	3,569	2,589	2,148	2,961	3,672	2,025

### 3. MATCHING-PARADIGMA

**Tabelle 3-6-3b:** Vergleich der optimierten Gewichte mit den initialen und gleichverteilten Werten bei konstanten MBs.

Szene	„Konstanter MB1“			„Konstanter MB2“		
	Simple	Initial	Optimiert	Simple	Initial	Optimiert
$w_1$	1	1	0,559	1	1	0,584
$w_2$	1	3	0,098	1	3	0,124
$w_3$	1	4	0,310	1	4	0,235
$w_4$	1	1	0,033	1	1	0,058
FzS	143	142	100	63	95	52
RzS	114	111	159	141	162	150
NeS	49	53	47	3	11	5
fSRE	0,824	0,862	0,725	2,49	1,539	1,269

#### 3.6.1.2 Getrennte qualitative Bewertung der geschätzten Gewichtsvektoren

Um ein Maß für eine qualitative Beschreibung der erzielten Verbesserungen zu formulieren, wird ein Gütekriterium  $G_Z(\mathbf{w})$  definiert, das von der Zuverlässigkeit der Zuordnungspaare abhängig ist. Die Zuordnungsgüte  $G_Z(\mathbf{w})$  ist so normiert, dass sie einen Wertebereich von  $[0,1]$  annimmt, wobei bei einem fehlerfreien Matching der Gütewert Null annimmt. Darüber hinaus wird ein zweites Gütemaß  $G_{SRE}(\mathbf{w})$  definiert, welches von dem *SRE*-Fehlermaß abhängig ist, um die Fehler der nicht zugeordneten Segmentpaare mit zu betrachten. Dieses Gütemaß gibt das durchschnittliche *SRE*-Fehlermaß pro falsches Zuordnungspaar an. Die Mengen  $Pa_r$  der richtigen und  $Pa_f$  der falschen Zuordnungspaare ergeben sich daher aus der Validierungsmatrix<sup>59</sup> ( $S_v$ ) und Grundwahrheitsmatrix ( $S_c$ ). Aus dem Vergleich mit der Zuordnungsmatrix der Grundwahrheit  $S_c$  werden die *falsch* und *richtig* zugeordneten Paarungen, sowie der Restverschiebungsfehler *fSRE* erfasst und in den Tabellen 3-6-2a, 3-6-2b, 3-6-3a und -3b dargestellt. Als Referenzwerte zu den optimierten Gewichten dienen der gleichförmige Gewichtsvektor  $\mathbf{w}=(1\ 1\ 1\ 1)^T$  (Simple) und der Vektor  $\mathbf{w}=(1\ 3\ 4\ 1)^T$  (Initial).

$$G_Z(\mathbf{w}) = \frac{1}{2} \cdot \left( 1 - \frac{|Pa_r| - |Pa_f|}{|Pa_r| + |Pa_f|} \right) \quad (3-6-3a)$$

$$G_{SRE}(\mathbf{w}) = \frac{\sum_v \sum_\mu SRE(p_v^k, p_\mu^{k+1})}{|Pa_f|} \quad (3-6-3b)$$

Der Wertebereich dieser zweiten Güte liegt im Intervall  $[0, \max_{v,\mu}(SRE(p_v^k, p_\mu^{k+1}))]$  d.h. bei

einem idealen Matching ist der Gütewert Null und im ungünstigsten Fall kann er den *SRE*-Fehlerwert des schlechtesten CSC-Matchingpaares annehmen.

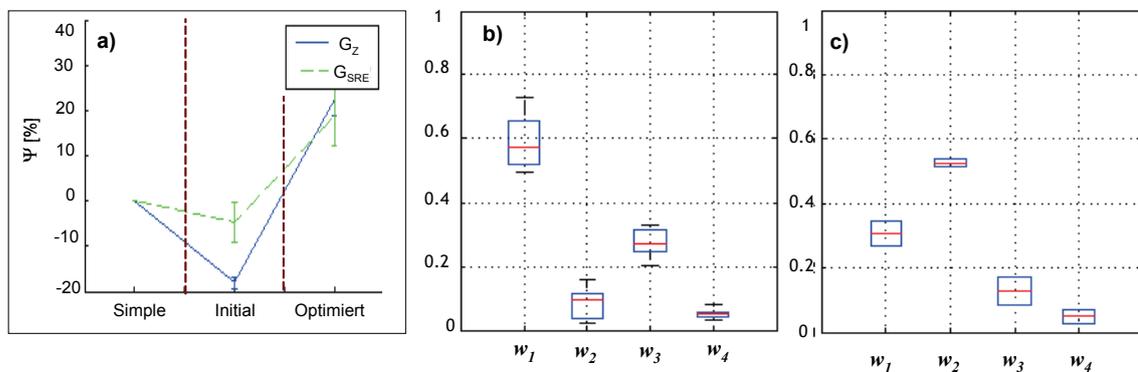
Abbildung (3-6-5c) zeigt den Vergleich der Güte des Matchingverfahrens mit den gleichverteilten, initialen und optimierten Gewichten. Sie sind auf den gleichgewichteten Fall normiert. Die Grafik bestätigt die im Einzelnen betrachteten Untersuchungen und zeigt eine signifikante Verbesserung des Matchingverfahrens mit optimierten Gewichten. Die Gesamtgüte wird aus dem statistischen Mittelwert der Gütemaße der

<sup>59</sup> Die Validierungsmatrix  $S_v$  wird in Abhängigkeit vom optimierten Gewichtsvektor  $\mathbf{w}$ , durch das SM- Gesamtähnlichkeitsmaß (Gl. 3-5-8) errechnet.

### 3. MATCHING-PARADIGMA

einzelnen Szenen, dargestellt in Tabelle 3-6-4, berechnet. Die Varianz der Gütemaße ist als Balken in das Diagramm eingetragen. Die Gütemaße werden auf den Fall „Simple“ normiert, um die relative Verbesserung abzubilden. Es zeigt sich auch, dass der fest vorgegebene initiale Gewichtsvektor in speziellen Fällen eine Verbesserung bewirkt. Allerdings verschlechtert sich die eindeutige Zuordnung der CSC-Segmente.

Abbildung 3-6-5c zeigt die Gesamtdarstellung aller optimierten Gewichte als Boxplot<sup>60</sup>. In den Versuchen zeigte sich, dass eine generelle Gewichtsconfiguration für ein optimales CSC-Matching vom Bildinhalt abhängig ist. Dafür zeigten die Versuche, dass sich die optimierten Gewichtsvektoren in zwei Klassen einteilen lassen. Der Klasse der konstanten MB-Konstellationen und der Klasse der veränderlichen MB-Konstellation. In der ersten Klasse ist die relative Position das dominierende Merkmal und in der zweite die Interframedistanz. Da die realen Bilder nicht gesättigte Farben enthalten, wird die Farbdistanz im Gesamtähnlichkeitsmaß durch die vorgestellte TAA-Optimierung nicht dominante Wichtungsfaktor aufweisen.



**Abbildung 3-6-5c:** a) Relative Verbesserung der Gütekriterien in Bezug auf den Fall „Simple“<sup>61</sup>. b) und c) Boxplot-Darstellung der optimierten Gewichte für konstante MBs (b) und veränderliche MBs.

**Tabelle 3-6-4:** Gütemaße für die betrachteten Szenen (s. Anhang 8.4).

	$G_Z(w)$			$G_{SRE}(w)$		
	Simple	Initial	Optimiert	Simple	Initial	Optimiert
Sz.1 „ein Auto“	0,232	0,326	0,168	0,827	0,974	0,742
Sz.2 „ein Auto“	0,492	0,628	0,349	0,717	0,878	0,466
Sz.3 „ein Auto“	0,329	0,424	0,213	0,856	1,003	0,660
Sz.1 „zwei Autos“	0,358	0,427	0,236	0,867	0,960	1,152
Sz.2 „zwei Autos“	0,574	0,651	0,390	0,748	0,845	0,492
Sz.3 „zwei Autos“	0,472	0,537	0,313	0,828	0,862	0,953
Sz.4 „Verschmelzung“	0,260	0,280	0,330	3,569	2,589	2,148
Sz.5 „Teilung“	0,273	0,300	0,243	2,961	3,672	2,025
Sz.6 „Konstanter MB1“	0,556	0,561	0,386	0,824	0,862	0,725
Sz.7 „Konstanter MB2“	0,309	0,370	0,257	2,490	1,539	1,269

<sup>60</sup> Der Box-Whisker-Plot dient zur Darstellung der Gewichte. Die rote Linie bezeichnet den Median, der Kasten die Varianz und der Balken das Minimum bzw. Maximum der Gewichte.

<sup>61</sup> Die relative Güte errechnet sich aus den Gütekriterien, um Bezug auf die Güte des Falls „Simple“. Aus Gründen der Anschaulichkeit wird diese Güte von eins abgezogen, um eine Verbesserung als wachsenden Graphen darzustellen. Es gilt daher:  $\psi = [1 - \text{relative Güte}]$ .

#### 3.6.2 Erweiterung des Ähnlichkeitskriteriums

Die Korrespondenzermittlung von der CSC-Matching erfolgte bisher durch das SM-Gesamtähnlichkeitsmaß, das sich aus der gewichteten Summe von merkmalspezifischen Einzelähnlichkeiten ergibt. Nun stellt sich die Frage, ob die Einbeziehung von weiteren Merkmalen in das Gesamtähnlichkeitsmaß zur weiteren Verbesserung der Qualität bei gestörten Bildsituationen führen kann. Aus diesem Grund werden zwei weitere merkmalspezifische Ähnlichkeitsmaße in Betracht gezogen, die den Umfang ( $SL_{UMF}$ ) und die Orientierung ( $SL_{ORI/EIG}$ ) eines Segmentes beinhalten.

- 1). Das erste neue merkmalspezifische Ähnlichkeitsmaß  $SL_{UMF}$  wird aus den Größenverhältnissen der Segmentumfänge berechnet. Dieses verhält sich ähnlich der Größe eines CSC-Segmentes. Es enthält die Längeninformaton der Kontur eines Segmentes und reagiert, wie die Größe, empfindlich auf das Ergebnis des Farbsegmentierungsalgorithmus. Es kennzeichnet sich durch Rotations- und Translationsinvarianz aus und es kann im Zusammenhang mit der Größe einen Merkmalsraum ähnlich der Kompaktheit aufspannen.
- 2). Das zweite neue merkmalspezifische Ähnlichkeitsmaß bedient sich der statistischen Analyse der Segmentierungsmaske. Es werden die Eigenwerte der Kovarianzmatrix aus der Hauptkomponentenanalyse für die CSC-Pixelmasken miteinander verglichen. Das Ähnlichkeitsmaß der Eigenwerte besitzt die Eigenschaft skalierung-, translations- und rotationsinvariant zu sein und besitzt einen Wertebereich von  $[0,1]$ .

Der Vorteil des Eigenwertemerkmals (ähnliches gilt für das Orientierungsmerkmal) ist es, dass es im Gegensatz zu allen anderen Merkmalen eine Information über die Segmentsausdehnung in Form der Pixelverteilung entlang der Hauptachsen liefert. Dabei wirkt sich eine Rotation, Translation oder die Skalierung des CSC-Segmentes, wie sie bei näher kommenden Fahrzeugen auftreten kann, nicht auf das Merkmal aus. Mit diesen beiden zusätzlichen Ähnlichkeitsmaßen ist es möglich, das dargestellte Gesamtgütekriterium (Gleichung 3-5-8 im Abschnitt 3.5) zu erweitern, wobei der Ähnlichkeitsraum nun auf den vier Merkmalsklassen basiert:

- Positionsmerkmale ( $SL_{POS}$ ,  $SI_{DIST}$ )
- Farbmerkmal ( $SC_{RGB}$ )
- Formmerkmale ( $SA_{SIZE}$ ,  $SL_{UMF}$ )
- Statistisches Merkmal ( $SL_{ORI/EIG}$ )

Nun soll die Erweiterung analysiert werden, wie sich die Einzelmaße auf das Multi-Matching auswirken. Dazu wird die Anzahl der Kriterien (AK) von zwei (AK1) bis sechs (AK5) erhöht. Die berücksichtigten Ähnlichkeitsmaße sind im folgenden zusammengefasst:

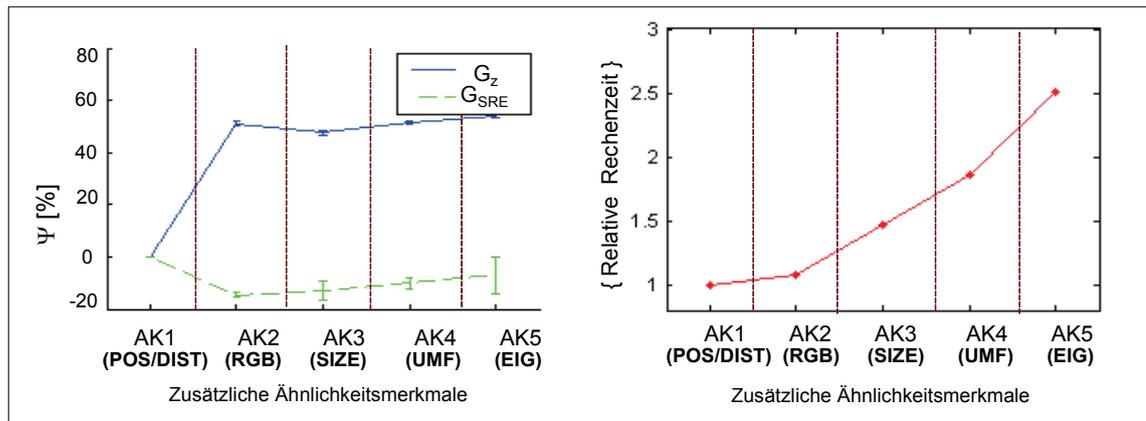
- AK1)  $SL_{POS}$  und  $SI_{DIST}$
- AK2)  $SL_{POS}$ ,  $SI_{DIST}$  und  $SC_{RGB}$
- AK3)  $SL_{POS}$ ,  $SI_{DIST}$ ,  $SC_{RGB}$  und  $SA_{SIZE}$
- AK4)  $SL_{POS}$ ,  $SI_{DIST}$ ,  $SC_{RGB}$ ,  $SA_{SIZE}$  und  $SL_{UMF}$
- AK5)  $SL_{POS}$ ,  $SI_{DIST}$ ,  $SC_{RGB}$ ,  $SA_{SIZE}$ ,  $SL_{UMF}$  und  $SL_{ORI/EIG}$

Für die Optimierung des Gesamtprozesses ist es zusätzlich notwendig den Rechenaufwand mit in Betracht zu ziehen. Dieser lässt sich in zwei Abschnitte unterteilen, der Berechnung der Parameter, wie der relativen Position und der

### 3. MATCHING-PARADIGMA

Berechnung der Korrelation der Merkmale. Ein gesteigerter Rechenaufwand für die Merkmalsextrahierung ergibt sich allein bei der Berechnung der Eigenwerte, durch die Berechnung der Kovarianzen. Die übrigen Merkmale ergeben sich während des Segmentierungsschrittes der CSC-Segmente und verlangen somit keine zusätzliche Rechenzeit. Zu dem ist der Rechenaufwand für die Gewinnung der Merkmalsparameter mit durchschnittlich 15% im Vergleich zu dem Rechenaufwand der CSC-Segmentierung sehr gering. Hingegen besitzt die Berechnung der Korrespondenztabelle mit durchschnittlich 60% im Vergleich zu der Rechenzeit des gesamten Matching, den größeren Anteil.

Abbildung (3-6-6a) zeigt die relative Rechenzeit in Bezug auf die Rechenzeit mit zwei Merkmalen. Der Rechenaufwand steigt durch arithmetische Operationen, wie die Berechnung der Quadratwurzel und die des absoluten Wertes. Durch die Farbinformation verbessert sich die Güte der richtigen Zuordnung um 50%, der Restfehler wird aber um ca. 10% größer. Es zeigt sich, dass die Merkmale einen unterschiedlichen Einfluss auf die korrekte Zuordnung und den  $fSRE$  des Matchings haben. Die Gütemaße wurden anhand der synthetischen Szenen bestimmt. Es wird gezeigt, dass im idealen Fall, mit der Zunahme der Anzahl der Merkmale, die Gesamtgüte ansteigt. Durch den erhöhten Rechenaufwand, den jedes Merkmal mit sich bringt, scheint es ratsam, nicht mehr als die relative Position (POS), die Interframedistanz (DIST) und die Farbinformation (RGB) zu verwenden. Der Rechenaufwand steigt mehr als der Faktor 0,5 bei der Einbeziehung von weiteren Merkmalen (Abb. 3-6-6a).



**Abbildung 3-6-6a:** Relative Verbesserung der Gütekriterien in Bezug auf die Güte mit zwei Merkmalen und die relative Rechenzeit in Bezug auf die Rechenzeit mit zwei Merkmalen, für synthetisch erzeugte Szenen.

**Tabelle 3-6-5a:** Gütemaße für die Erweiterung der Ähnlichkeitsmerkmale am Beispiel idealer Szenen

Ähnlichkeitsmaße	$G_z(w)$					$G_{SRE}(w)$				
	AK1	AK2	AK3	AK4	AK5	AK1	AK2	AK3	AK4	AK5
Sz. 1 „ein Auto“	0,395	0,151	0,161	0,155	0,154	0,712	0,745	0,768	0,700	0,703
Sz.2 „ein Auto“	0,543	0,318	0,338	0,293	0,295	0,419	0,514	0,536	0,504	0,512
Sz.3 „ein Auto“	0,473	0,189	0,198	0,190	0,197	0,569	0,664	0,641	0,650	0,792
Sz.1 „zwei Autos“	0,503	0,215	0,249	0,227	0,186	1,225	1,356	0,982	1,066	0,818
Sz.2 „zwei Autos“	0,636	0,366	0,367	0,376	0,347	0,411	0,529	0,553	0,528	0,507
Sz.3 „zwei Autos“	0,516	0,298	0,322	0,276	0,259	0,839	0,876	0,966	0,944	0,772

### 3. MATCHING-PARADIGMA

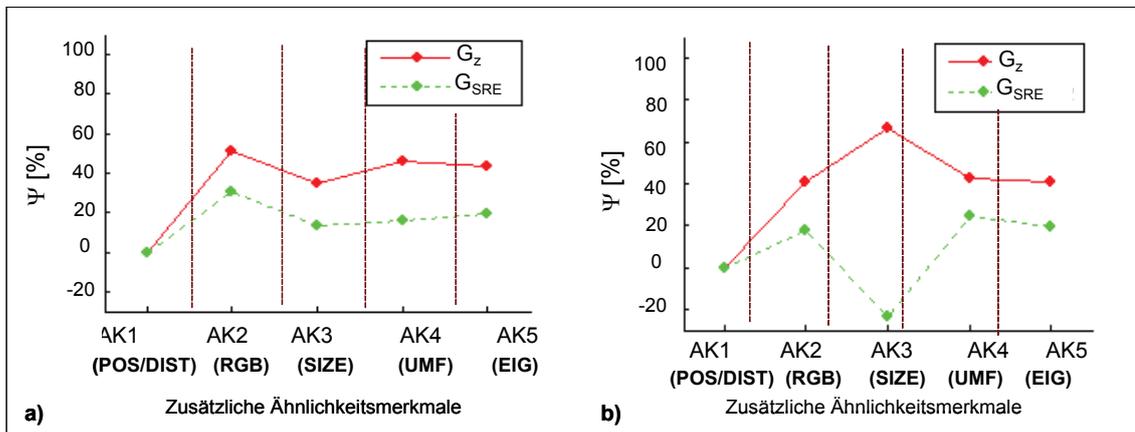
Die Erweiterung der Gütemerkmale führt im realen Fall nicht zwingend zu einer Verbesserung der Güte. Mit der Verwendung zusätzlicher Merkmale kann es vermehrt zu Doppeldeutigkeiten kommen, was sich negativ auf das Matching auswirkt, wie in Abbildung (3-6-6b) zu sehen ist. Die Güte der Restfehler verschlechtert sich mit dem Hinzufügen des Größenmerkmals, aber in Kombination mit dem Umfangmerkmal ist eine signifikante Verbesserung erkennbar.

**Tabelle 3-6-5b:** Gütemaße für die Erweiterung der Ähnlichkeitsmerkmale am Beispiel realer Szenen.

Ähnlichkeitsmaße	$G_Z(w)$					$G_{SRE}(w)$				
	AK1	AK2	AK3	AK4	AK5	AK1	AK2	AK3	AK4	AK5
Sz.4 „Verschmelzung“	0,474	0,23	0,309	0,258	0,27	2,143	1,691	2,05	2,09	1,956
Sz.6 „Konstanter MB1“	0,614	0,36	0,205	0,351	0,36	0,656	0,827	1,92	0,72	0,798

Im Gegenteil zu der Güte des Restfehlers vergrößert sich die Zuordnungsgüte. Dies deutet darauf, dass sich zwar die Gesamtanzahl der korrekt zugeordneten Segmentpaare vergrößert aber diejenigen, die falsch zugeordnet wurden, besitzen einen größeren Fehler. Dies kann dazu führen, dass die Segmenttrajektorien mit höherer Wahrscheinlichkeit in das fremde Objekt aufgrund der Segmentlage springen, was die Objektverfolgung stört.

Wird das Umfangsmerkmal hinzugenommen, so werden weniger korrekte Segmentpaare gefunden. Die falschen Paare liegen aber näher bei den korrekten, da der Restfehler geringer ist, und somit sinkt die Wahrscheinlichkeit, dass die Segmenttrajektorien auf das fremde Objekt überspringen. Die Objektverfolgung wird trotz falsch zugeordneter Paare robuster. Das zeigt die stabilisierende Wirkung der Kombination der Merkmale Umfang und Größe.



**Abbildung 3-6-6b:** Relative Verbesserung der Gütekriterien in Bezug auf die Güte mit zwei Merkmalen, für die Szenen „Verschmelzung“ (a) und „konstanter MB1“ (b).

Bei der Betrachtung der Gütewerte in Abbildung 3-6-6b (links) erkennt man, dass die Einbeziehung der zusätzlichen Größenmerkmale aufgrund der Ungenauigkeit im Optimierungsschritt zur Verschlechterung der beiden Gütemerkmale geführt hat. Dieser Fall zeigt die Schwierigkeiten bei der Optimierung<sup>62</sup> der nichtdifferentiellen Fehler-

<sup>62</sup> Das Optimierungsverfahren ist in diesem Fall in einem lokalen Minimum stecken geblieben. Der minimierte Gewichtsvektor beträgt  $(0,242 \ 0,551 \ 0,135 \ 0,071)^T$  mit den Gütewerten  $G_Z=0,309$  und  $G_{SRE}=2,050$ . Für den

### 3. MATCHING-PARADIGMA

funktion  $E(\mathbf{w})$ . Während die Positions- und Farbmerkmale für ein genaues CSC-Matching ausschlaggebend sind, wirken sich die Formmerkmale stabilisierend auf den Restfehler des Matching aus.

In Hinblick auf den Rechenaufwand stellt sich heraus, dass sich die Kombination aus Fläche und Umfang besonders gut eignet, da diese sich untereinander zusätzlich stabilisieren. In Abbildung (3-6-6c) sind die optimierten Gewichte für veränderliche und konstante MBs angegeben. Die Darstellung beschränkt sich auf die letztendliche Optimierung der Gewichte, auf die Positionsmerkmale POS ( $w_1$ ) und DIST ( $w_2$ ), das Farbmerkmal RGB ( $w_3$ ) und auf die Formmerkmale SIZE ( $w_4$ ) und UMF ( $w_5$ ), da die Verwendung des Eigenwertmerkmals ineffizient ist.

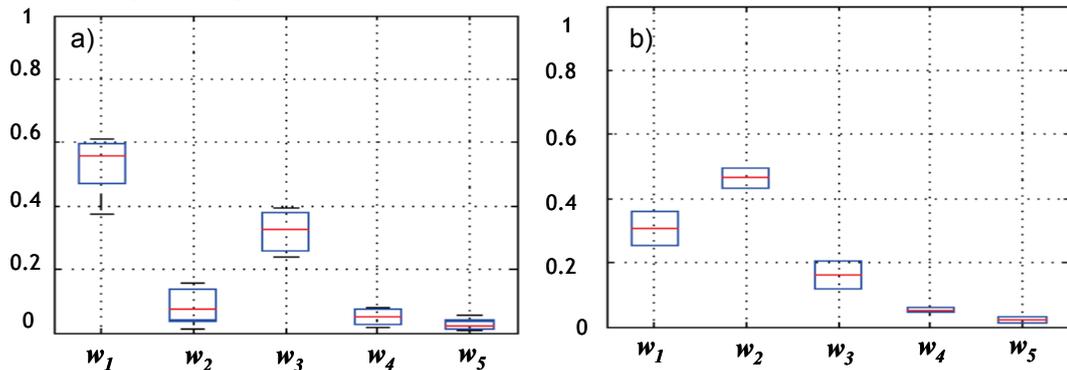


Abbildung 3-6-6c: Boxplot-Darstellung der optimierten Gewichte für konstante MBs links und veränderliche MBs rechts.

Des Weiteren hat sich die Einteilung in zwei Klassen auch hier als sinnvoll erwiesen. Es hat sich auch herausgestellt, dass sich im Hinblick auf den Rechenaufwand die Verwendung des Eigenmerkmals nicht rentabel ist. Durch einen etwas erhöhten Rechenaufwand in der Korrelationsberechnung können so starke fehlerbehaftete falsche Zuordnungen verringert werden.

**Schlussfolgerung:** Die grundlegende Idee bei der Generierung des SM-Gesamtkriteriums (Gl. 3-5-8) besteht darin, dass die Zuordnung der CSC-Segmente durch die definierten merkmalspezifischen Ähnlichkeitsmaße hergestellt wird. Anhand der Wichtungsfaktoren  $w_i$  können diese Ähnlichkeitsmaße über ihrer Zuverlässigkeit verknüpft werden. Dies führt zur Auslegung, dass nicht nur das SM-Gesamtkriterium in Bezug auf die Lokalisierung des Extremwertes verbessert wurde, sondern auch die Reduktion des Einflusses von nicht zuverlässigen Komponenten gewährleistet wurde. Allerdings hängt die Reduktion des Einflusses von unzuverlässigen Komponenten von mehreren Faktoren ab, die zeit-, lage- und störabhängig sind. Deshalb wurde das vorgeschlagene Gesamtkriterium einer Optimierung in Hinsicht auf Genauigkeit und Stabilität des gesamten Matching-Paradigmas benötigt, da sich Fehler in der Berechnung fortsetzen und verstärken je länger die Verfolgung andauert. Die Optimierung hat sich somit auf die Wichtungsfaktoren  $\mathbf{w} = (w_1 w_2 w_3 w_4)$  und die Anzahl der einbezogenen Ähnlichkeitsmaße im SM-Gesamtkriterium bezogen. Dafür wurden in diesem

---

Gewichtsvektor  $(0,349 \ 0,442 \ 0,208 \ 0)^T$  werden die Gütewerte  $G_Z=0,232$  und  $G_{SRE}=1,691$  erreicht. Da beide Werte kleiner sind, beweist dies, dass ein kleinerer Wert für die Fehlerfunktion  $E(\mathbf{w})$  existiert, als der als optimal angegebene. Es zeigt sich also, dass nicht davon ausgegangen werden kann, dass der optimierte Gewichtsvektor im globalen Minimum liegt. Dennoch beeinflusst dies nicht die Verbesserungen, die durch die Optimierungen erreicht wurden. Es beweist nur, dass möglicherweise optimale Gewichtsvektoren existieren.

Kapitel ein Optimierungsverfahren vorgestellt, das keine vereinfachenden Einschränkungen bezüglich des Verkehrsszenarios und der Art der Verkehrsteilnehmer voraussetzt.

Aus diesem Grund wurde ein *TAA* angewendet, der eine Abwandlung des *Simulted Annealing Algorithmus* darstellt. Dabei stellte sich heraus, dass zwei Klassen von Optimierungen abhängig von den Abschnitten und gestörten Situationen zu finden sind, für die es jeweils angepasste Gewichte gibt. Diese Klassen sind die Klasse der konstanten MB-Konstellationen und die Klasse der veränderlichen MB-Konstellation.

In der ersten Klasse ist die relative Position das dominierende Merkmal und in der zweiten Klasse, aufgrund der Abhängigkeit der relativen Position von den MB-Maßen, die Interframedistanz. Es zeigte sich auch, dass die Gewichte innerhalb einer Klasse variieren können. Begründet ist dieser in der Abhängigkeit der Gütemaße vom Bildinhalt, wie z.B. der Farbigkeit der Szene und der Konstanz der CSC-Segmentierung. Dennoch lässt sich eine grundlegende Tendenz der dominierenden Gewichte bestimmen. Die Optimierung der Gewichte hat sich als sinnvoller Schritt herausgestellt, wodurch Verbesserungen im Bereich des eindeutigen als auch uneindeutigen Matchings erreicht wurden.

Eine Weitere Zielsetzung zur Verbesserung des CSC-Matchings beinhaltet die Einbeziehung von zusätzlichen Merkmalen in dem SM-Gesamtkriterium. Durch die schrittweise Erweiterung des Gesamtähnlichkeitsmaßes wurde gezeigt, wie sich die Einzelmaße auf das Matching auswirken. Dazu wird die schrittweise Anzahl der Ähnlichkeitsmaße von zwei auf sechs erhöht. Es stellt sich heraus, dass das Umfangsmerkmal in Kombination mit der Fläche eine Verbesserung bewirkt, da sich diese untereinander zusätzlich stabilisieren.

Als Ausblick auf weitere Optimierungsschritte des CSC-Matching würde sich ein künstlich neuronales Netz eignen, um die Gewichte analog zu [Al-Hamadi 01] zu berechnen. Es können aus den Ähnlichkeitsmaßen und weiteren Kriterien, wie z.B. der Farbigkeit der Szene oder der Kompaktheit der CSC-Segmente, über ein Training mit einer Grundwahrheit, die optimalen Gewichte geschätzt werden.

Ein weiterer Ansatz wäre die Ausführung des CSC-Matchings über eine Fuzzy-Logik. Demnach gäbe es keine eindeutige Zuordnung eines Segmentes zu einen anderen mehr. Die Zugehörigkeit würde durch eine Fuzzy-Wahrscheinlichkeit gegeben werden. Die Übertragungsfunktion müsste dann mit Hilfe von Expertenwissen über Fuzzy-Mengen modelliert werden. Mit diesem Verfahren könnte eine Möglichkeit zur Verbesserung der Effektivität und der Robustheit der Objektverfolgung bestehen.

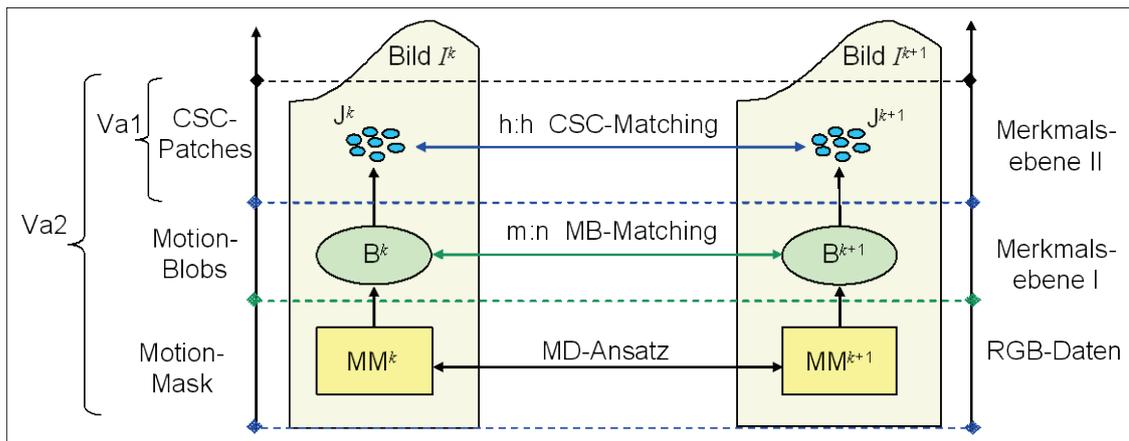
### 3.7 Validierung des Gesamtmatching-Prozesses

Zur Überprüfung der grundlegenden Funktionsfähigkeit des Gesamtmatching-Prozesses eignen sich künstlich erzeugte Eingangsdaten, bei denen die Bewegungsparameter als Grundwahrheit bekannt sind. Die Korrelationsgewichte können entweder unter objektiver Überprüfung ihrer Wirkung oder mittels der vorgestellten im Abschnitt 3.6 Ansätze bestimmt werden. Die Verschiebungsvektoren leiten sich aus Korrespondenzen zwischen CSC-Segmenten ab, die mit Hilfe des beschriebenen zweistufigen Matchings ermittelt werden. Um Aussagen über die Genauigkeit der Ergebnisse des Gesamtmatchings treffen zu können, wurden zwei grundlegende Experimente durchgeführt (Abb. 3-7-1):

**Va1:** Funktionsfähigkeit des 1:1 Matchings mit konstanten Segmenten

**Va2:** Funktionsfähigkeit des gesamten Matchingsprozesses mit variablen Segmenten

In Teil (Va1) wurde die grundsätzliche Funktionsfähigkeit des 1:1 Matchings der CSC-Segmente geprüft. In Teil (Va2) wird der gesamte Prozess zur Ermittlung der Bewegungsparameter ausgehend von der differenzbildbasierten Bewegungserfassung, über die zweistufige Merkmalerfassung und das zugehörige Matching untersucht.



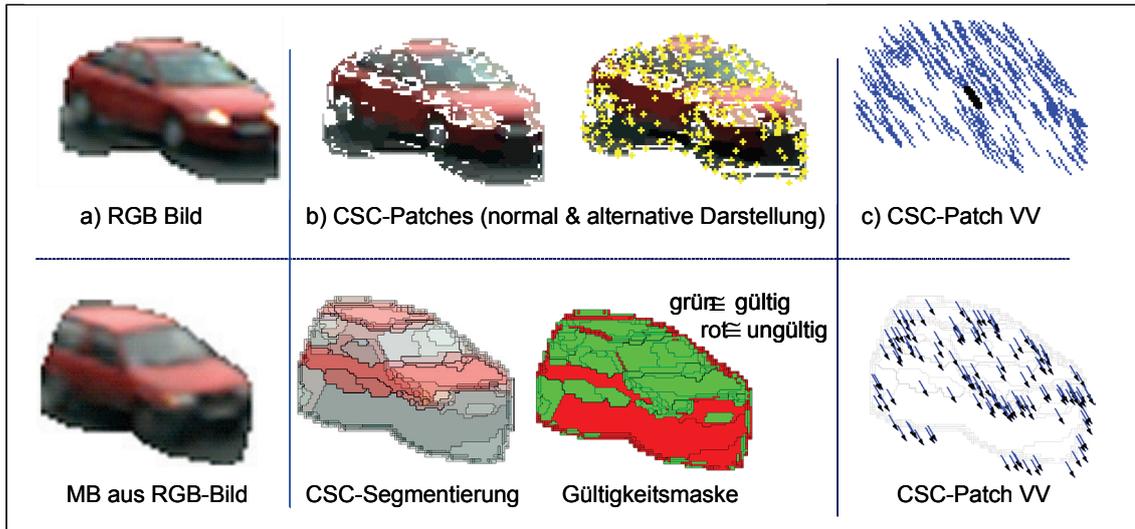
**Abbildung 3-7-1:** Nachweis der Funktionsfähigkeit des Matching-Prozesses in zwei Schritten

**Va1:** Überprüfung der Funktionsfähigkeit des 1:1 Matchings bei konstanten Segmenten

Das auf Ähnlichkeitsmaßen basierende 1:1 Merkmals-Matching (h:h CSC-Matching) repräsentiert die Kernfunktion des gesamten Paradigmas. Um eine separate, von Nebeneffekten unbeeinträchtigte Prüfung des 1:1 Matchings durchzuführen, müssen Variationen, die bei der bewegungsbasierten Segmentierung und Erzeugung der MB-Strukturen auftreten können, vermieden werden. Um dies zu realisieren, erfolgte eine Generierung von MB-Regionen aus konstanten RGB-Bildern. Im Anschluss wurden die MB-Regionen auf einem kreisförmigen Bewegungspfad bewegt. Aus dieser Ausgangssituation erfolgten die CSC-Segmentierung und das Matching. Die CSC-Segmentierung konstanter Bilder ergibt natürlich in jedem neuen Frame dasselbe Ergebnis, da der CSC-Algorithmus selbst deterministisch ist. Ein Beispiel mit 262 Segmenten ist in Abb. 3-7-2 dargestellt. Neben den CSC-Patch Segmenten sind auch die Verschiebungsvektoren der einzelnen Segmente zu sehen.

### 3. MATCHING-PARADIGMA

Die dargestellten Verschiebungsvektoren „auch als *Korrelationsvektoren* bezeichnet“ berechnen sich aus der Differenz der Mittelpunkte von jeweils zwei korrelierten CSC-Patches. Neben den CSC-Segmenten sind auch der Gültigkeitsstatus, welcher Auskunft über die Matching-Tauglichkeit der Segmente gibt sowie die völlig identischen Verschiebungsvektoren der einzelnen Segmente zu sehen. Hierzu gibt es 130 gültige Segmente, die sich aus einer Reduzierung auf ca. 50% der Segmente durch diesen Gültigkeitsstatus ergeben. Diese Entscheidung wird in Abhängigkeit von der Segmentgröße und Kompaktheit getroffen. Ein Auszug aus der berechneten Korrelationstabelle, die zu Abb. 3-7-2 gehört, ist in Tabelle 3-7-1 zu sehen.



**Abbildung 3-7-2:** Beispiel RGB Bild, CSC-Segmentierung und Korrelationsvektoren. Dazu ist der korrelierte konstante MB mit 130 gültigen Segmenten zu erkennen.

Die markierten Felder in der Korrelationstabelle/Zuordnungsmatrix zeigen die CSC-Patch Paare, die die größte Übereinstimmung haben. Diese sind im Fall der künstlich konstruierten Eingangsdaten natürlich auf den Diagonalen zu finden, da zu jedem Segment  $p_v^k$  genau das Segment  $p_\mu^{k+1}$  korrespondiert. Dennoch weist der Korrelationswert SM keine Übereinstimmung zu 100 Prozent auf. Diese Tatsache lässt sich durch die Betrachtung der Einzelkriterien erklären, aus denen sich das Gesamtähnlichkeitsmaß herleitet (Tabelle 3-7-2).

**Tabelle 3-7-1:** Auszug aus der 130 x 130 Elemente großen Korrelationstabelle

SM	$p_0^1$	$p_1^1$	$p_2^1$	$p_3^1$	...
$p_0^0$	0.997	0.546	0.529	0.473	
$p_1^0$	0.544	0.999	0.929	0.783	
$p_2^0$	0.523	0.929	0.998	0.818	
$p_3^0$	0.463	0.782	0.819	0.998	
...					...

Wie man in Tabelle 3-7-2 sehen kann, sind die Ähnlichkeitsmaße für die relative Segmentposition, -größe und -farbe bei einer Verschiebung eines konstanten Motion-

### 3. MATCHING-PARADIGMA

Blobs unbeeinträchtigt. Das Merkmal der Interframe-Distanz ist dies natürlich nicht. Bei diesem Ähnlichkeitsmaß steigt die Übereinstimmung zwischen zwei Segmenten, wenn sich deren Entfernung verringert.

**Tabelle 3-7-2:** Ausschnitt aus den Korrelationstabellen der Einzelkriterien (Beispiel aus 5-16)

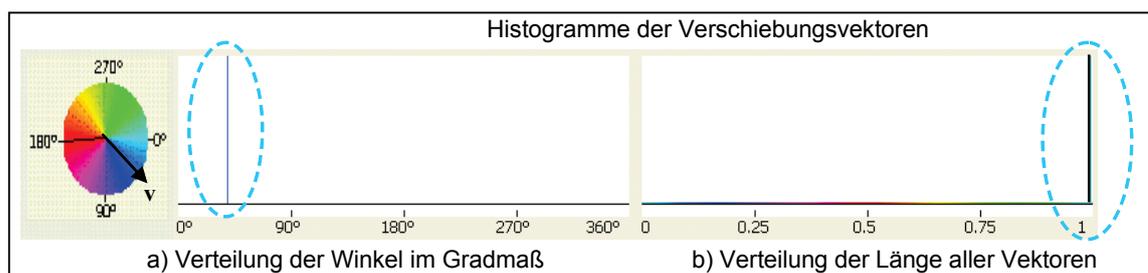
$SL_{v,\mu}$	$p_0^1$	$p_1^1$	$p_2^1$	$p_3^1$	$SI_{v,\mu}$	$p_0^1$	$p_1^1$	$p_2^1$	$p_3^1$
$p_0^0$	1	0	0	0	$p_0^0$	0.985	0.791	0.801	0.816
$p_1^0$	0	1	0.794	0.554	$p_1^0$	0.78	0.993	0.988	0.969
$p_2^0$	0	0.794	1	0.74	$p_2^0$	0.759	0.983	0.992	0.984
$p_3^0$	0	0.554	0.74	1	$p_3^0$	0.742	0.962	0.987	0.991

$SA_{v,\mu}$	$p_0^1$	$p_1^1$	$p_2^1$	$p_3^1$	$SC_{v,\mu}$	$p_0^1$	$p_1^1$	$p_2^1$	$p_3^1$
$p_0^0$	1	0.001	0.001	0.003	$p_0^0$	1	0.757	0.725	0.623
$p_1^0$	0.001	1	0.857	0.5	$p_1^0$	0.757	1	0.967	0.864
$p_2^0$	0.001	0.857	1	0.428	$p_2^0$	0.725	0.967	1	0.895
$p_3^0$	0.003	0.5	0.428	1	$p_3^0$	0.623	0.864	0.895	1

Da sich die Segmente aber fortbewegen, existiert eine Distanz  $d > 0$  zwischen ihnen und der Ähnlichkeitswert ist nicht 1. Dieses anscheinend kontraproduktive Maß, welches den Gesamtkorrelationswert SM zweier zugehöriger Segmente zu reduzieren vermag, macht in der Kombination mit weiteren Ähnlichkeitsmaßen dennoch Sinn, da es zum Ausschluss vieler anderer Segmente führen kann, die nicht miteinander korrelieren. Aufschluss über die Genauigkeit der Zuordnung aller CSC-Patches, gibt ein manueller Vergleich der Segmente bzw. eine Untersuchung der Korrelationsvektoren. Ein rein subjektiver Vergleich ist jedoch nur in begrenztem Umfang möglich und im Falle sehr kleiner Segmente fast ausgeschlossen. Ein manueller optischer Vergleich macht lediglich Sinn, um zu prüfen, ob auffällige Fehlzuordnungen bestehen. Ein automatischer Vergleich bietet sich durch die Untersuchung der Verschiebungsvektoren an. Da normalerweise jeder dieser Vektoren von der Güte der Segmentierung der zugrunde liegenden CSC-Patches  $p_v^k$  und  $p_\mu^{k+1}$  abhängig ist, stellt ein solcher Vergleich nur eine indirekte Überprüfung der Genauigkeit des Matchings dar.

Ein Überblick über die Verteilung der Winkel- und Längenwerte, wird durch die VV-Histogramme gegeben. Die Histogramme des letzten Beispiels finden sich in Abb. 3-7-3. Zur Berechnung des ersten Histogramms werden die Winkel in Schritten von einem Grad diskretisiert, woraus 360 Intervalle hervorgehen. Zur Berechnung des zweiten werden die Längen aller Vektoren auf Werte zwischen 0 und 1 skaliert und ebenfalls in 360 diskreten Intervallen zusammengefasst.



**Abbildung 3-7-3:** Histogramme der Korrelationsvektoren des Beispiels in Abb.3-7-2

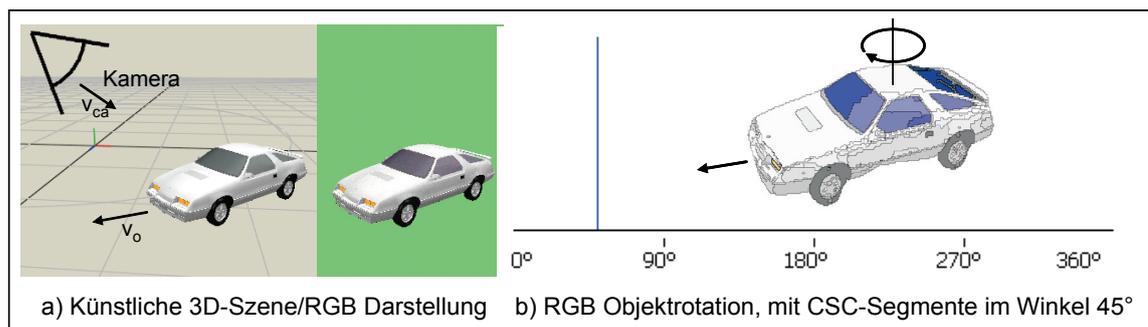
### 3. MATCHING-PARADIGMA

Die Konstellation, dass die dargestellten Histogramme über Winkel und Länge beide einen einzigen Peak enthalten, führt dazu, dass alle VV identisch sind und weiterhin, dass das Matching der CSC-Patches prinzipiell korrekt gearbeitet hat. Es hat sich in allen Matching-Tests, die unter denselben Konstellationen durchgeführt wurden, gezeigt, dass die CSC-Segmente zu 100 % korrekt korreliert wurden, unabhängig vom Segmentierungsgrad bzw. der Größe der MB-Regionen. Die Zuordnung funktioniert bei allen Segmentgrößen, auch im Falle sehr kleiner Segmente, die die Mindestgröße  $A_{min}$  mit einem Flächenmaß von nur 5 Bildpunkten aufweisen. Aus diesen Ergebnissen und dem Sachverhalt folgt, dass das Konzept, der auf Ähnlichkeitsmaßen beruhenden Korrelation von Bildmerkmalen prinzipiell gültig ist.

#### **Va2:** Funktionsfähigkeit des gesamten Matching-Prozesses mit variablen Segmenten

Nachdem gesichert ist, dass das 1:1 Matching, im Falle einfacher Bewegungen konstanter MBs und CSC-Patches richtig arbeitet, soll nun geprüft werden, wie genau das Verfahren bei veränderlichen MB- und damit auch CSC-Segmentierungen arbeitet, d. h. wie gut der gesamte entwickelte Verarbeitungsprozess unter möglichst idealen Bedingungen arbeitet.

Hierzu sollen die Vorteile künstlicher Bildsequenzen ausgenutzt werden, bei denen die Bewegungsparameter bekannt sind<sup>63</sup>. Eine beispielhafte Darstellung der Situation ist in Abbildung 3-7-4 zu sehen, in der die Kamera einen Sichtvektor  $v_{ca}$  besitzt, und ein Objekt eine Bewegung in Richtung des Vektors  $v_o$  ausführt. Das 3D-Modell eines Objektes wurde auf einer kreisförmigen Bahn durch den virtuellen Raum bewegt und dessen 2D-Bildprojektion ausgewertet. Die Erzeugung der Bilder von 3D-Modellen erfolgt mit Hilfe der Grafikschnittstelle OpenGL, die es unter geringem Aufwand ermöglicht Daten zu generieren, die sich für eine Überprüfung des Segmentierungs- und Matchingverfahrens eignen. Hierfür wird eine virtuelle Aufnahmeszene entworfen, in der die zu erfassenden Objekte eine fest definierte Position einnehmen, bzw. sich auf einer vordefinierten Bahn bewegen.



**Abbildung 3-7-4:** Generierung einer künstlichen Testszene, mit Kamera

Weiterhin wird eine Kamera erschaffen, die bestimmte Aufnahmeparameter besitzt, welche benötigt werden, um die Szene darzustellen und die virtuellen 3D-Bewegungsparameter auf die Bildfläche zu transformieren. Bedingt durch die

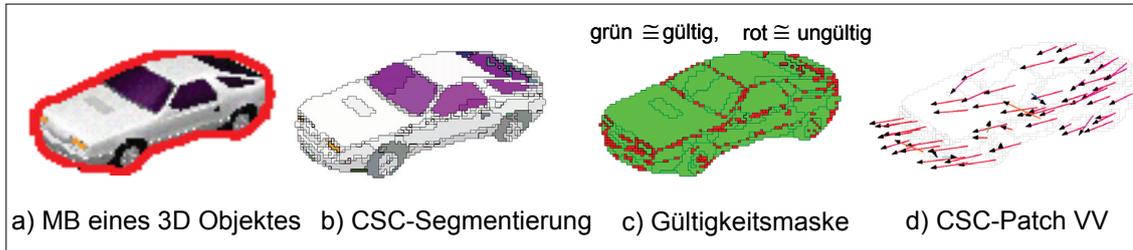
<sup>63</sup> Es ist an dieser Stelle jedoch nicht sinnvoll, wie in der ersten Testphase, mit RGB Abbildern von Objekten zu arbeiten, da eine Erfassung eines Abbildes im Bild, keine realistische Überprüfung der Funktionsfähigkeit des Segmentierungs-, bzw. des Multi-Matching-Verfahrens darstellt. Daher wird mit 3D-Modellen gearbeitet, deren räumliche Bewegungsparameter auf die Bildfläche projiziert werden können, um diese als repräsentative Grundwahrheit, qualitativ mit den berechneten CSC-Segmenten, Verschiebungsvektoren zu vergleichen.

### 3. MATCHING-PARADIGMA

Rotationsbewegung und die Perspektive besteht zwischen den Bildern eine markante Formvariation des Objektes. Daher enthält auch diese künstliche Szene Merkmale einer einfachen nicht-kooperativen Bildaufnahmesituation<sup>64</sup>.

Ziel der nun durchgeführten Analyse ist es, den Grad der Funktionsfähigkeit des 1:1 Matchings, bei mäßig verändertem Bildinhalt, insbesondere bei variierenden CSC-Patch Segmenten zu messen. Zur Durchführung der Untersuchung bewegt sich ein Objekt auf einer kreisförmigen Bahn und befindet sich dabei ständig im Blickfeld der Kamera. Aufgrund der Objektbewegung erfolgen bei der Abbildung Änderungen, welche durch Translation, Rotation und perspektivisch bedingte Größenverzerrung entstehen. Außerdem bestehen leichte Helligkeitsvariationen der Objektoberfläche, die durch ein einfaches Beleuchtungsmodell hervorgerufen werden. Die genannten Änderungen führen zu Variationen bei der Segmentierung der MB und der CSC-Patches.

Die Erfassung des Objektes erfolgt durch das vorgeschlagene bewegungsbasierte Verfahren sehr genau (Abb. 3-7-5a). Ein vergleichbares Ergebnis wurde für alle Bilder der Sequenz erzielt. Bedingt durch die sehr homogenen Flächen, welche aus dem mangelnden Texturgehalt des künstlichen Objektes resultieren, wurden einige sehr große Segmente durch den CSC-Algorithmus ermittelt. Diese werden dazu führen, dass das Verschiebungsvektorfeld (VVF) an einigen Stellen weniger dicht ausfällt (Abb. 3-7-5d). In der untersuchten künstlichen Sequenz besitzt im Mittel jedes zweite Segment einen positiven Gültigkeitsstatus (Abb. 3-7-5c). Solche Segmente zeigen eine sehr kleine Fläche oder einen hohen Kompaktheitswert, was auf eine zerklüftete, entartete Form hindeutet. Deshalb ist es sinnvoll diese als ungültig zu klassifizieren und zur präventiven Fehlervermeidung von der Korrelation auszuschließen.



**Abbildung 3-7-5:** Ausschnitt eines Bildpaares aus der Korrelationsprüfsequenz

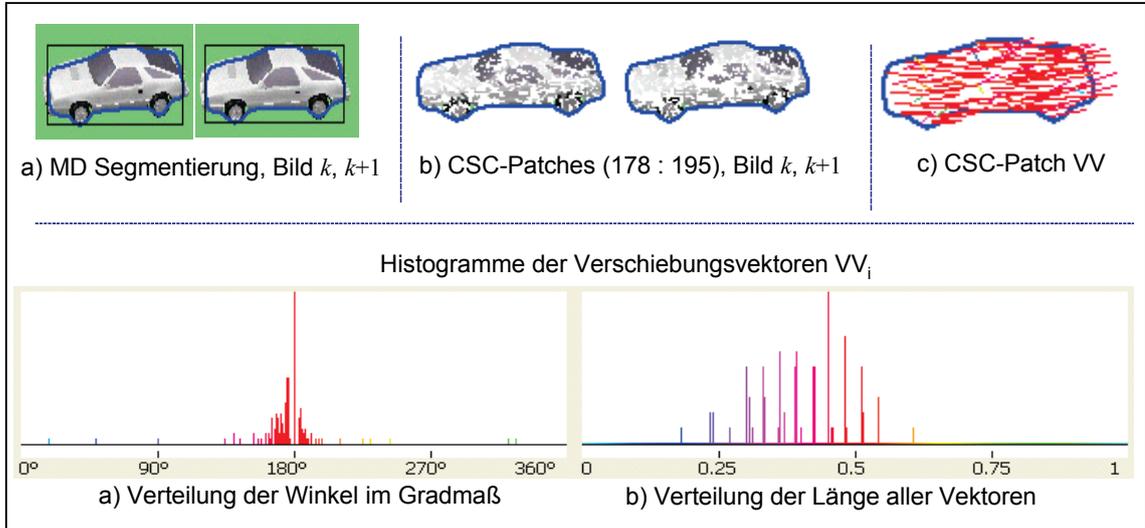
Die Prüfung, ob die Korrelation, der in den Folgebildern erzeugten Segmenten, korrekt arbeitet, soll durch quantitative und qualitative Messungen erfolgen, bei denen die bekannten Bewegungsparameter zum Vergleich eingesetzt werden. Zur Veranschaulichung des Ganzen, wird in Abbildung 3-7-6 ein Bildpaar aus der Sequenz dargestellt, für das die Korrelationsvektoren der CSC-Patches berechnet wurden.

In diesem Beispiel ist das Ergebnis mit hohem Segmentierungsgrad (siehe Abb. 3-7-6b), welches eine mäßige Translation und eine schwache Rotation, sowie Größenänderung beinhaltet. Das Histogramm der ermittelten Verschiebungsvektoren zeigt eine ausgeprägte Häufung der Winkel bei ca. 180 Grad (Abb. 3-7-6a und b unten). Das Histogramm der VV-Länge, zeigt eine etwas stärkere Streuung der Werte. Aus der Häufung im Histogramm der Verschiebungsvektorenwinkel ergibt sich, dass die

<sup>64</sup> Wie in Abb. 3-7-4 zu erkennen ist, wird in den künstlichen 3D-Szenen kein Hintergrund modelliert. Ein Hintergrund ist für die, an dieser Stelle durchzuführenden Tests nicht unbedingt notwendig, da die bewegungsbasierte MB-Erfassung und damit auch die CSC-Patch-Erstellung, gut ohne auskommt. Wichtig ist allerdings, dass die künstlichen Objekte über genug Textur verfügen, um im Differenzbild nicht herausgefiltert zu werden.

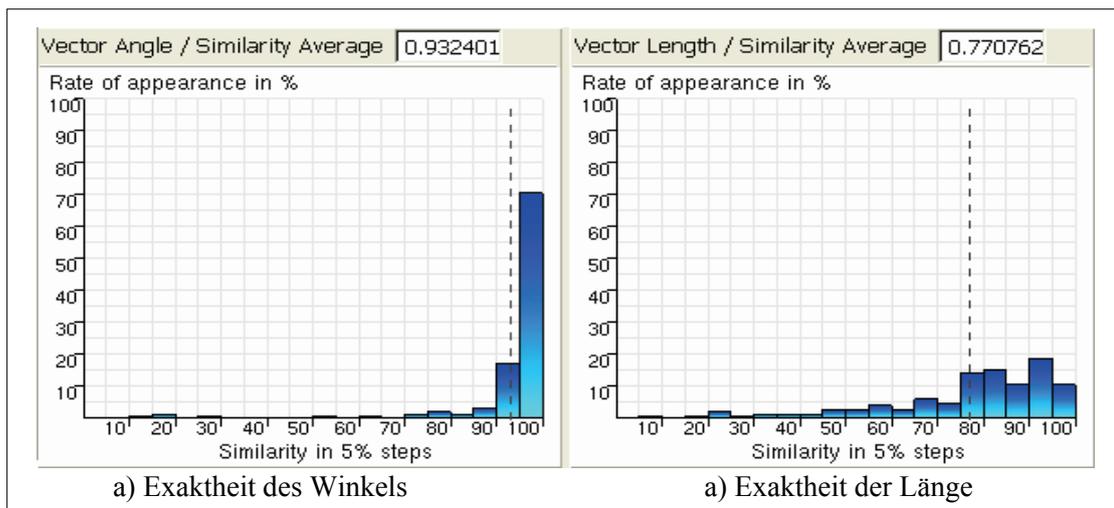
### 3. MATCHING-PARADIGMA

Korrelationspartner der meisten CSC-Patches in ein und derselben Richtung zu finden sind, was positiv zu bewerten ist. Aus der Streuung der aufgetretenen Vektorenlängen folgt jedoch, dass die Verschiebungsvektoren vieler CSC-Patches eine weniger genaue Übereinstimmung ihrer Länge, mit der tatsächlichen Länge des Verschiebungsvektors aufweisen.



**Abbildung 3-7-6:** Histogramme der Korrelationsvektoren des Beispiels in Abb. 3-7-5

Zur Illustration dieser Schlussfolgerung erfolgt in Abb. 3-7-7 eine Darstellung der Häufigkeit des Auftretens von Verschiebungsvektoren (VV), in Abhängigkeit vom Grad der Exaktheit, d.h. in Abhängigkeit vom Grad der Übereinstimmung zum tatsächlichen Bewegungsvektor. Der Grad der Übereinstimmung erfolgt jeweils in 5% Schritten. Ein Beispiel wie dieses spezielle Histogramm zu lesen ist, ist folgendes: “Die Verschiebungsvektoren, die eine Übereinstimmung des Winkels zu 95-100% aufweisen, haben einen Anteil von 70 % an der Gesamtmenge aller Verschiebungsvektoren“.



**Abbildung 3-7-7:** Häufigkeit des Auftretens von VV in Abhängigkeit von der Exaktheit

Der Übereinstimmungswert zweier Vektoren  $\mathbf{v}_1^k = (v_x, v_y)$  und  $\mathbf{v}_2^k = (v_x, v_y)$ , wird hierbei prinzipiell durch den Vergleich der Winkel  $\varphi_1$  und  $\varphi_2$  bzw. der Vektorlänge  $l_1$  und  $l_2$

### 3. MATCHING-PARADIGMA

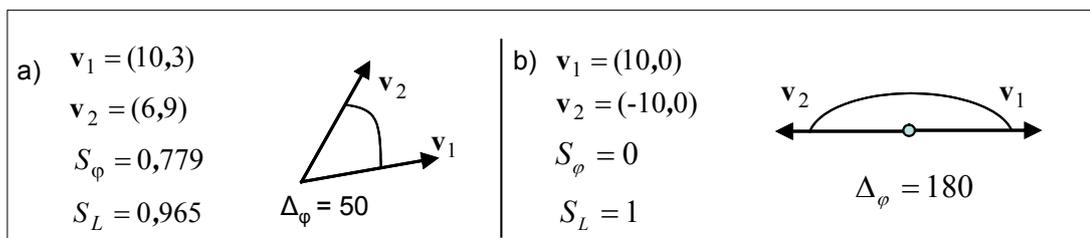
errechnet. Die aus dem Vergleich berechnete Ähnlichkeit des Winkels  $S_\varphi$  ergibt sich aus Gleichung (3-7-2b) und die der Länge  $S_L$  aus Gleichung (3-7-2c).

$$\varphi_i = \arctan\left(\frac{v_y}{v_x}\right) \text{ und } l_i = \sqrt{v_x^2 + v_y^2} \quad (3-7-2a)$$

$$S_\varphi = \begin{cases} 1 - \frac{360 - |\varphi_1 - \varphi_2|}{180}, & \text{falls } |\varphi_1 - \varphi_2| > 180 \\ 1 - \frac{|\varphi_1 - \varphi_2|}{180}, & \text{sonst} \end{cases} \quad (3-7-2b)$$

$$S_L = \begin{cases} \frac{l_1}{l_2}, & \text{falls } l_2 > l_1 \\ \frac{l_2}{l_1}, & \text{sonst} \end{cases} \quad (3-7-2c)$$

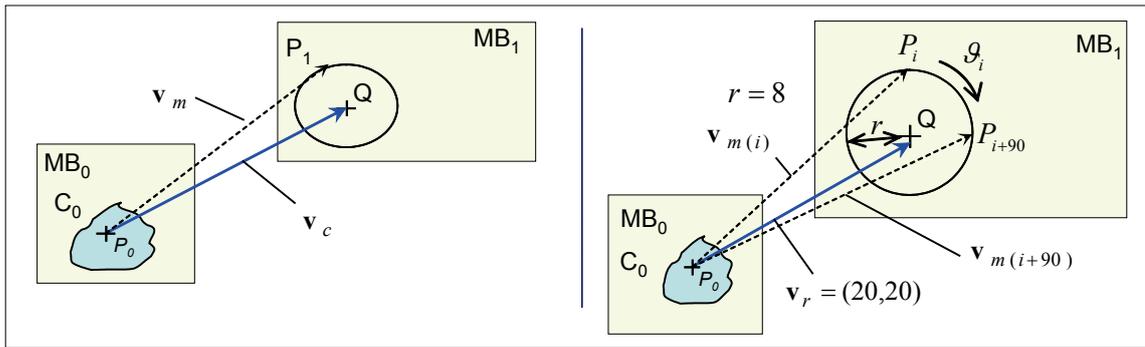
Beispiele für die Ähnlichkeiten  $S_\varphi$  und  $S_L$  sind in Abb. 3-7-8 dargestellt. Aus der Sichtweise in Abb. 3-7-8 wird deutlich, dass die Übereinstimmung der VV-Winkel, mit dem Winkel des tatsächlichen Bewegungsvektors stärker ausgeprägt ist, als die der Vektorlänge. Diese Ergebnisse bestätigen die vorangegangene Histogrammuntersuchung, bei der die Winkel der Vektoren das bessere Resultat, mit geringerer Streuung lieferten. Ähnliche Ergebnisse wie in diesem Beispiel, wurden in zahlreichen Tests des Multi-Matching Verfahrens erzielt, in denen künstlich generierte Objekte, im Zusammenhang mit einer sehr kleinen Segmentgröße verwendet wurden. Darüber hinaus ist auch in Untersuchungen realer Szenen, welche im Abschnitt 3.8 betrachtet werden, festzustellen, dass die Histogramme der VV-Längen eine stärkere Streuung aufweisen.



**Abbildung 3-7-8:** Beispiele für die Ähnlichkeitsbestimmung von Vektoren

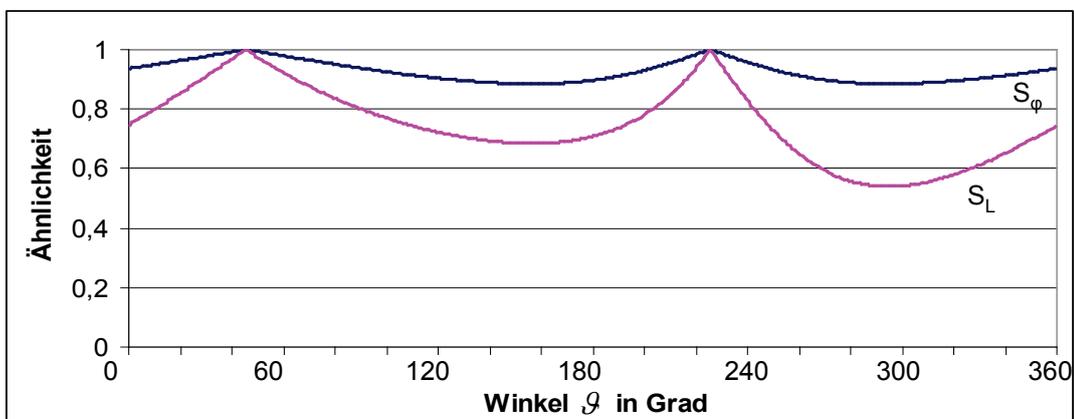
Das Phänomen der generell ungünstigeren Verteilung der VV-Länge, erscheint für die Überprüfung des gesamten Verfahrens interessant und soll daher etwas näher betrachtet werden. Ursächlich ergibt sich dieser Effekt aus den Variationen, die während der CSC-Patch-Segmentierung auftreten. Die CSC-Segmente in Folgebildern, bzw. in korrelierten MB, weisen zwar bestimmte Ähnlichkeitsmerkmale auf, welche die Basis für das 1:1 Matching darstellen, dennoch haben sie Unterschiede in Form, Position und auch Farbe. Der relevante Parameter für die Ermittlung des VV eines CSC-Patches  $p_v$  ( $C_0$ ) ist dessen ursprüngliche Position  $P_0$ .

### 3. MATCHING-PARADIGMA



**Abbildung 3-7-9:** Jeder Korrelationsvektor  $\mathbf{v}_m$  eines CSC-Patches  $C_0$  liegt im Umfeld des idealen Vektors  $\mathbf{v}_c$  und auf der rechten Seite. Beispiel für die Simulation einer kreisförmig verteilten Streuung der Korrelationsvektor  $\mathbf{v}_m$ .

Die ideale Position für den  $p_v^k$  Nachfolger  $p_u^{k+1}$  sei der Punkt  $Q$  (siehe Abb. 3-7-9). Dieser ideale Punkt  $Q$  ergibt sich aus  $p_v^k$  und der Verschiebung um den normalerweise unbekannt, tatsächlichen Bewegungsvektor  $\mathbf{v}$ . Durch die Variationen bedingt, stimmt die Position von  $p_v^k (C_0)$ , welche  $P_1$  genannt wird, im Allgemeinen nicht mit  $Q$  überein. Stattdessen befindet sich  $P_1$  irgendwo innerhalb eines gewissen Umfeldes von  $Q$ . Daher existiert für jeden CSC-Patch  $C_i$  eine andere Abweichung des entsprechenden  $\mathbf{v}_m$  von idealen  $\mathbf{v}_r$ . Weiterhin bedeutet dies, dass es im Histogramm der  $\mathbf{v}_m$  eine gewisse Streuung gibt. Berücksichtigt man nun, dass die  $\mathbf{v}_m$  eine Ablenkung, innerhalb eines zufälligen, durch die Segmentierung hervorgerufenen Umfeldes erfahren, ergeben sich hieraus Implikationen für die Verteilung der  $\mathbf{v}_m$ -Winkel und Längen. Ferner lässt sich durch die Berücksichtigung der besonderen Variation der  $\mathbf{v}_m$  auch die ungleichmäßige Verteilung aus Abb. 3-7-9 erklären. Dies soll nun an einem Beispiel (Abb. 3-7-10) verdeutlicht werden, in dem eine kreisförmige Streuung der  $\mathbf{v}_m$  simuliert wird. Hierzu werden  $n=260$  virtuelle Korrelationsvektoren  $\mathbf{v}_m$  zwischen je einem festen Ausgangspunkt  $P_0$  und einem weiteren Punkt  $P_i$ , welcher auf der Peripherie eines Kreises mit dem Radius  $r$  liegt, erzeugt. Der Winkel  $\vartheta_i$  gibt die Position des Punktes  $P_i$  auf der Kreisperipherie an. Weiterhin wird in der Simulation, der tatsächliche Bewegungsvektor  $\mathbf{v}_c$  als gegeben angenommen.



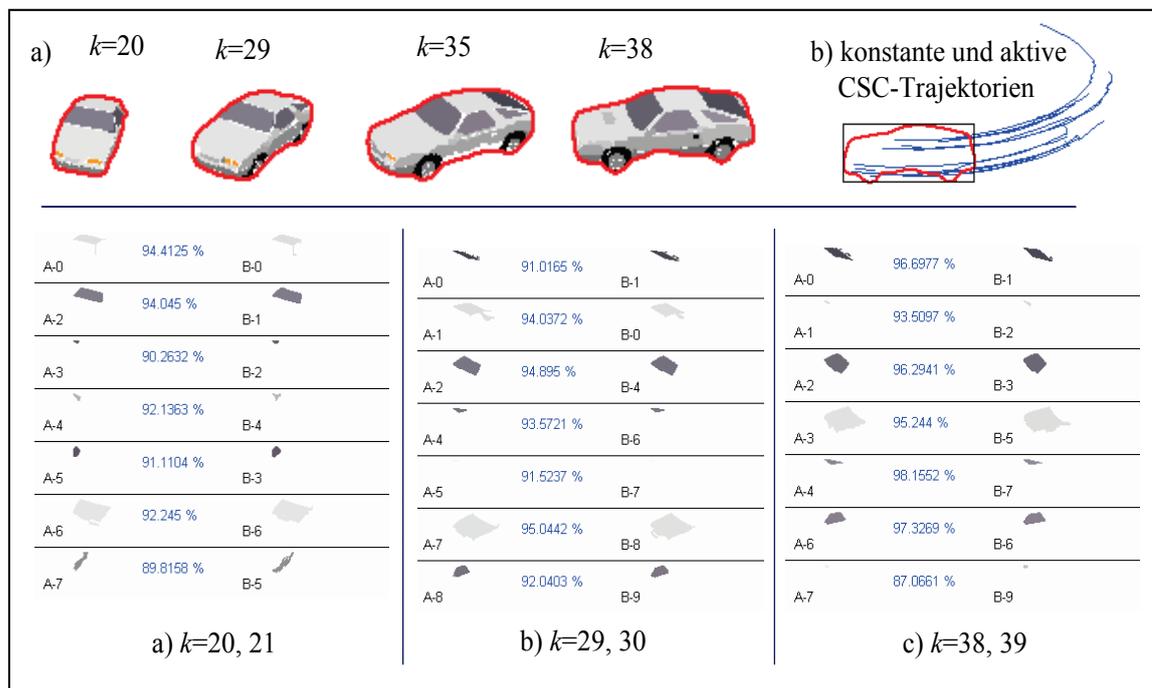
**Abbildung 3-7-10:** Verteilung der Winkel- und Längenähnlichkeiten der Vektoren

### 3. MATCHING-PARADIGMA

Die verwendeten Größen  $r$  und  $v_c$  entsprechen den Parametern einer durchschnittlichen Aufnahmesituation. Betrachtet man nun die Ähnlichkeitswerte zwischen  $v_c$  und  $v_m$  in Abhängigkeit des Winkels  $\vartheta_i$ , so ergibt sich für die Winkel- und Längenähnlichkeiten ( $S_\varphi$  und  $S_L$ ) folgende interessante Darstellung (Abb. 3-7-10).

Das Ergebnis der simulierten, kreisförmigen Streuung der  $v_m$  macht deutlich, dass die Übereinstimmung des Winkels, zwischen dem tatsächlichen Vektor  $v_r$  und den fiktiven Korrelationsvektoren  $v_{m(i)}$ , stets höher ist, als die Übereinstimmung der Länge. Weiterhin weist das Ergebnis der Winkelähnlichkeit  $S_\varphi$  eine geringere Schwankung auf, was einer geringeren Streuung in der Verteilung des Ergebnisses entspricht. Auch wenn diese Simulation eine Vereinfachung, der in tatsächlichen Messungen zufälligen, evtl. nicht gleich verteilten Streuung des Korrelationsvektors  $v_m$  darstellt, so wird dennoch gezeigt, wieso es zu auffälligen Unterschieden bei der Genauigkeit der Winkel- und Längenähnlichkeitswerte kommt.

Als Resultat dieser Untersuchung kann die Feststellung getroffen werden, dass Variationen der  $v_m$ -Länge, bis zu einem bestimmten Grad normal sind und diese daher nicht automatisch die Qualität des 1:1 Matchings herabsetzen. Die Variationen sind stets vorhanden, da sie aus den natürlichen Abweichungen der CSC-Segmentierung hervorgehen. Weiterhin ist es normal, dass die Qualität der  $v_m$ -Winkel im Allgemeinen höher ist als die der  $v_m$ -Länge. Folglich geben die  $v_m$ -Winkel ein gleichförmigeres Bild vom Korrelationsergebnis. Durchgeführte Tests mit künstlichen Objekten ergaben stets ähnliche Ergebnisse wie im Beispiel aus Abbildung 3-7-11.



**Abbildung 3-7-11:** Sequenz mit extra großen CSC-Segmenten bei der Objektbewegung. Unten ist ein Screenshots vom 1:1 CSC-Patch Matching, innerhalb der Bildsequenz dargestellt.

Weiterhin wurde die Qualität der  $v_m$  in längeren künstlichen Sequenzen betrachtet, in denen eine große Anzahl von CSC-Patch Segmenten einander zugeordnet worden sind. Hier hat sich in diesen Sequenzen gezeigt, dass die paarweise Korrelation von CSC-Segmenten eines abgebildeten Objektes invariant gegenüber Objektrotation und Vergrößerung ist, d.h. dass in zwei Folgebildern stets CSC-Segmente gefunden werden,

### 3. MATCHING-PARADIGMA

die einander zugeordnet werden können. Eine Veranschaulichung mit Ausschnitten aus einer Sequenz erfolgt in Abbildung 3-7-11.

Man kann in Abbildung 3-7-11 deutlich erkennen, dass die großen Segmente in den Zeitraffer-Folgebildern ihre Form beträchtlich verändern und zum Teil auch verschwinden. Die Veränderung zwischen zwei Bildern fällt jedoch meist nur gering aus. Weiterhin sind in Abbildung 3-7-12 die Trajektorien von unmittelbaren CSC-Patch Nachfolgern zu sehen, die beim 1:1 Matching einfach miteinander verkettet wurden (eine Verkettung funktioniert z.B. so: „ $C_1$  wird mit  $C_2$  korreliert,  $C_2$  wiederum mit  $C_3$ ,  $C_3$  mit  $C_4$ , usw.“).

Anhand der langen Trajektorien in Abbildung 3-7-11 erkennt man, dass trotz der perspektivischen Verzerrung, Rotation, etc. eine gewisse Permanenz/ Konstanz bei der Segmentzuordnung besteht. Diese Beständigkeit ist auch bei sehr kleinen Segmenten zu finden. Die Gleichförmigkeit der CSC-Segment-Trajektorien ist ein starker Hinweis auf die Korrektheit des hier vorgestellten Multi-Matching Verfahrens. Beispielhaft kann die Übereinstimmung korrelierter Segmente betrachtet werden. Die Ergebnisse des Matchings führen in der Sequenz zu einem saubereren, ungestörten Verschiebungsvektorfeld (Abb. 3-7-12). Die Dichte des Verschiebungsvektorfeldes (VVF) ist jedoch an solchen Positionen geringer, an denen das Objekt weiter entfernt ist und folglich durch weniger CSC-Segmente abgebildet wird.

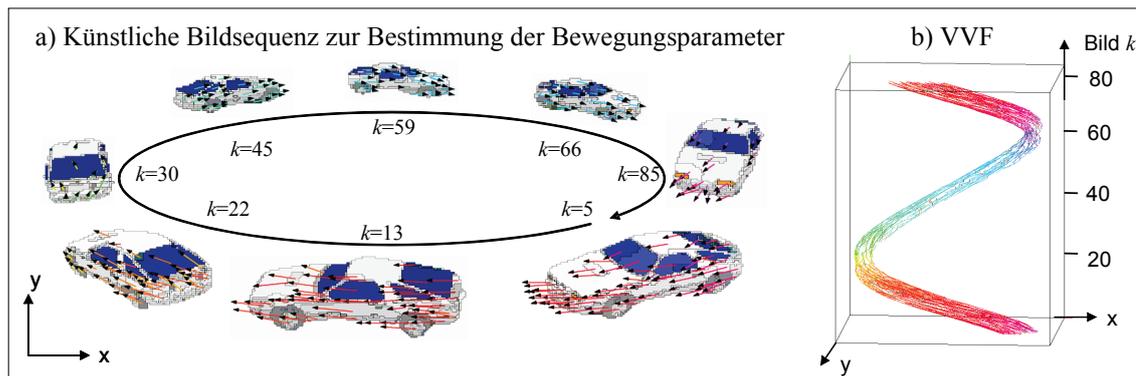


Abbildung 3-7-12: Berechnung des Verschiebungsvektorfeldes.

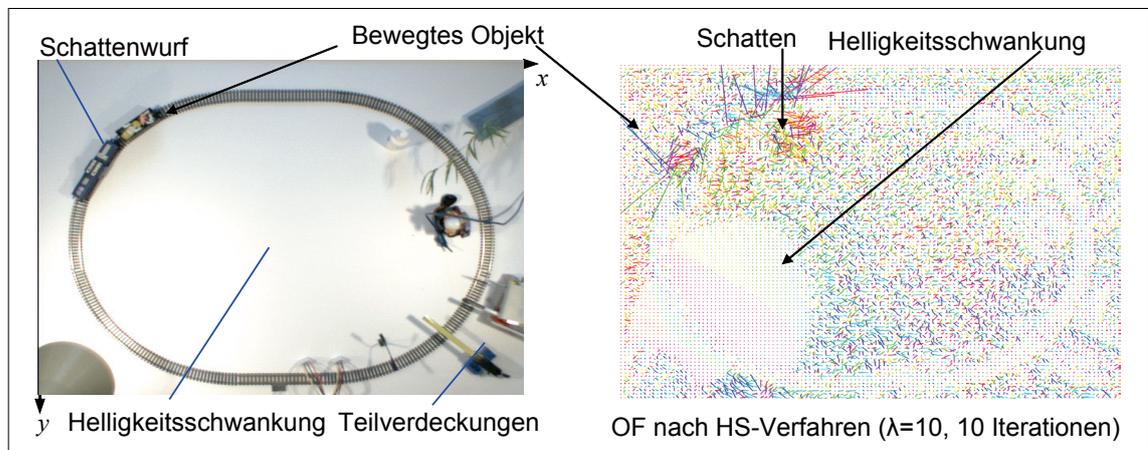
Die durch künstlich generierte Eingangsdaten gewonnenen Erkenntnisse über das Multi-Matching geben genügend Motivation, um das Verfahren mit realen Farbbildsequenzen zu testen. Im Anschluss werden Ergebnisse aus Untersuchungen „anspruchsvollerer“, realer Bildsequenzen vorgestellt, die nicht-kooperative Messsituationen enthalten.

## 3.8 Experimentelle Ergebnisse

Im Folgenden werden Untersuchungsergebnisse vorgestellt, die durch Anwendung des entwickelten Matching-Paradigmas gewonnen wurden. Dazu wurden Resultate aus künstlichen Bildsequenzen im Abschnitt (3.7) betrachtet, bevor auf reale, gestörte Datensätze von Verkehrsaufnahmen eingegangen wird. Die Korrespondenzprobleme herkömmlicher intensitätsbasierter Verfahren (s. Kapitel 2), die aus den beschriebenen bildspezifischen Störungen resultieren, werden vergleichend mit den Ergebnissen des entwickelten Lösungsansatzes beschrieben, um die Leistungsfähigkeit und Effizienz des entwickelten Matching-Paradigmas bei der Analyse von realen Bildsequenzen, nämlich genau bei dem Einfluss von nicht-kooperativen Messsituationen, zu demonstrieren.

### 3.8.1 Einfluss von nicht-kooperativen Messsituationen

Die erste Bildsequenz entstand durch die Aufnahme einer Modelleisenbahn, bei der durch verschiedene Maßnahmen versucht wurde, ausgewählte Störungen zu simulieren. Diese Testszene enthält nicht-kooperativen Messsituationen, wie starke Beleuchtungsänderungen, Schattenwurf, größere Verschiebungen, die Teilverdeckungen und Formveränderung des bewegten Objektes durch Rotation (Abb. 3-8-1). Dazu kommen auch leichtes Rauschen und homogene Bildbereiche. Vorteilhafte Eigenschaften für eine Bewegungsanalyse sind hingegen die hohe Bildhelligkeit und eine kleine perspektivisch bedingte Größenverzerrung des einzigen Objektes.



**Abbildung 3-8-1:** Testszene Modelleisenbahn und Variationen der Bildhelligkeit durch eine Lampe führen zu falschen Verschiebungsvektoren bei der Verwendung des im Kapitel 2 vorgestellten optischen Flussverfahren (z.B. OF nach HS-Verfahren).

Ausgehend von der Tatsache, dass die Überlagerung dieser realen Bildsequenz von nicht-kooperativen Messsituationen eine starke Modifikationen der Intensitätswerte verursacht, resultieren hier großflächige Variationen der Helligkeitswerte aufgrund der intensiven Beleuchtungsschwankungen sowie den durch Schattenwurf bedingten Verdunklungen. Diese lösen ein unvermeidliches Korrespondenzproblem aus, welches zur falschen Bestimmung der Verschiebungsvektoren führen wird. Zur Bewertung der Analyseleistung des vorgeschlagenen Matching-Paradigmas soll erst die Analyse mittels der klassischen intensitätsbasierten Verfahren im Folgenden dargestellt werden.

### 3. MATCHING-PARADIGMA

---

Tendenziell können mit den differentiellen Verfahren *nur* recht kurze VV korrekt bestimmt werden, solange die Beleuchtung konstant ist und der Grauwertverlauf linear bleibt, was in dieser realen Sequenz jedoch nicht zu gewährleisten ist (s. Kapitel 2). Will man trotzdem ein intensitätsbasiertes Verfahren wie das HS-Verfahren oder das LK-Verfahren zur Bestimmung von Verschiebungsvektoren einsetzen, sollte man mit Fehlinterpretationen aufgrund der starken Veränderung der Beleuchtung und der größeren Verschiebungen rechnen. Die Veränderung der Beleuchtung führt zu einer Änderung der Bildirradienzen und wird bei der Auswertung ein mittels differentiellen Verfahren geschätztes optisches Verschiebungsvektorfeld ergeben, das nicht unbedingt die tatsächliche Bewegung beschreibt (Abb.3-8-1).

Schließt man nun von optischen Flussvektoren auf eine Bewegung des Gegenstandes, kommt es zu einer Fehleinschätzung der realen Umgebung. Dazu ist in der Regel mit systematischen Fehlern vor allem an den Objektändern oder in den Bereichen mit einem harten Objektübergang (ohne Unschärfe) zu rechnen. Dieses Problem entsteht unter anderem durch den sog. *Short-Range-Mechanismus* des angewandten differentiellen Verfahrens, d.h. durch die Tatsache, dass nur eine minimale Nachbarschaft bei der Korrelation der Bildwerte berücksichtigt wird. Trotz der Verwendung mehrerer Iterationen zur Glättung der Verschiebungsvektoren ist ein Teil der Falschdetektionen in Szene 3-8-1 auf größere Verschiebungen und Störungen in den Bildbereichen zurückzuführen, was folglich die Qualität der ermittelten Bewegungstrajektorien verschlechtert (Abb. 3-8-1). Interessant sind auch die starken Variationen der Verschiebungsvektoren im Bereich der homogenen Regionen in der Abbildung, die auf ein Blendenproblem hindeuten.

Im Gegensatz zu den differentiellen Verfahren arbeiten die Korrelationsverfahren (wie Blockmatching) zuverlässiger, weil sie erstens die Schätzung großer Verschiebungen erlauben und zweitens eine höhere Stabilitätsanforderung aufgrund der einbezogenen ausgeprägten Primitive aufweisen (s. Kapitel 2). Da aber die berechneten Verschiebungsvektoren bei Blockmatching auf einem flächenbasierten Ähnlichkeitsmaß basieren, ist hier mit einer zuverlässigen Lokalisierung des Extremwertes des Ähnlichkeitsmaßes in realen einfachen Sequenzen zu rechnen (Abb. 3-8-2a). Die Verbesserung ist auf den so genannten *Long-Range-Mechanismus* zurückzuführen, der darauf beruht, dass eine maximale Nachbarschaft bei der Korrelation berücksichtigt wird, die auch ausgeprägte Primitive beinhalten kann.

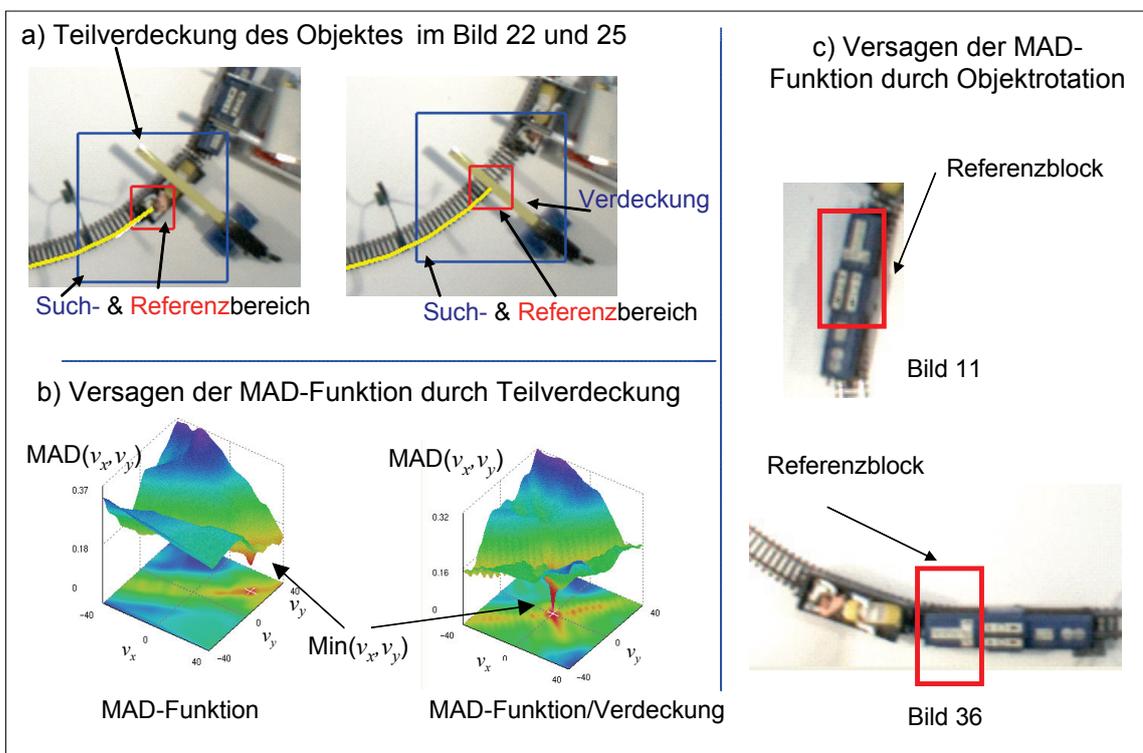
Allerdings lösen die nicht-kooperativen Messsituationen wie beispielsweise inhomogene Beleuchtungsänderungen und Teilverdeckungen hier ein Korrespondenzproblem aus, das zur Anhebung oder Absenkungen des Wertebereiches (abhängig von dem Helligkeitsübergang) sowie zur Deformationen der berechneten Ähnlichkeitsmaßen führt. Dadurch können mehrere Nebenminima oder Nebenmaxima (je nachdem, was für ein Ähnlichkeitsmaß man verwendet hat) auftreten, die das richtige dominieren.

Die Einbeziehung von Farbinformation kann zur Verbesserung der Verschiebungsvektorermittlung mittels Blockmatching eingesetzt werden. Allerdings beruht die Verbesserung der Matchinggüte erstens auf der optimalen Farbbeschreibung, die ein Multi-Matching in ikonischer Ebene ermöglichen wird und zweitens auf der konkreten Aggregation der Farbkomponente, die eine helligkeitsunabhängige adaptive Komponente ermöglicht. Immerhin sollte man hier mit höherem Rechenaufwand und Komplexität rechnen [Al-Hamadi 03g, 03f].

Weiterhin entsteht ein weiterer Störeffekt durch die mehrfachen Teilverdeckungen des Objektes in dieser Sequenz, die während der Bildfolge eintreten (Abb. 3-8-2a). Diese

### 3. MATCHING-PARADIGMA

Teilverdeckungen bewirken Deformationen wie z.B. mehrere Minima in der angewandten MAD-Funktion (Abb. 3-8-2a/b), wodurch fehlerhafte Ergebnisse im resultierenden Verschiebungsvektor entstehen. Das führt bei der Analyse dazu, dass der betrachtete Bildblock scheinbar von dem ursprünglich verfolgten Block wegdriftet (Abb. 3-8-2a/b). Eine ausschließliche Suche des Minimums „aufgrund der angewandten MAD-Funktion“ detektiert in diesem Fall fälschlicherweise den Vordergrund (Abb. 3-8-2a). Die Einbeziehung solcher fehlerhaften Messergebnisse verursacht eine Verfälschung nachfolgender Analyseergebnisse, oder bedingt weiteren Rechenaufwand, wenn mit zusätzlichen Verfahren versucht wird, solche Fehlmessungen nachträglich zu eliminieren bzw. zu korrigieren. Zu diesem Zweck könnte ein rekursives Schätzverfahren, zum Beispiel der Kalman-Filter, eingesetzt werden [Mecke 99, Al-Hamadi 03f, Pathan 09c], das aber eine gewisse Einschränkungen bezüglich des Rechenaufwandes, der Linearität und Flexibilität beinhaltet [Al-Hamadi 03g].



**Abbildung 3-8-2:** Abbruch der Blockmatching-Objektverfolgung wegen Teilverdeckung und c) Referenzblock wird durch Rotation ungültig.

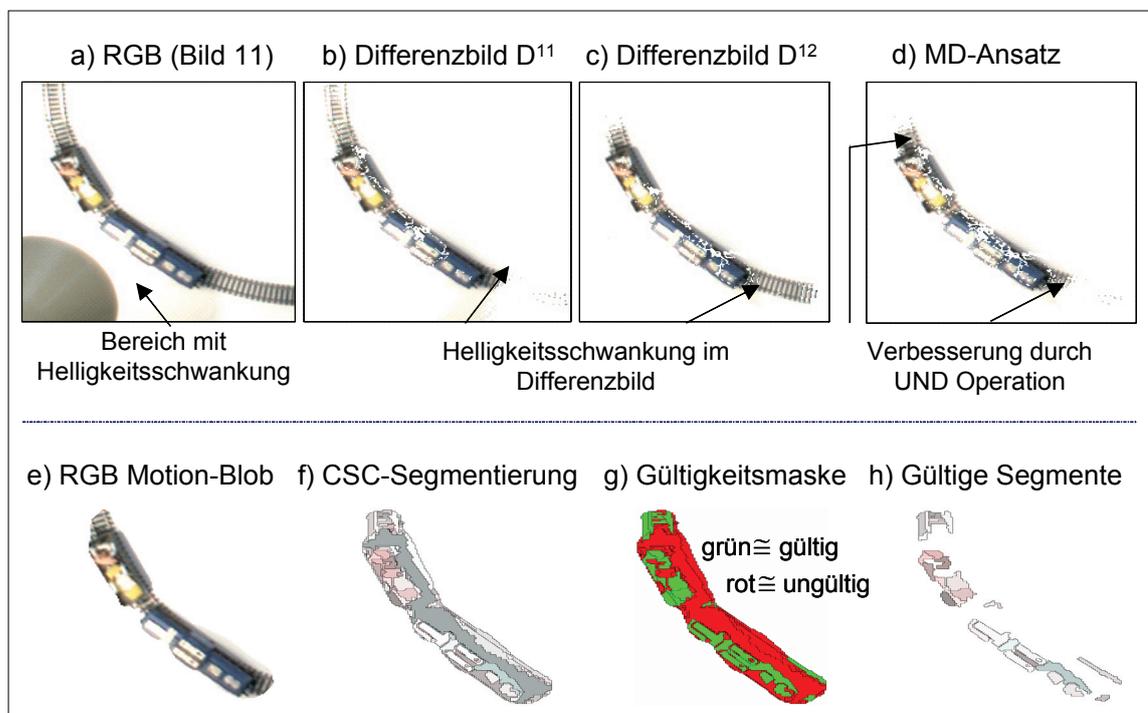
Weiterhin verursacht die durch die Rotationsbewegung bedingte Lageveränderung bzw. Formveränderung des Objektes eine Störung von Korrelationsverfahren, die mit einem festen Blockraster arbeiten (Abb. 3-8-2c). Durch den Eingriff des Anwenders kann die Blockgröße optimiert werden. Dies ist aber unter dem Aspekt der technischen Realisierbarkeit schwierig, da es angesichts der Mannigfaltigkeit realer Messobjekte und möglicher Einflussfaktoren nicht praktikabel erscheint, für jede Konstellation (Objekt und Szene) den subjektiven Eingriff des Anwenders zu realisieren. Vielmehr erscheint es sinnvoll, eine adaptive Blockgröße vorzunehmen, die aber eine automatische Vorsegmentierung voraussetzt. Generell wird eine Verbesserung bei diesen Verfahren dadurch erreicht, dass mehrere Informationen zur Berechnung der VV

### 3. MATCHING-PARADIGMA

herangezogen werden. Ein Nachteil ist jedoch der hohe Rechenaufwand, der keine echtzeitfähige Bearbeitung der Daten erlaubt.

Nun soll mittels des vorgeschlagenen Matching-Paradigmas in zwei Merkmalsebenen die echtzeitfähige Berechnung und Auswertung von Verschiebungsvektoren für das gesamte Bild trotz des Einflusses von nicht-kooperativen Messsituationen gewährleistet werden. Dieses Matching-Paradigma umgeht die durch Helligkeitsschwankungen, Teilverdeckungen und Schattenwurf bedingten Störungen. Die Zuverlässigkeit der Berechnung beruht dadurch erstens auf der robusten automatischen Erfassung von bewegten Regionen und zweitens auf der eleganten Trennung der Bildmerkmale und effizienter Merkmalszuordnung in zwei hierarchischen Ebenen. Da durch die vorgeschlagene modifizierte Differenzbildtechnik (MD) das gesamte bewegte Objekt erfasst wird und nicht nur kleine Teile betrachtet werden (wie im Fall des Blockmatchings). Beispielsweise kann das Blendenproblem aus rein lokaler Sicht praktisch nicht mehr auftreten, welches aus der Betrachtung eines homogenen bzw. texturlosen Blocks eines Objektes resultieren kann.

Der differenzbildbasierte Segmentierungs- und Nachverarbeitungsalgorithmus liefert in allen Bildern der untersuchten Sequenz exakte Ergebnisse (Abb. 3-8-3d/e). Die Nulldurchgänge in dem resultierenden Differenzbild (MD) geben die genaue Position des bewegten Objektes an, womit das Problem der einfachen Differenzbildtechnik (z.B. Ghostingeffekt) klar beseitigt wird. Die enthaltenen Beleuchtungsschwankungen führen zu großflächigen und kurz andauernden Veränderungen der Bildhelligkeit. Zwar bedingt dies eine Fehlsegmentierung bei der Differenzbildung (Abb. 3-8-3 b/c). Durch die UND-Verknüpfung der Differenzbilder wird diese Störung jedoch entfernt (Abb. 3-8-3 d). Durch den farbbasierten Schattenfilter werden die Effekte des Schattenwurfes eliminiert.



**Abbildung 3-8-3:** Bewegungs-basierte Segmentierung bei Beleuchtungsschwankungen Beispielsegmentierung mittels CSC (sehr grobe Segmentierungsstufe).

### 3. MATCHING-PARADIGMA

Das Ergebnis des Nachverarbeitungsschrittes (morphologische Operationen und CSC-Schattenfilterung) ist ein Motion-Blob, der das bewegte Objekt enthält. Dieser Motion-Blob stellt die erste Merkmalsebene im aktuellen Bild dar (Abb. 3-8-3). Die auf der Grundlage dieser Blobregion durchgeführte CSC-Segmentierung führt zu einer Menge von Segmenten, die in Abhängigkeit von bestimmten Kompaktheits- und Größenattributen als gültig bzw. ungültig klassifiziert werden (Abb. 3-8-3-f/g). Die Gültigkeit wird benutzt, um unzuverlässige Regionen für die weitere Analyse auszuschließen und dadurch die Störsicherheit des merkmalsbasierten Matchings zu erhöhen (Abschnitt 3.4.2). Die auftretenden Beleuchtungsänderungen zwischen den Folgebildern beeinflussen die CSC-Segmentierung nur wenig, da in der Aufnahme ein ausreichender Kontrast vorhanden ist und der CSC-Algorithmus unter dieser Voraussetzung stabil segmentiert. Das Matching arbeitet in dieser, von der Anzahl der enthaltenen Objekte, „einfachen“ Sequenz robust.

Die Korrelation der Motion-Blobs beschränkt sich im Wesentlichen auf die Zuordnung einer einzigen Region und ist daher trivial. Die Zuordnung der CSC-Patch-Segmente arbeitet aufgrund der guten Segmentierung und der nicht auftretenden Verschmelzungen bzw. Teilungen der MB-Regionen sehr stabil (Abb. 3-8-4). Die Reduzierung der Segmentgröße bis zu einem gewissen Grad (ca.  $C_{SEG}=10$ ) führt zu einem dichten, dennoch genauen Verschiebungsvektorfeld (VVF) und dadurch haben die unvermeidlichen Vektorausreißer keinen Einfluss auf die Gesamteinschätzung.

Ein größerer CSC-Segmentierungsgrad führt dazu, dass viele, dafür aber kleine CSC-Segmente berechnet werden. Aus der größeren Anzahl von Segmenten folgt, dass viele Verschiebungsvektoren ermittelt werden können. Besonders im Falle kleiner MB-Regionen (Abb. 3-8-4b/c) verringert die höhere Informationsmenge des Verschiebungsvektorfeldes die Fehlerwahrscheinlichkeit bei der Bewegungsschätzung. Allerdings kann die Reduzierung der Segmentgröße nicht weiter fortgesetzt werden, da dies einer Korrelation im Pixelbereich nahe kommen würde, für die das Multi-Matching aufgrund ihrer Arbeitsweise nicht robust funktioniert (Abschnitt 3.5.2).

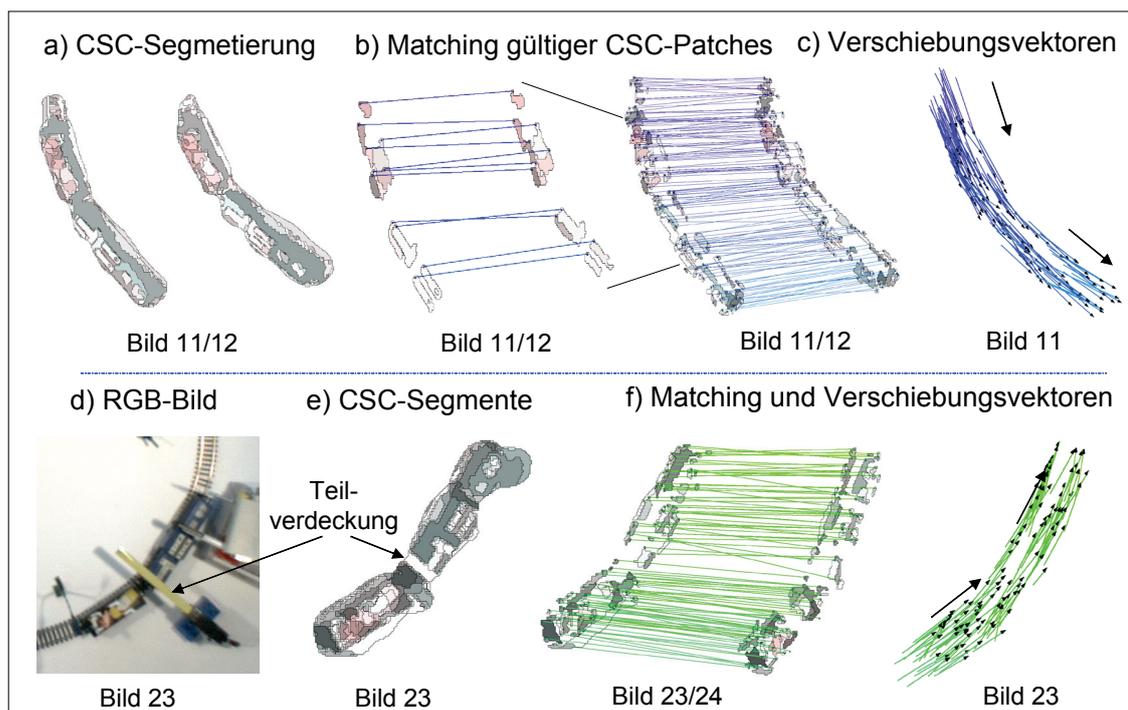
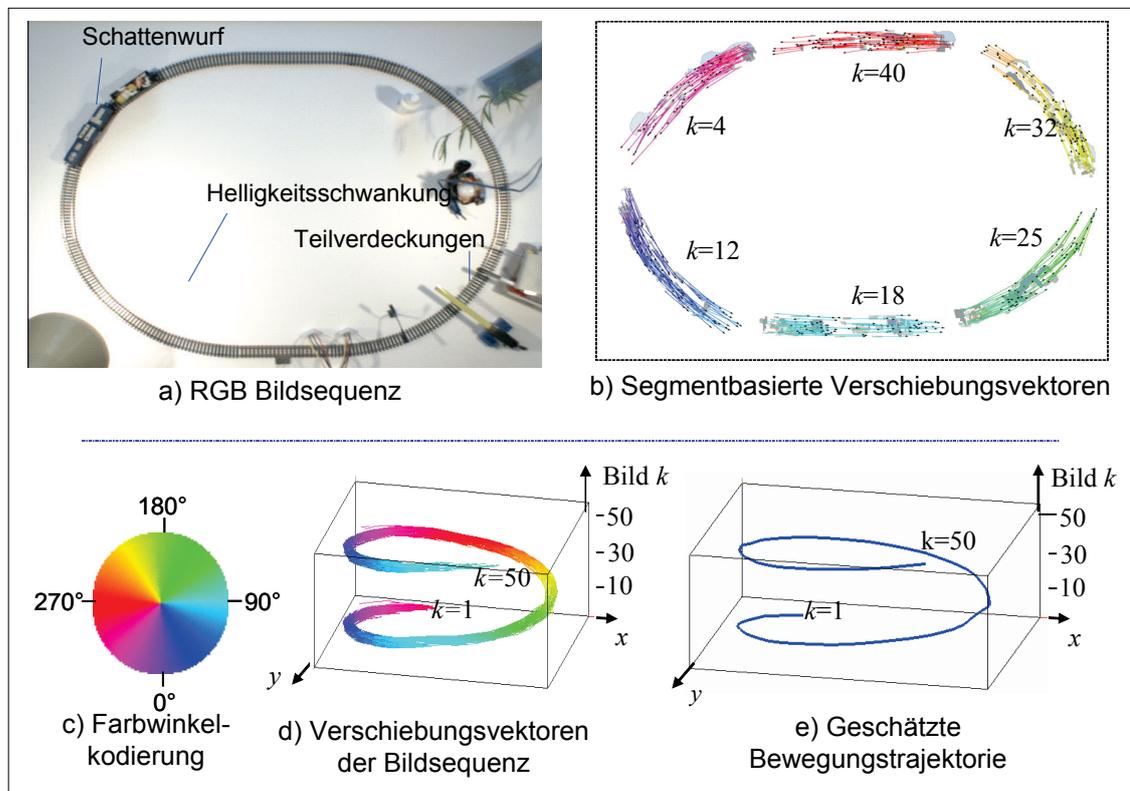


Abbildung 3-8-4: Beispiel für das Matching der CSC-Patch-Segmente bei Teilverdeckung.

### 3. MATCHING-PARADIGMA

Anhand der qualitativ hochwertigen Verschiebungsvektoren der CSC-Segmente (Abb. 3-8-4c/f) ist zu erkennen, dass das untersuchte Objekt im vorderen und hinteren Bereich verschiedene Bewegungsrichtungen aufweist. Dies wird durch die Rotationsbewegung des Objektes verursacht. Die Teilverdeckungen, die in dieser Szene präsent sind, beeinträchtigen die Korrespondenzermittlung nur minimal. Durch morphologische Operationen während der bewegungsbasierten Segmentierung wird der Verdeckungsbe-  
reich mitsegmentiert, d.h. zur Differenzbildregion hinzugefügt (Abb. 3-8-4 d/e/f). Das Multi-Matching der CSC-Segmente gemäß Gleichung (3-5-8) wird durch diese Störregionen nur wenig beeinflusst (Abb. 3-8-4f), d.h. an den Stellen der Objektverdeckung wird kein Verschiebungsvektor berechnet. Dies hat jedoch keinen Einfluss auf das Resultat der gesamten Auswertung (Abb. 3-8-5d/e). Die Menge sämtlicher Vektoren aller Bildpaare zeigt die gesamte gemessene Bewegungsinformation in der Bildsequenz. In der Darstellung der Vektoren wird die Bewegungsrichtung durch eine entsprechende Farbe abgebildet (Abb. 3-8-5c/d). Der fließende Farbübergang zeigt, dass die kreisförmige Bewegung des Objektes sehr genau erfasst wurde.



**Abbildung 3-8-5:** Farbkodierte Verschiebungsvektoren und die geschätzte Bewegungstrajektorie des Objektes der gesamten Bildsequenz unter dem Einfluss von gestörten Bildsituationen.

Eine automatisierte Aussage darüber, wie viele bewegte Objekte in der Szene enthalten sind und wie deren Bewegungstrajektorien definiert sind, ist dennoch nicht trivial und kann nur unter bestimmten Annahmen getroffen werden. Eine mögliche Annahme ist, dass ein Objekt durch eine Masse von benachbarten, ähnlichen VV ausgemacht werden kann. Ein Algorithmus der diese Aufgabe bewältigt, wurde im Rahmen der Forschungsarbeiten mittels des Algorithmus in Abschnitt 3.6 entwickelt und wird daher hier nicht näher beschrieben. Es werden jedoch einige vorläufige Ergebnisse dargestellt, um weitere Forschungen zu motivieren. Da in der untersuchten Sequenz eine große Menge

### 3. MATCHING-PARADIGMA

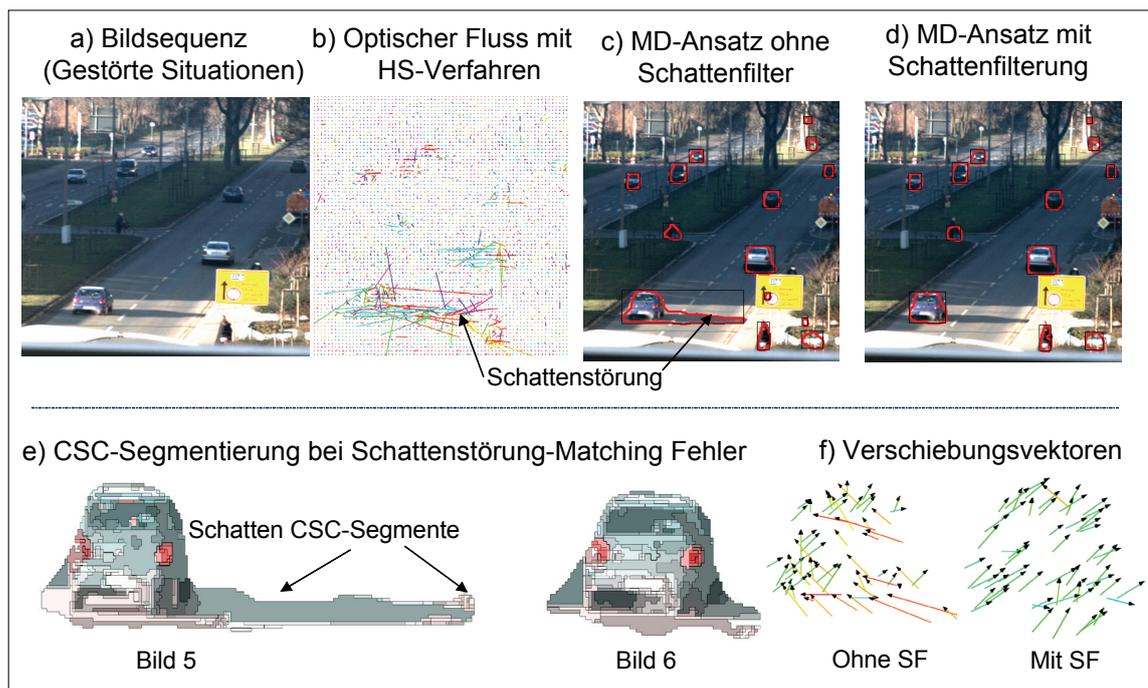
---

kontinuierlicher, dicht beieinander liegender, in dieselbe Richtung zeigender VV vorkommt, kann eine verhältnismäßig sichere Bewegungspfadsschätzung mittels der vordefinierten Attribute der Korrelationen von den Trajektorien gemäß Abschnitt 3.6 erfolgen (Abb. 3-8-5e).

Der Rechenaufwand beträgt in dieser Sequenz durchschnittlich 200ms je Bild (608x472 Pixel, PC 2 GHz). Damit ist das entwickelte Verfahren im Vergleich zur Berechnung des Optischen Flusses nach dem vorgestellten Verfahren in Kapitel 2 (mit  $\lambda=10$ , 10 Iterationen) ca. acht mal schneller. Zudem sind die VV des Optischen Flusses in dieser Sequenz ohne erhebliche Verbesserungsschritte unbrauchbar, um eine Bewegungsschätzung durchzuführen. Die nachweislich aufgrund der Funktionsweise des Matching-Paradigmas erzielten positiven Resultate lassen die Schlussfolgerung zu, dass in vergleichbaren Sequenzen ähnlich robuste Ergebnisse zu erwarten sind. Das Verhalten des Verfahrens in Situationen, in denen mehrere bewegte Objekte vorkommen, war Gegenstand der Untersuchung einer weiteren Bildfolge.

#### 3.8.2 Sequenz II – Geringer Straßenverkehr, Schattenwurf

Die nachfolgend beschriebene Sequenz ist eine Verkehrsaufnahme, die mehrere bewegte Objekte beinhaltet. Enthaltene Störungen sind starker Schattenwurf der Fahrzeuge im vorderen Bildbereich und die spontane Änderung der Helligkeit der Fahrzeuge bei Eintritt in die Schattenregion (Abb. 3-8-6a). Dazu kommen die geringen Grauwertgradienten der Fahrzeuge innerhalb der Schattenregion und die perspektivisch bedingte Änderung der Fahrzeuggröße. Diese bildspezifischen Störungen verursachen ein Korrespondenzproblem, das die Bestimmung der Verschiebungsvektoren erschwert. Die Schattenregion verursacht beispielsweise eine starke Abdunkelung von mehr als 40 %. Dies führt durch ungleichmäßige Intensitätsänderungen zu Deformationen des angewandten Ähnlichkeitsmaßes und somit zu einer falschen Einschätzung der Verschiebungsvektoren. Dazu kommt, dass wenige Grauwertgradienten in den dunklen kontrastarmen Bildbereichen vorkommen, wodurch kaum mit einem eindeutigen Extremwert aufgrund der Absenkung des Luminanzdifferenz zwischen Vordergrundobjekte und Hintergrundobjekte zu rechnen ist. Weiterhin löst der Schattenwurf des Fahrzeugs eine Fehlsegmentierung durch den vorgeschlagenen MD-Ansatz (Abb. 3-8-6c) aus, falls kein Vorverarbeitungsschritt zur Schattensegmentierung angewandt wird (Abschnitt 3.3).



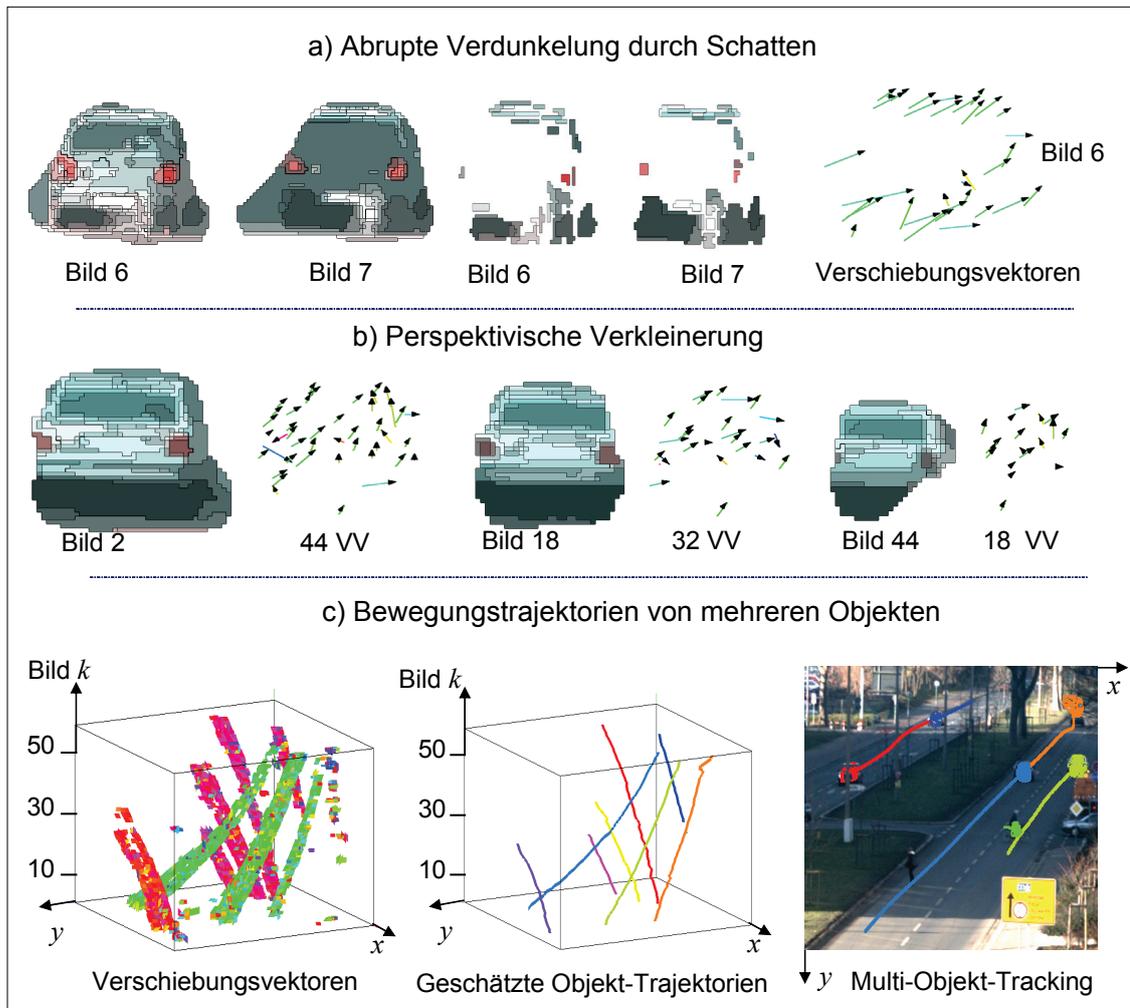
**Abbildung 3-8-6:** a) Schattenwurf bedingt fehlerhafte Verschiebungsvektoren bei Optischem Fluss bzw. b) Fehlsegmentierung und c) mit Schattenfilterung. Verschiebungsvektoren a) ohne und b) mit Schattenfilter (SF).

Wird die Schattenregion nicht entfernt, kommt es zu Fehlzuordnungen beim Matching der CSC-Segmente und somit zur Verfälschung der Verschiebungsvektoren (Abb. 3-8-6e/f). Weiterhin wird die Korrespondenzermittlung durch die abrupte, starke Objektverdunklung bei dessen Eintritt in die Schattenregion erschwert (Abb. 3-8-7a). Eine extreme Verdunklung führt zu einer Verschlechterung der Segmentierungsergebnisse.

### 3. MATCHING-PARADIGMA

Bedingt durch die geringe Helligkeit werden nur wenige, vorwiegend große Segmente berechnet. Diese entziehen die Grundlage für ein stabiles Matching, welches eine Vielzahl von VV liefert.

Im Beispiel in Abb. 3-8-6/7 wurde dennoch ein ausreichendes Ergebnis erzielt, da das segmentierte Fahrzeug zumindest an einigen Stellen genügende Helligkeit und damit auch Segmente aufweist. Die auftretende, perspektivisch bedingte Verkleinerung entfernter Gegenstände ist generell kein einfaches Problem. Einfache Blockmatching Verfahren mit einem festen Referenzbereich haben zum Beispiel Schwierigkeiten bei der Korrespondenzermittlung, da dieser bei der Verkleinerung des Objektes keine genaue Grundlage für das Matching darstellt. Durch die Objektverkleinerung in einem Referenzbereich steigt auch die Wahrscheinlichkeit für eine Koexistenz von Objekten. Das merkmalsorientierte Matching liefert hingegen auch in dieser Sequenz ein dichtes, relatives geordnetes Verschiebungsvektorfeld. Die perspektivische Objektverkleinerung führt jedoch zu einer verminderten Segmentanzahl und damit zu einer Reduzierung der VV (Abb. 3-8-7). Bedingt durch die Verkleinerung sich entfernender Objekte verringert sich auch der Betrag der entsprechenden VV. Dadurch wird eine mögliche Schätzung der Geschwindigkeit des bewegten Objektes erschwert.



**Abbildung 3-8-7:** Starke Verdunkelung führt zu Beeinträchtigung der Segmentierung(a). Dies führt zu weniger Matching-Partner und somit weniger Verschiebungsvektoren. B) Perspektivische Verkleinerung führt zur Segmentabnahme, Verschiebungsvektoren zur Verkürzung. C) Ergebnisse des Matchings.

Die Menge aller Verschiebungsvektoren kann auch in dieser Sequenz erfolgreich genutzt werden, um eine Schätzung wahrscheinlicher Objekt-Trajektorien durchzuführen (Abb. 3-8-7c). Diese Aufgabe ist im vorliegenden Fall recht einfach, da die Verschiebungsvektoren in gut abgegrenzten Häufungsbereichen vorliegen (Abb. 3-8-7c).

Der Rechenaufwand beträgt in dieser Sequenz durchschnittlich 190ms je Bild (360x368 Pixel, PC 2 GHz). Die Resultate zeigen, dass das entwickelte Matching prinzipiell auch in Situationen mit ausgeprägtem *Schattenwurf* und starken *Helligkeitsvariationen* robust arbeitet und die Bewegungstrajektorie die tatsächliche Bewegung der Objekte beschreibt. Es hat sich jedoch gezeigt, dass eine starke Reduzierung der Bildhelligkeit zur Verschlechterung der Matching-Bedingungen führt, was sich auf die schwach ausgeprägten Farbsegmentierungsergebnisse zurückführen lässt. Dieses Problem kann nicht auf der Softwareebene gelöst werden, da der Informationsgehalt der Eingangsdaten im Falle sehr dunkler Bilder bzw. Bildbereiche zu gering ist. Eine Verbesserung könnte hier durch den Einsatz von Kameras mit höherer Helligkeitsdynamik erreicht werden.

#### 3.8.3 Sequenz III – Dichter Straßenverkehr

Die Leistungsfähigkeit des vorgeschlagenen Matching-Paradigmas zur Bestimmung von Bewegungstrajektorien in Situationen, in denen eine größere Anzahl sich *dicht nebeneinander bewogender* Fahrzeuge vorkommt, war Untersuchungsgegenstand weiterer Bildfolgen (Abb. 3-8-8a). Enthaltene Störungen sind dichter Straßenverkehr mit leichten Teilverdeckungen der PKWs, Reflexionen auf der Strasse und perspektivisch bedingte Änderung der Objektgröße. Dazu kommen die Verschmelzung und das Zerfallen von Objekten aufgrund der auftretenden Störungen (Stau und Reflexionen).

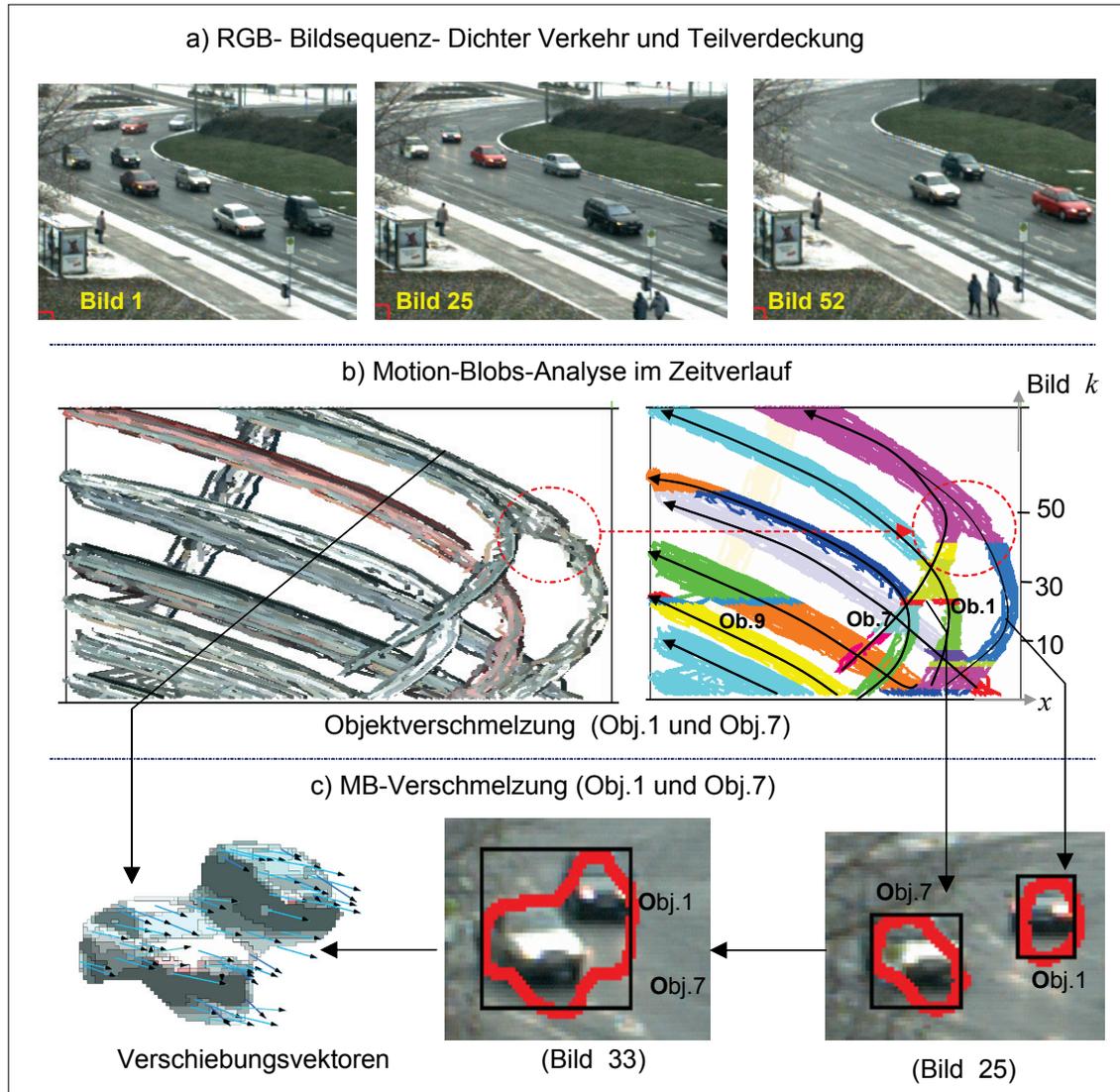
Die große Anzahl dicht nebeneinander fahrender PKWs bewirkt bei der automatischen Objekterfassung und deren Nachbearbeitung eine Verschmelzung von Objektregionen (Abb. 3-8-8b/c). Hierdurch werden mehrere Objekte aufgrund der Verschmelzung auf eine MB-Region abgebildet. Auch wäre zum anderen Zeitpunkt ein Split, also ein Auftrennen des Objekts in zwei Objekte möglich. Dies wird durch unterschiedliche Bewegungsvektoren und Attribute erkannt, die von dem gleichen Objekt ausgehen.

Da ein MB bei Verschmelzung mehrere Objekte beinhalten kann, wird hier das Korrespondenzproblem in der MB-Korrelationsebene nicht gelöst. Hier ändern sich die Attribute des jeweiligen Motion-Blobs im Zeitverlauf aufgrund der gestörten Bildsituationen abrupt (s. Abb. 3-8-8b). In der Abbildung ist die Verschmelzung von Objekt 1 und Objekt 7 zum Zeitpunkt  $k=33$  zu erkennen. Die Zerlegung des Motion-Blobs in Segmente und die Berechnung von jeweiligen Verschiebungsvektoren sind in (Abb. 3-8-8c) dargestellt. Hier hat sich gezeigt, dass die Analyse von VV zwischen zwei aufeinander folgenden Bildern in den meisten Fällen keine ausreichende Information liefert, um ein einzelnes Objekt durch abgeleitete Attribute aus VV-Verteilungen zu lokalisieren und voneinander zu trennen. Die Histogramme über Winkel und Länge der einzelnen Vektoren liefern nicht genug ausgeprägte Häufungen bei der Betrachtung von einen einzelnen Zeitpunkt in realen Bildszene, um Rückschlüsse auf einzelne Objektbewegungen zu ziehen (Abb. 3-8-9b). Häufungen sind jedoch auch nur zu erwarten, wenn Objekte eine sehr unterschiedliche Bewegungsrichtung aufweisen und einen Mindestabstand voneinander haben, was leider in den realen Bildsequenzen nicht zu gewährleisten ist.

### 3. MATCHING-PARADIGMA

Bedingt durch natürliche Variationen bei der Segmentierung weisen die VV stets leichte Variationen auf. Bewegen sich zwei Objekte annähernd in die gleiche Richtung, ist daher keine klare Unterscheidung auf der Grundlage der Vektoren möglich.

Es besteht jedoch die Möglichkeit, Objekte, welche sich im Verlauf einer Bildfolge angenähert haben, so dass sie zu einer Region verschmolzen wurden, wieder zu trennen. Dies kann durch die zusätzliche Einbeziehung der Segmentvorgeschichte und somit auch die aktiven Bewegungstrajektorien (s. Abschnitt 3.6) realisiert werden.

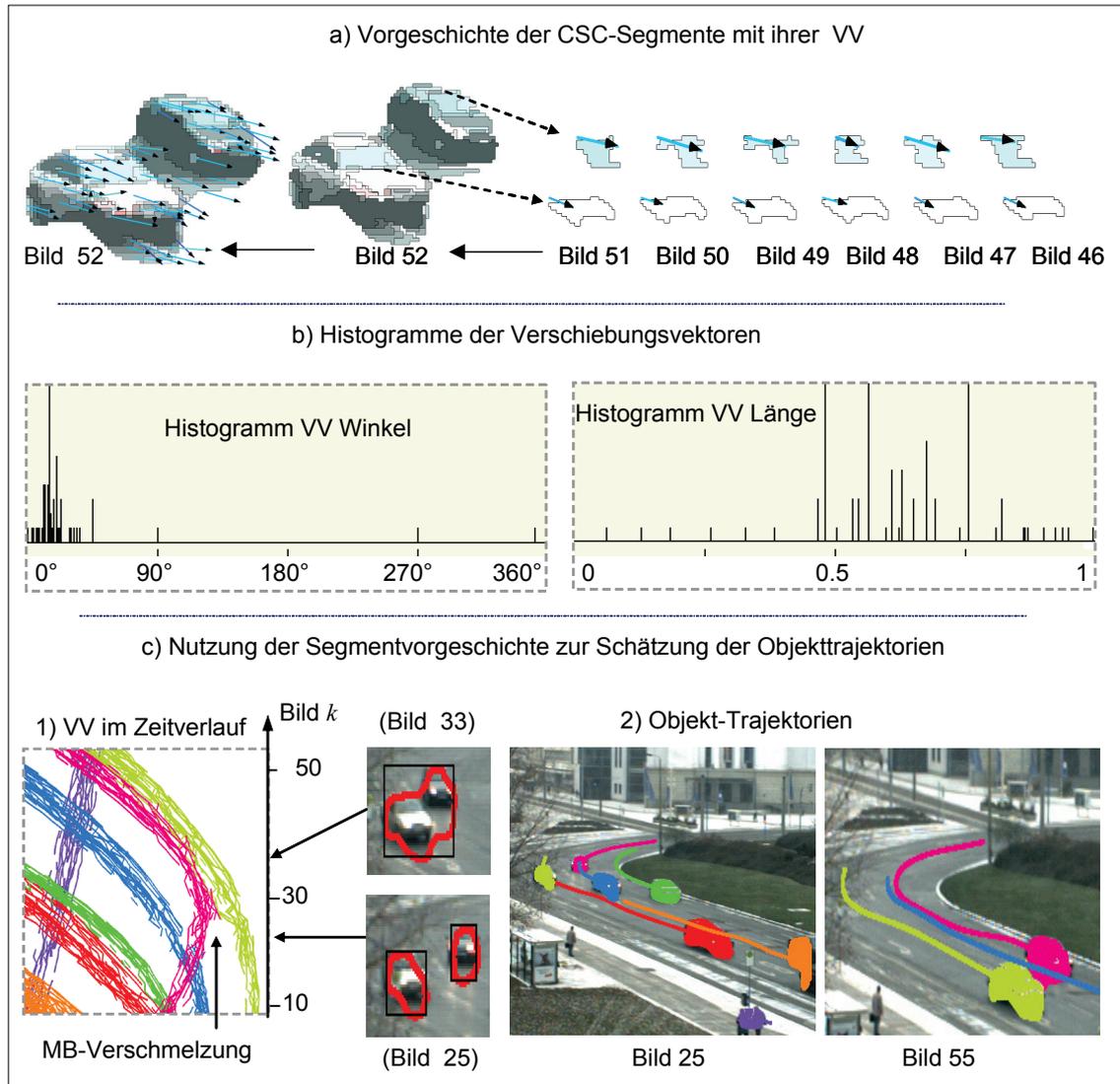


**Abbildung 3-8-8:** Straßenverkehrsaufnahme b) Verschmelzung dicht fahrender bzw. sich verdeckender PKWs, c) Ausschnitt aus der Vorgeschichte zweier CSC-Segmente.

Die Vorgeschichte ergibt sich aus der Matchinginformation jedes CSC-Segmentes und kann je nach Situation, d.h. abhängig vom Grad der Variationen bei der Segmentierung, unterschiedlich lang sein (Abb. 3-8-9a). Die Variationen, die am häufigsten zu einem Abbruch in der Vorgeschichte eines Segmentes führen, entstehen durch Objektrotation, Verdeckung oder starke Beleuchtungsänderungen. Aus den zeitlichen Segmentzusammenhängen lassen sich im Falle einer Verschmelzung von MB-Regionen, die darin enthaltenen CSC-Segmente weiterverfolgen (Abb. 3-8-9a/b/c). Dies bedeutet, dass ein Objekt, von dem angenommen wird, dass es durch eine bestimmte Menge von

### 3. MATCHING-PARADIGMA

Bildmerkmalen, d.h. CSC-Segmenten bzw. deren VV repräsentiert wird, auch in Situationen, in denen diese Merkmale stark gehäuft auftreten, verfolgt werden kann. Die Segmentvorgeschichte stellt daher eine nützliche Information bei der Auswertung des VVF dar. Im Falle von größeren oder totalen Objektverdeckungen, die eine Unterbrechung in der Vorgeschichte der CSC-Segmente bewirken, muss jedoch ein geeigneter Schätzalgorithmus (z.B. mittels Kalman-Filter oder Partikelfilter [Al-Hamadi 03f, 07a, 08b]) entwickelt werden, um eine Objektverfolgung zu realisieren. Die Entwicklung eines solchen Verfahrens stellt eine der Hauptmotivationen für weitergehende Untersuchungen mittels Partikelfiltern in Kapitel 5 dar.



**Abbildung 3-8-9:** Histogramme der Verschiebungsvektoren und Nutzung der Segmentvorgeschichte zur Schätzung der Objekt-Trajektorien.

Eine schwer handhabbare Störung, die in der untersuchten Sequenz präsent ist, sind Lichtreflexionen auf dem nassen Asphalt, welche durch PKW Scheinwerfer hervorgerufen werden. In Bildanalyse-Systemen, die Außenaufnahmen auswerten, sind derartige nicht-kooperative Störungen von großer Relevanz. Analog zum Schattenwurf erfasst die Differenzbildtechnik zum großen Teil auch solche Regionen, die Lichtreflexionen beinhalten. In der vorliegenden Sequenz sind diese Störungen jedoch verhältnismäßig schwach ausgeprägt, daher werden das Matching und damit das berechnete Verschie-

### 3. MATCHING-PARADIGMA

bungsvektorfeld nur unwesentlich beeinflusst. Allerdings sollte für eine sichere Anwendung des differenzbildbasierten Matching-Verfahrens zur Vermeidung des Problems der Lichtreflexionen ein gesonderter Arbeitsschritt bei der Ermittlung der bewegten Regionen eingebracht werden. Ein solcher Schritt könnte zum Beispiel Formveränderungen der Differenzbildregionen feststellen, die im Zusammenhang mit erhöhten Bildhelligkeitswerten auftreten. Die geringe Größe entfernter Objekte stellt eine weitere Schwierigkeit der untersuchten Bildsequenz dar. Diese führt zu einer Verringerung der Anzahl an Segmentprimitiven, wodurch es letztlich zu weniger Genauigkeit bzw. Sicherheit bei der Korrespondenzermittlung kommt. Um dieses Problem zu vermeiden, bietet es sich an, während der Bildaufnahme die zu untersuchende Region möglichst vergrößert aufzunehmen.

Die Resultate dieser Bildsequenz zeigen, dass das entwickelte Matching-Verfahren auch in gestörten bildspezifischen Situationen arbeitet und ein dichtes VVF liefert, in denen mehrere bewegte Objekte vorhanden sind, welche eine geringe Distanz zueinander haben. Natürlich hat das Verfahren auch Grenzen, wenn eine bestimmte Anzahl von Objekten das Bild füllt. Der Berechnungsaufwand beträgt in dieser Sequenz im Mittel 70ms je Bild (312x280 Pixel, PC 2 GHz).

#### 3.9 Auswirkung verschiedener Segmentierungsgrade

Im Abschnitt 3.1 ist erwähnt, dass die Verbesserung der Matchinggüte von der effizienten Erzeugung von qualitativ hochwertigen Farbsegmentprimitiven abhängig ist. Die qualitativ hochwertigen Segmente sollen eine möglichst gleichmäßig runde, wenig zerklüftete Form aufweisen. Eine solche Eigenschaft kann genutzt werden, um stark formentartete Segmentprimitive zu eliminieren, welche meist nicht für ein Matching geeignet sind und daher möglichst selten segmentiert werden sollten. Dies kann dadurch erreicht werden, dass bei der Anwendung ein bestimmter Segmentierungsgrad  $C_{SEG}$  eingestellt wird. Der Segmentierungsgrad  $C_{SEG}$  bestimmt die CSC-Segmentgröße während der Erstellung der zweiten Merkmalsebene und beeinflusst daher die Ergebnisse des Matchings in grundlegender Weise. In Abbildung 3-9-1 werden die Auswirkungen verschiedener Segmentierungsgrade beispielhaft demonstriert.

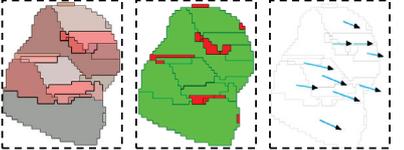
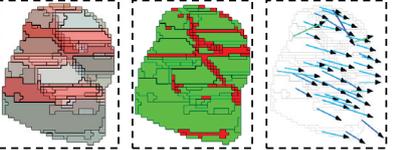
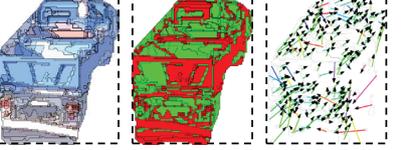
<p>a)</p> <p>36x46 Pixel</p> 	<p><math>C_{SEG} = 45 \rightarrow 23</math> CSC-Patches, 10 VV</p> 	<p><math>C_{SEG} = 10 \rightarrow 153</math> CSC-Patches, 34 VV</p> 
<p>b)</p> <p>110x114 Pixel</p> 	<p><math>C_{SEG} = 45 \rightarrow 203</math> CSC-Patches, 20 VV</p> 	<p><math>C_{SEG} = 10 \rightarrow 804</math> CSC-Patches, 232 VV</p> 
<p>RGB</p>	<p>CSC-Patches Gültigkeit VV</p>	<p>CSC-Patches Gültigkeit VV</p>

Abbildung 3-9-1: Beispiele verschiedene Segmentierungsgrade  $C_{SEG}$

Ein geringer CSC-Segmentierungsgrad (z.B.  $C_{SEG}=45$ ) führt dazu, dass wenige, dafür aber große CSC-Segmente berechnet werden. Aus der geringen Anzahl von Segmenten folgt, dass nur wenige Verschiebungsvektoren ermittelt werden können. Besonders im Falle kleiner MB-Regionen (Abb. 3-9-1a) erhöht die geringe Informationsmenge des Verschiebung die Fehlerwahrscheinlichkeit bei der Bewegungsschätzung. Auch bei größeren MB-Regionen führt ein geringer Segmentierungsgrad nur zu einem dünnen Verschiebungsvektorfeld (Abb. 3-9-1 a/b,  $C_{SEG}=45$ ). Es hat sich in den Untersuchungen gezeigt, dass eine Reduzierung der Segmentgröße bis zu einem gewissen Grad (ca.  $C_{SEG}=10$ ) zu einem dichten, dennoch genauen VVF führt. Dabei ist es unerheblich wie groß die zu segmentierende Region ist (Abb. 3-9-1 a/b,  $C_{SEG}=10$ ). Die Reduzierung der Segmentgröße kann jedoch nicht weiter fortgesetzt werden, da diese einer Korrelation im Pixelbereich nahe kommen würde, für die die Matching-Ähnlichkeitsmaße aufgrund ihrer Arbeitsweise nicht funktionieren.

#### 3.10 Diskussion, Schlussfolgerung und Grenzen

Die Untersuchungen mit realen Testsequenzen haben gezeigt, dass die angestrebten Zielstellungen der Forschungsarbeit mit dem vorgeschlagenen Matching-Paradigma grundsätzlich erreicht wurden und sich wie folgt darstellen lassen.

- Der angewandte MD-Ansatz einschließlich Nachbearbeitung (morphologischen Operationen und Schattenfilter) hat es ermöglicht, auf geeignete Weise die Bilddaten so zu reduzieren, dass ausschließlich bewegte Regionen in der Bewegungsanalyse betrachtet werden. Es konnte durch die Kombination zweier Differenzbilder eine Verbesserung gegenüber der einfachen Differenzbildtechnik erzielt werden, welche nicht in der Lage ist genaue Konturen von bewegten Objekten zu ermitteln. Weiterhin wurde mit der Verwendung des CSC-Farbsegmentierungsalgorithmus ein geeigneter Weg gefunden, um im Anschluss an die Differenzbildtechnik erstens eine Schattenfilterung zu realisieren und zweitens effizient flächenhafte Bildprimitive zu erzeugen, welche zweckmäßige Eigenschaften für ein merkmalsbasiertes Matching aufweisen. Insbesondere wurde hierzu eine Untersuchung durchgeführt, bei der Segmentprimitive dreier Farbsegmentierungsalgorithmen im Hinblick auf Eigenschaften, die für das Matching relevant sind, verglichen wurden.
- Die Trennung der bewegten Bilddaten in Motion-Blobs und CSC-Farbsegmente hat es erfordert, zwei merkmalsorientierte Ähnlichkeitsmaße zu definieren, um einerseits ein m:n Matching der Motion-Blobs, andererseits ein 1:1 Matching der CSC-Patches durchzuführen. Das Ähnlichkeitsmaß der Motion-Blobs beruht auf der topologischen Nachbarschaft und den Flächenmaßen, das zu einer globalen Zuordnung der Regionen dient. Die genau detaillierte Zuordnung erfolgt in der zweiten CSC-Ebene mittels der extrahierten Attribute. Das Ähnlichkeitsmaß der wesentlich kleineren und zahlenmäßig (im Mittel 10-1000fach) häufiger vorkommenden CSC-Patch Segmente ist dementsprechend detaillierter und setzt sich aus vier unterschiedlich gewichteten Einzelähnlichkeitsmaßen zusammen. Die verwendeten Gewichte wurden auf der Basis der Aussagekraft des jeweiligen Kriteriums festgelegt. Die Aussagekraft wurde dabei durch die Funktionsweise begründet (Abschnitt 3.5.2). Weiterhin wurde, wenn möglich, die Verwendung der Ähnlichkeitsmaße von der Merkmalszuverlässigkeit abhängig gemacht (siehe SA- Ähnlichkeit der relativen CSC-Patch Position, Abschnitt 3.5.2).

- Die Optimierung der Gewichte hat sich als sinnvoller Schritt herausgestellt, wodurch signifikante Verbesserungen im Bereich des eindeutigen sowohl uneindeutigen Matchings erreicht wurden. Allerdings sollten weitere Untersuchungen zur Verallgemeinerung der Gewichtsfaktoren unternommen werden. Dazu wurde durch die schrittweise Erweiterung des Matching gezeigt, wie sich die Einzelmaße auf das Matching auswirken. Diese führte dazu, dass die Einbeziehung von weiteren Merkmalen nicht signifikant zur Verbesserung der Störsicherheit beiträgt. Durch etwas erhöhten Rechenaufwand in der Korrelationsberechnung können so starke fehlerbehaftete falsche Zuordnungen verringert werden. Als Ausblick auf weitere Optimierungsschritte des CSC-Matching würde sich ein künstlich neuronales Netz eignen, um die Gewichte analog zu [Al-Hamadi 01a, 03f, 03g] zu berechnen. Ein weiterer Ansatz wäre die Ausführung des CSC-Matchings über eine Fuzzy-Logik. Demnach gäbe es keine eindeutige Zuordnung eines Segmentes zu einem anderen mehr. Die Zugehörigkeit würde durch eine Fuzzy-Wahrscheinlichkeit gegeben werden. Die Übertragungsfunktion müsste dann mit Hilfe von Expertenwissen über Fuzzy-Mengen modelliert werden. Durch diese Erweiterungen könnte die Robustheit der Objektverfolgung erheblich profitieren (vgl. Kapitel 5).
- Die Ergebnisse der durchgeführten Bildfolgenuntersuchungen bestätigten die Funktionsfähigkeit und Korrektheit des merkmalsorientierten Ähnlichkeitsmaßes sowohl in der grundlegenden Ebene der MB-Regionen als auch in der Ebene der CSC-Patch-Segmente. Die Betrachtung der VV, insbesondere der VVF hat die Richtigkeit des Matchings deutlich aufgezeigt. Es war sowohl in künstlichen, ungestörten Bildfolgen als auch in durch verschiedene Störungen betroffenen realen Bildfolgen, möglich dichte VVF zu generieren.
- Bei der Bestimmung des Verschiebungsvektorfeldes konnte eine hohe Robustheit bei ausgeprägten Helligkeitsschwankungen durch Beleuchtungsänderungen bzw. Schattenwurf sowie bei Objektrotation, Deformation und Teilverdeckung erreicht werden. Die Robustheit resultiert zum einen aus der hierarchischen Merkmalsstruktur, zum anderen aus der damit verbundenen Reduzierung der Bildmerkmale. Weiterhin ist die Robustheit auf das Vorhandensein von Farbinformation zurückzuführen, die die Grundlage für die Segmentierung der CSC-Patches darstellt, aber auch als ein zentrales Ähnlichkeitsmaß für das Matching fungiert. Ebenfalls trägt die Anwendung des auf Farbbilddaten arbeitenden Schattenfilters zur Erhöhung der Stabilität bei der Bestimmung der Bewegungsparameter bei (Abschnitt 3.2.3).
- Die in den untersuchten Bildfolgen ermittelten VVF konnten genutzt werden, um Aussagen über den Bewegungspfad einzelner Objekte zu treffen. Es wurde demonstriert, dass die Bewegung einzelner Objekte durch Häufungen von VV im räumlich-zeitlichen Kontext repräsentiert wird. Es hat sich jedoch auch gezeigt, dass Objekte, beispielsweise Fahrzeuge, die sich sehr dicht nebeneinander bewegen, aufgrund der alleinigen Betrachtung einer kleinen Anzahl von VV nicht detektiert werden können. Für eine solche Aufgabe ist die Auswertung eines größeren räumlich-zeitlichen Zusammenhangs notwendig, bei dem z.B. die Vorgeschichte einzelner Farbsegmente hilfreich sein kann. In den experimentellen Untersuchungen haben sich jedoch auch Einschränkungen des Verfahrens in seiner jetzigen Form gezeigt, die Erweiterungen bzw. Veränderungen erfordern.

#### 3.10.1 Besonderheit des vorgeschlagenen Matching-Paradigmas

- Ein wesentlicher Vorteil des vorgeschlagenen Matching-Paradigmas gegenüber herkömmlichen Verfahren besteht darin, dass keine kalibrierte Kamera, kein Straßenmodell und keine Kenntnisse über das Aussehen der zu detektierten Objekte erforderlich sind. Darüber hinaus lassen sich Objekte mit signifikanter Relativbewegung zur Kamera zuverlässig detektieren. Als weiterer Vorteil des Verfahrens gilt, dass in der Bewegungsbestimmung keine Parameter einzustellen sind. Hinzu kommt, dass das Verfahren aufgrund der erfassten kleinen Segmente, die auch eine Bewegungsbestimmung ermöglichen, robust gegenüber Teilverdeckung und Verschmelzung oder Zerfallen von Objekten ist.
- Weitere entscheidende Vorteile liegen in der robusten Detektion und Analyse von formveränderlichen Objekten sowie die Erfassung von Schattenwurf. Darüber hinaus liefern die Objekthypothesen eine gute Grundlage zur Beschreibung der Objekte. Eine Klassifikation des Fahrzeugtyps (PKW, LKW, Motorrad) bei vorbeifahrenden Fahrzeugen kann auf diese Grundlagen implementiert (Support-Vector-Machine oder Multi-Layer-Perzeptron) werden.
- Weitere Vorteile wiederum liegen im geringen Rechenaufwand und in der Komplexität, welche ein echtzeitfähiges System ermöglichen können (auf einem PC mit einem Ein-Prozessor System konnte hier eine Bildwiederholrate von 6-8 Bilder pro Sekunde erzielt werden).

#### 3.10.2 Grenzen des Verfahrens

Die Ergebnisse des Matchings können an Genauigkeit verlieren bzw. falsch sein, wenn es zu komplizierten Aufnahmesituationen kommt. Es hat sich gezeigt, dass eine sehr homogene Textur großer Objektflächen, Fehler bei der Erfassung von bewegten Regionen verursachen kann. Die Ursache dieses Effekts liegt in dem MD-Ansatz. Die Störung entsteht, wenn die Differenz zweier ähnlicher Intensitätswerte unter eine festgelegte Mindestschwelle fällt, wodurch Löcher im Differenzbild entstehen (Abb. 3-10-1). Dies kann möglicherweise zu gestörten VV führen, zumindest aber zu einem weniger dichten VVF.

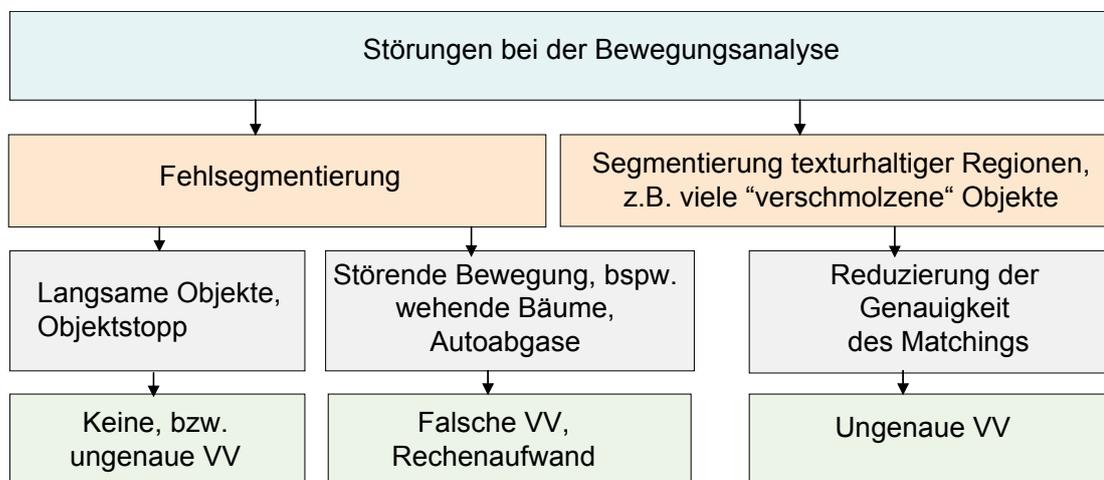


Abbildung 3-10-1: Grenzen des vorgeschlagenen Paradigmas

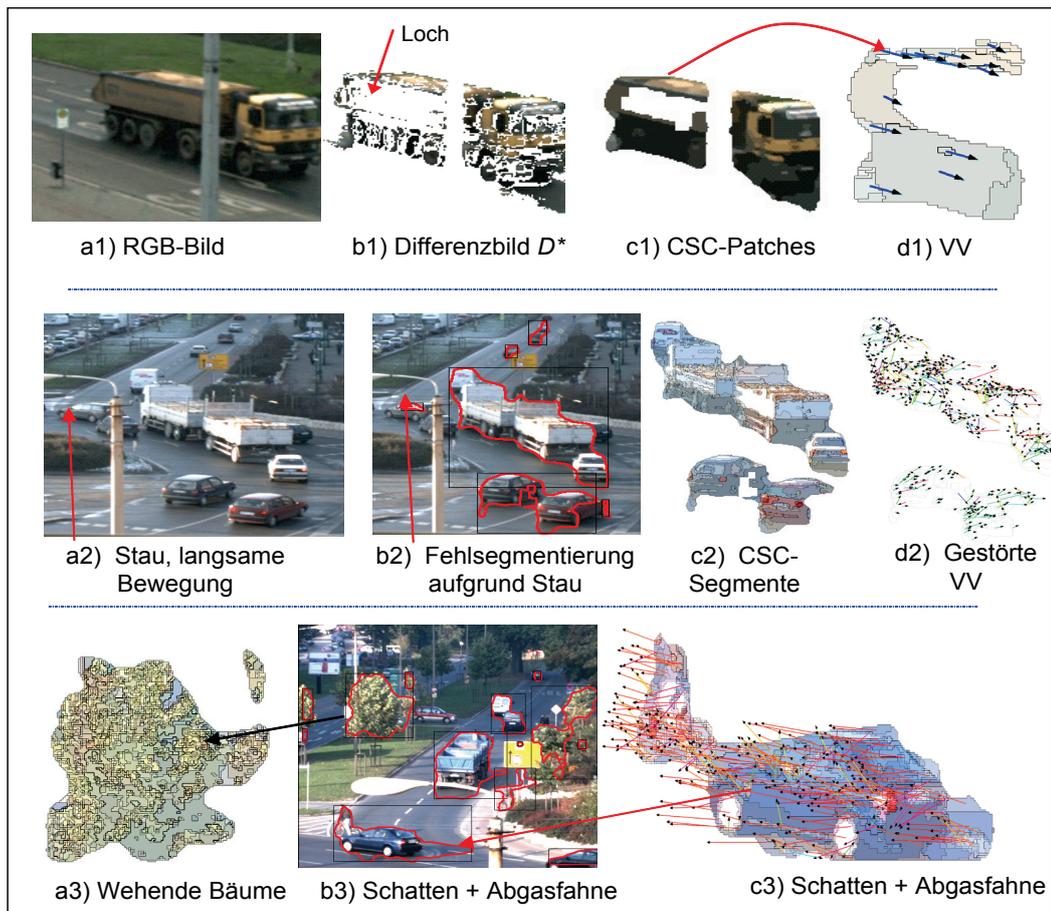
### 3. MATCHING-PARADIGMA

Das Problem der homogenen Textur besteht jedoch nicht nur im Falle der Differenzbildtechnik. Bei Korrelationsverfahren kann dies zum Blendenproblem führen und auch eine Berechnung des Optischen Flusses wäre an den betroffenen Stellen kritisch. Ein weiterer Störeffekt, der beobachtet wurde, ist die Segmentierung redundanter Bewegungsinformation, welche zur Beeinträchtigung der Gesamtleistung des Analysesystems führen kann. Es hat sich in Untersuchungen gezeigt, dass z.B. die segmentierte Bewegung wehender Bäume sehr störend sein kann (Abb. 3-10-2).

Zur Beseitigung einer derartigen „Übersegmentierung“ müssen gesonderte Strategien entwickelt werden, wie etwa eine anwendungsspezifische Auswahl bestimmter Bildbereiche. Eine generellere Möglichkeit wäre die Ähnlichkeit der VV im zeitlichen Zusammenhang auszuwerten und die Bildbereiche zu deaktivieren, die über längere Zeit eine chaotische Bewegung aufzuweisen scheinen, wie im Falle der Blätter- bzw. Baumbewegung. Eine weitere Alternative ist die Definition von erhabenen Objekten.

Die Genauigkeit des Matchings kann abnehmen, wenn extrem große Bildbereiche bei der Bewegungssegmentierung ermittelt werden, die viele „verschmolzene“ Einzelobjekte mit eigener Bewegungsrichtung beinhalten (siehe Abb. 3-10-3).

Generell steigt mit der Zunahme an CSC-Segmenten die Gefahr von Fehlzuordnungen. Die Effizienz und Genauigkeit des Matching-Paradigmas, das darauf beruht, eine möglichst stark reduzierte Menge an Bildmerkmalen zu korrelieren, werden hierdurch gefährdet. Daher ist das Matching in seiner hier vorgestellten Variante weniger für Bildsequenzen geeignet, in denen bildfüllende Bewegungen vorkommen.



**Abbildung 3-10-2:** Löcher im Differenzbild. Die großen MB-Regionen führen zur gestörten Verschiebungsvektoren, a3) Störungen bei der Bewegungssegmentierung.

#### 3.11 Zusammenfassung und Ausblick

Die Aufgabe dieses Kapitels bestand in der Entwicklung eines neuen merkmalsorientierten Multi-Matching-Paradigmas, welches zur Lösung des Korrespondenzproblems in der Bildfolgenanalyse eingesetzt werden kann. Die allgemeine Lösung des Korrespondenzproblems besteht in der Zuordnung von Bildprimitiven, wie etwa Punkten, Kanten oder Regionen. Um eine hohe Robustheit und Effizienz zu erreichen, sollten insbesondere Farbinformationen sowie eine differenzbildbasierte Bewegungssegmentierung Anwendung finden. Diese Kombination stellt ein Konzept dar, welches so bisher nicht in der Literatur beschrieben wurde.

Um eine möglichst genaue Segmentierung bewegter Objekte zu erreichen, wurde ein MD-Ansatz verwendet, bei der jeweils zwei Einzeldifferenzbilder kombiniert werden. Im Zusammenhang mit einer umfangreichen Nachbearbeitung konnten genaue Segmentierungen erzielt werden. Da durch die Bewegungssegmentierung auch bewegte Schattenregionen erfasst werden, welche die Bewegungsanalyse des gesamten Bildes stören können, wurde weiterhin ein Algorithmus zur nachträglichen Segmentierung und Entfernung von Schatten entworfen.

Die als Motion-Blobs vorgestellten Regionen, welche durch den Vorgang der Bewegungssegmentierung ermittelt werden, können recht groß sein und die Abbildungen mehrerer Objekte beinhalten. Motion-Blobs besitzen daher generell mehr als nur einen Verschiebungsvektor. Um diesem Sachverhalt gerecht zu werden, werden MB-Regionen in kleinere Komponenten unterteilt, was durch den CSC-Farbsegmentierungsalgorithmus realisiert wird. Aus der Menge der ermittelten Farbsegmente und der übergeordneten MB-Regionen gehen zwei Merkmalsebenen hervor, welche genutzt werden, um in der Kombination das Korrespondenzproblem zu lösen.

Anders als in bisherigen merkmalsorientierten Verfahren kann durch die Betrachtung zweier Merkmalsebenen die hohe Anzahl an Bildmerkmalen sehr einfach, schnell und exakt reduziert werden. Dies führt zu einer erheblichen Beschleunigung des gesamten Verfahrens und zur Verbesserung der Genauigkeit beim Matching der Farbsegmente. Das Matching wird in separaten Durchgängen realisiert und basiert jeweils auf einer Reihe von Ähnlichkeitsmaße. Die durch den Matching-Prozess ermittelten Korrespondenzen zwischen den Farbsegmenten führen zu Verschiebungsvektoren, welche prinzipiell genutzt werden können, um Aussagen über Objektbewegungen zu treffen. Es wurde anhand realer, gestörter Beispiele demonstriert, dass die Bewegung einzelner Objekte durch Häufungen von Verschiebungsvektoren im räumlich-zeitlichen Kontext repräsentiert wird. Bei der Berechnung der Vektoren konnte eine hohe Robustheit gegenüber in realen Situationen häufig kombiniert auftretenden Problemen, wie Beleuchtungsschwankungen, Teilverdeckungen, Schattenwurf, Objektdeformation durch perspektivisch bedingte Größenveränderung und Rotation, erreicht werden.

In den Untersuchungen realer Bildfolgen hat sich gezeigt, dass das entwickelte merkmalsorientierte Matching-Konzept eine echtzeitfähige Bewegungsanalyse generell ermöglicht. Dies stellt einen klaren Fortschritt gegenüber aufwändigen intensitätsorientierten Ansätzen dar. Weiterhin liegt die Qualität, der in den Experimenten erzielten Ergebnisse, über den Möglichkeiten der störanfälligen, intensitätsorientierten Verfahren, wie zum Beispiel optischer Fluss.

### 3. MATCHING-PARADIGMA

---

Das Verfahren wird jedoch durch die vorausgesetzte Verwendung einer statischen Kamera in seinen Einsatzgebieten eingeschränkt.

Das entwickelte Verfahren ermöglicht an mehreren Stellen Erweiterungen und Verbesserungen. Diese hängen jedoch sehr vom jeweilig geplanten Einsatzgebiet ab. Eine praktische Aufgabe könnte sich zum Beispiel zur Parkplatzüberwachung oder Ampelsteuerung ergeben. Bei derartigen Anwendungen müsste eine Erweiterung realisiert werden, die es ermöglicht, Fahrzeuge auch beim Stoppen und Parken zu erfassen. Da die Differenzbildtechnik hierzu nicht in der Lage ist, müssten die kurz vor dem Halten erfassten Fahrzeugregionen gesondert verwaltet werden.

Eine bereits motivierte Aufgabe, ist die Entwicklung eines Verfahrens zur Auswertung der räumlich-zeitlichen Veränderungen des Vektorfeldes, mit dem Ziel einer stabilen Objektverfolgung. Dabei wird davon ausgegangen, dass ein Objekt durch eine Menge von Segmenten bzw. durch die Verschiebungsvektoren der Segmente beschrieben wird. In einer solchen Anwendung könnten insbesondere Schätzverfahren wie etwa der Kalman Filter benutzt werden, um auch in der Situation einer totalen Objektverdeckung, die zu einem vorübergehenden Ausfall der Verschiebungsvektoren eines Objektes führt, eine Verfolgung zu erreichen.

Eine weitere Aufgabe zur Verbesserung des Ergebnisses der Bewegungssegmentierung in Außenaufnahmen liegt in der automatischen Erfassung und Entfernung von Störregionen, die sich etwa durch Bewegungen von Bäumen bzw. Blättern, o.ä. ergeben. In Untersuchungen hat sich gezeigt, dass solche Regionen einen sehr hohen Texturgehalt sowie ein chaotisches Vektorfeld aufweisen. Auf dieser Grundlage könnten die Regionen klassifiziert und entfernt werden. In ähnlicher Weise könnten auch andere unvorhergesehene Bewegungsquellen bzw. Störungsquellen, die in der jeweiligen Anwendung auftreten, entfernt werden. Es ist allerdings klar, dass derzeit kein System alle möglichen Ursachen, die in der Realität zu Fehlern bei der Bewegungserfassung führen können, beherrscht.

Die in dieser Arbeit geschaffene Technik zur Bewegungserfassung und merkmalsorientierten Bestimmung der entsprechenden Bewegungsparameter, d.h. der Verschiebungsvektoren, dürfte jedoch eine brauchbare Grundlage in vielen Situationen darstellen. Die Ergebnisse, die in experimentellen Untersuchungen erreicht wurden, geben dafür genug Grund zur Annahme.

Das Verfahren benötigt keine kalibrierte Kamera, keine Straßenmodell und keine Kenntnisse über das Aussehen der zu detektierten Objekte. Damit lassen sich Objekte mit signifikanter Relativbewegung zur Kamera zuverlässig detektieren. Dazu ist das Verfahren robust gegenüber Teilverdeckung und Verschmelzung oder Zerfallen von Objekten aufgrund der erfassten kleinen Segmente, die auch eine Bewegungsbestimmung ermöglichen. Im Unterschied zu anderen Verfahren ist auch bei Bildobjekten, die keinen ausgesprochenen Flächencharakter besitzen, eine Bewegungsbestimmung möglich. Das Verfahren eignet sich zur Detektion von formveränderlichen Objekten und kann sich Objekte und Schattenwurf unterscheiden. Mit der Erweiterung auf den Meta-Objekten können Objekthypothesen generiert werden, die eine gute Beschreibung der Objekte liefern. Eine Klassifikation des Fahrzeugtyps (PKW, LKW, Motorrad) bei seitlich vorbeifahrenden Fahrzeugen kann auf diese Grundlagen implementiert mittels Support-Vector-Machine (SVM) oder Multi-Layer-Perception (MLP) realisiert werden. Der Rechenaufwand ist für ein Verfahren der Farbbildfolgenanalyse relative klein (Auf einem PC mit einem Prozessor System konnte eine Bildwiederholrate von 6 Bildern pro Sekunde erzielt werden).

### 4. Objektverfolgung auf der Basis von Datenassoziationen

Der Hauptaspekt der Betrachtungen in diesem Kapitel ist die Verwendung eines Partikelfiltersystems in Verbindung mit einer Datenassoziation für die Multi-Objektverfolgung in realen Bildsequenzen. Die Grundlage dafür ist das bekannte Bayes-Theorem, das im Folgenden hergeleitet und dessen verwendete Größen eingeführt werden. Die Anforderungen an das Verfolgungssystem werden anhand der Hypothesen, die aus dem Rahmen der Arbeit resultieren, und der zuvor ermittelten Problemsituationen aufgestellt. Anschließend werden die einzelnen Schritte in der Verarbeitungskette des entwickelten Partikelfiltersystems [Al-Hamadi 07a, 08b] betrachtet. Das Bildmaterial im Fokus dieser Arbeit wird auf bildspezifische Problemsituationen wie bspw. Teilverdeckung untersucht und verallgemeinert.

Als Grundlage für das Partikelfilter-Verfahren [Isard 98, Martnez 04, Vermaak 04] wird der MD-Ansatz [Al-Hamadi 03a, 05b] verwendet und hier kurz beschrieben. Darauf aufbauend werden die Initialisierung und die Zustandsbeschreibung vorgestellt, die der Formulierung des Bewegungs- und Beobachtungsmodells dienen. Um die Bewegungsgrößen eines Objektes bestimmen zu können, werden zwei Prozesse in Betracht gezogen. Zum einen wird ein adaptives Bewegungsmodell zur Beschreibung der verfeinerten Partikelentwicklung aufgestellt und zum anderen wird eine Differenzierung der Objekte bezüglich Korrespondenzanalyse durch eine Datenassoziation erfolgen. Objektkonstellationen aufgrund von Verschmelzung, Zerfallen oder Teilverdeckung in realen Bildsequenzen werden untersucht und für die Datenassoziation klassifiziert. Die vorgeschlagene Datenassoziation wird mit den Vorteilen herkömmlicher Assoziationsverfahren aufgestellt und mit einem Verwaltungsschritt abgerundet.

#### 4.1 Intuitive Einführung

Für relativ einfache Videoszenen stehen heute leistungsfähige Verfahren zur Berechnung von Bewegungsparametern bzw. für das Tracking von Objekten zur Verfügung, die meist beim Einsatz einigen Einschränkungen unterliegen [Rehrmann 94, Mecke 99, Brandner 02, Nummiaro 02, Köhler 05, Brox 04, Coifman 98, Pathan 09a, Vermaak 03, Sidenbladh 03]. Hier sind beispielsweise entweder die zu verfolgenden Objekte starken Bewegungseinschränkungen unterworfen oder der Bereich der Überwachung ist starr und darf sich nicht verändern. Kritische Szenen sind auch solche mit dichtem Verkehr, die bedingt zur schwierigen Merkmalsextrahierung führen und demzufolge ein Korrespondenzproblem in Bildsequenzen verursachen. Auch wenn der Hintergrund den eigentlichen Objekten recht ähnlich ist oder die sich bewegenden Objekte nicht lineare Bewegungen aufweisen, haben herkömmliche Verfahren große Schwierigkeiten und können die Zielobjekte verlieren. Dies führt zur Begrenztheit der herkömmlichen Verfahren zur Bewegungsschätzung. Ausgehend davon sind noch viele Probleme zu lösen, bis sich die bekannten Verfahren robust und routinemäßig anwenden lassen.

Ziel dieses Kapitels ist es, ein Verfahren zu entwickeln, das auf der Basis von Partikelfiltern die Verfolgung multipler Objekte sowie die stabile Ermittlung ihrer Trajektorien auch in Konfliktszenen ermöglicht. Die Herausforderung besteht demzufolge sowohl in der Initialisierung eines Partikelfilters für jedes Objekt und

Zustandsbeschreibung für die Verarbeitungsschritte eines Partikelfilters. Dazu werden drei Hauptprobleme aus der Objektverfolgung hier die *Zustandsprädiktion*, *Datenzuweisung* der Prädiktionen und Messungen und *Korrektur* des prädizierten Objektzustandes sein. Die dynamische Formulierung der Prädiktion stellt für eine sich verändernde Anzahl unabhängiger Objekten eine komplexe Herausforderung dar. Hierbei spielt das Wissen über Objektfluktuationen zur Erleichterung der Modellierung des Schätzprozesses eine entscheidende Rolle.

Um das Verständnis zu erleichtern, beschreibt dieser Abschnitt kurz die Theorie der Bayes-Filter<sup>65</sup> [Gnedenko 65, Doucet 03, Maskel 01, Stüker 04], da sie die Grundlage für die vorliegende Arbeit in diesem Kapitel darstellt. Dafür werden ein Bewegungsmodell, ein Beobachtungsmodell und eine A-priori-Wahrscheinlichkeit<sup>66</sup> der Zustände durch das Bewegungsmodell als bekannt angenommen. Um die Verfolgung der Objekte zu vereinfachen, wird von folgenden Bedingungen ausgegangen, die dem ganzen Verfolgungsprozess die Struktur eines Markov-Modells geben:

- Die Daten in Form von Bildsequenzen werden von einem Kamerasystem geliefert, wobei jedes Bild einen Zeitschritt darstellt. Die Objektzustände (z.B. Position eines Objektes sowie dessen Bewegungsgeschwindigkeit) zum Zeitpunkt  $k$  werden im Folgenden diskret im *Zustandsvektor*  $\mathbf{z}_k = [z_1, z_2, \dots, z_n]^T$  zusammengefasst. Die Messungen oder die Beobachtungen sind im *Messvektor*  $\mathbf{m}_k = [m_1, m_2, \dots, m_n]^T$  angegeben.
- Für die dynamische Beschreibung des Systems gilt, dass nur die direkte Vergangenheit gemäß der folgenden Beschreibung  $p(\mathbf{z}_k | \mathbf{z}_1, \dots, \mathbf{z}_{k-1}) = p(\mathbf{z}_k | \mathbf{z}_{k-1})$  berücksichtigt wird.
- Neue Messungen  $\mathbf{m}_k$  sind nur vom aktuellen Zustand  $\mathbf{z}_k$  abhängig. Für die Messungen  $\mathbf{m}_k$  wird angenommen, dass diese sowohl untereinander voneinander stochastisch unabhängig sind, als auch im Hinblick auf den dynamischen Prozess. Formal stellt sich das wie folgt:

$$p(\mathbf{m}_{k-1}, z_k | \mathbf{z}_{k-1}) = p(z_k | \mathbf{z}_{k-1}) \cdot \prod_{i=1}^{n-1} p(m_i | z_i) \quad (4-1-1a)$$

$$p(\mathbf{m}_k | \mathbf{z}_k) = \prod_{i=1}^{n-1} p(m_i | z_i) \quad (4-1-1b)$$

Gesuchte Informationen über Objektzustände im Folgebild einer Sequenz werden nun mit der A-posteriori-Wahrscheinlichkeit  $p(\mathbf{z}_k | \mathbf{m}_k)$  ermittelt, die man mit der Verknüpfung von A-priori- und A-posteriori-Wahrscheinlichkeit gemäß dem Bayes-Gesetz erhält. Daher ist die A-posteriori-Wahrscheinlichkeit  $p(\mathbf{z}_k | \mathbf{m}_k)$  proportional zum Produkt von A-priori-Wahrscheinlichkeit und der aus der Beobachtung resultierenden Likelihood-Funktion  $p(\mathbf{m}_k | \mathbf{z}_k)$  [Koch 00, Doucet 01].

$$p(\mathbf{z}_k | \mathbf{m}_k) = \frac{p(\mathbf{m}_k | \mathbf{z}_k) \cdot p(\mathbf{z}_k | \mathbf{m}_{k-1})}{p(\mathbf{m}_k)} = \frac{p(\mathbf{m}_k | \mathbf{z}_k) \cdot p(\mathbf{z}_k | \mathbf{m}_{k-1})}{\int p(\mathbf{m}_k | \mathbf{z}_k) \cdot p(\mathbf{z}_k | \mathbf{m}_{k-1}) dz_k} \quad (4-1-1c)$$

<sup>65</sup> Bayes-Filter sind Wahrscheinlichkeitsschätzer von Zuständen in dynamischen Systemen, die die A-posteriori-Wahrscheinlichkeit des zuschätzenden Zustandes aufgrund beobachteter Sensordaten und Handlungen in jedem Zeitschritt ermitteln. Daran anlehnend werden die auf Ereignisse bezogenen Betrachtungen für die Verteilungen stetiger ( $t$ ) und diskreter ( $k$ ) Zufallsvektoren verallgemeinert und für die Verwendung in zeitlichen Bildsequenzen formuliert.

<sup>66</sup> Das Vorwissen legt einen A-priori-Raum aller möglichen Werte des zu schätzenden Parameters fest.

Aus dieser Gleichung geht hervor, welche Größen benötigt werden, um die A-posteriori-Wahrscheinlichkeit  $p(\mathbf{z}_k | \mathbf{m}_k)$  in Abhängigkeit von der A-priori-Wahrscheinlichkeit<sup>67</sup>  $p(\mathbf{z}_k | \mathbf{m}_{k-1})$  berechnen zu können. Die Wahrscheinlichkeit  $p(\mathbf{m}_k)$  für eine Beobachtung des aktuellen Zeitschritts ist konstant und wird lediglich die Aufgabe der Normierung besitzen, die dazu führt, dass sich die einzelnen Wahrscheinlichkeiten des Zustandsraums zu „1“ aufsummieren. Allerdings ist sie wegen des Integrals im Nenner des rechten Terms von Gleichung (4-1-1c) oft schwierig zu bestimmen. Im Falle eines dynamischen Prozesses kann die analytische Berechnung von  $p(\mathbf{z}_k | \mathbf{m}_k)$  rekursiv in *zwei Schritten* erfolgen [Maskell 01].

Im Prädiktionsschritt erfolgt die Schätzung der neuen Zustandsparameter anhand eines erstellten Bewegungsmodells mit der Auftrittswahrscheinlichkeit  $p(\mathbf{z}_k | \mathbf{z}_{k-1})$ , mit dem die Schätzung über die möglichen Veränderungen des Objektzustands  $\mathbf{z}$  von Schritt  $k-1$  zu  $k$  aufgestellt werden können. Das Ergebnis der Prädiktion ist die effektive A-priori-Wahrscheinlichkeit  $p(\mathbf{z}_k | \mathbf{m}_{k-1})$ , die abhängig von der alten A-posteriori-Wahrscheinlichkeit  $p(\mathbf{z}_{k-1} | \mathbf{m}_{k-1})$  ist. Mit anderen Worten gesagt, lässt sich die Wahrscheinlichkeit des Systemzustands  $\mathbf{z}_k$  in Abhängigkeit von den Messungen  $\mathbf{m}_{k-1}$  bis zum letzten Zeitpunkt  $k-1$  dadurch gewinnen, dass man alle Hypothesen mit Statusvektor  $\mathbf{z}_{k-1}$  mit den Auftrittswahrscheinlichkeiten  $p(\mathbf{z}_k | \mathbf{z}_{k-1})$  des Bewegungsmodells gewichtet.

$$\text{Prädiktion:} \quad p(\mathbf{z}_k | \mathbf{m}_{k-1}) = \int p(\mathbf{z}_k | \mathbf{z}_{k-1}) \cdot p(\mathbf{z}_{k-1} | \mathbf{m}_{k-1}) d\mathbf{z}_{k-1} \quad (4-1-1d)$$

Im zweiten Schritt wird die A-posteriori-Wahrscheinlichkeit  $p(\mathbf{z}_k | \mathbf{m}_k)$  dadurch berechnet, dass man die ungewichtete A-priori-Wahrscheinlichkeit  $p(\mathbf{z}_k | \mathbf{m}_{k-1})$  durch die Ähnlichkeitsfunktion  $p(\mathbf{m}_k | \mathbf{z}_k)$  wichtet und gemäß Gleichung 4-1-1e normiert. Diese stellt nun wiederum Grundlage für die neue A-priori-Verteilung<sup>68</sup> dar.

$$\text{Aktualisierung:} \quad p(\mathbf{z}_k | \mathbf{m}_k) = \frac{p(\mathbf{m}_k | \mathbf{z}_k) \cdot p(\mathbf{z}_k | \mathbf{m}_{k-1})}{\int p(\mathbf{m}_k | \mathbf{z}_k) \cdot p(\mathbf{z}_k | \mathbf{m}_{k-1}) d\mathbf{z}_k} \quad (4-1-1e)$$

Um diese Gleichung auszuwerten, bedarf es einer geeigneten Repräsentation dieser Wahrscheinlichkeitsdichten, die ausreichend genau und auch schnell zu ermitteln sind. Dafür gibt es mehrere Lösungsansätze, die z.B. auf der Gaußverteilung wie beim Kalman-Filter [Kalman 85, Stüker 04, Fox 03] oder auf Partikeln wie beim Partikelfilter [Martnez 04] basieren. Die Herangehensweise zur Lösung dieser Gleichungen hängt von der Art der Prädiktion und Aktualisierung ab. Der einfachste Fall tritt auf, wenn das Bewegungsmodell einer linearen Transformation entspricht und die Streuung der Zustandsgrößen durch eine Normalverteilung sowie die Messungengenauigkeit durch eine Gaußverteilung beschrieben werden kann. Werden diese Bedingungen erfüllt, so kann eine optimale Lösung mit dem einfachen Kalman-Filter erzielt werden, auf den kurz eingegangen wird.

#### 4.1.1 Kalman-Filter

Der Kalman-Filter (KF) stellt eine A-posteriori-Verteilung durch eine parametrisierte Gaußverteilung dar, die allein durch die Vorgabe von statistischen Größen wie Mittelwert und Kovarianz das Bewegungs- und Beobachtungsmodell beschreibt. Dies

<sup>67</sup> Es wird zur Zeit  $k=0$  eine Initialverteilung  $p(\mathbf{z}_0)$  über dem Objektzustand angenommen.

<sup>68</sup> Die A-posteriori-Verteilung wird auch als Filterverteilung bezeichnet.

#### 4. OBJEKTVERFOLGUNG AUF DER BASIS VON DATENASSOZIATIONEN

führt dazu, dass in jedem Schritt (sei es nach der Prädiktion oder nach der Aktualisierung) die Zustände  $\mathbf{z}_k$  durch unimodale, normalverteilte Verteilungen  $N$  angenähert werden können. Die Schätzung der aktuellen Objektzustände<sup>69</sup> ergibt sich aus einer gewichteten Mittelung der aktuellen Messwerte und der vorherigen Schätzwerte. Die Gewichte werden aus den Unsicherheiten gebildet, die aus den neuen Mittelwerten und der Kovarianzmatrix  $\Sigma_k$  des  $n_x$ -dimensionalen Zustands folgen. Beim Kalman-Filter wird die Zustandschätzung für einen Zustand  $\mathbf{z}_k$  und die Integration der Messung  $\mathbf{m}_k$  durch folgende Gleichungen ermittelt.

$$\mathbf{z}_k = A\mathbf{z}_{k-1} + B\mathbf{u}_k + \omega_k \quad (4-1-2a)$$

$$\mathbf{m}_k = H\mathbf{z}_k + \zeta_k \quad (4-1-2b)$$

$$p(\mathbf{z}_k | \mathbf{u}_k, \mathbf{z}_{k-1}) \approx N(\mathbf{z}_k; A\mathbf{z}_{k-1} + B\mathbf{u}_k, R_k) \quad (4-1-2c)$$

Gleichung (4-1-2a) entspricht hierbei dem Bewegungsmodell, das seinen linearen Charakter der Matrix  $A$  mit der Dimension  $n_x \times n_x$  verdankt. Diese beschreibt, wie der  $n_x$ -dimensionale Zustandsvektor  $\mathbf{z}_k$  zu einem diskreten Zeitpunkt  $k$  vom Zustandsvektor  $\mathbf{z}_{k-1}$  und dem  $n_u$ -dimensionalen Steuereingangsvektor  $\mathbf{u}_k$  zum Zeitpunkt  $k-1$  propagiert. Die Steuermatrix  $B$  mit der Dimension  $n_x \times n_u$  berücksichtigt Steuerinformationen  $\mathbf{u}_k$ , die hier nicht gegeben sind. Diese legen den Einfluss des alten Zustands und des Steuereingangs auf den Folgezustand fest. Beim Tracking von Objekten im Straßenverkehr wird oft auf den Steuervektor verzichtet, da dieser nicht zur Verfügung steht. Der zeitdiskrete Term  $\omega_k$  ( $R_k$ ) steht für das Prozessrauschen mit der Dimension  $n_x$  und wird aus einer Normalverteilung nach  $N(\mathbf{z}_k, 0, \Sigma_k)$  mit der Kovarianzmatrix  $\Sigma$  bestimmt. Dieser enthält alle Einflüsse auf  $\mathbf{z}_{k+1}$  die nicht durch  $A$  und  $B$  modelliert sind. Um den Informationsgehalt von Messungen des Systemzustands für einen Kalman-Filter zu beschreiben, wird neben dem Bewegungsmodell noch ein Beobachtungsmodell benötigt (Gl. 4-1-2b). Dieses wird letztlich durch die Matrix  $H$  festgelegt und durch ein Messwertrauschen  $\xi_k$  nach  $N(\mathbf{z}_k, 0, Q_k)$  mit der Kovarianzmatrix  $Q$  und erwartungswertfrei ( $\mu_k = 0$ ) überlagert. Für die Herleitung dieser Gleichungen sowie deren einzelne Schritte der Prädiktion und Korrektur sei auf die Literatur [David 03, Mecke 00, Fox 03, Stüker 04, Schwarzenberg 05] verwiesen.

$$p(\mathbf{m}_k | \mathbf{z}_k) \approx N(\mathbf{m}_k; H\mathbf{z}_k, Q_k) \quad (4-1-2d)$$

$$p(\mathbf{z}_k | \mathbf{m}_{k-1}) \approx N(\mathbf{z}_k; \bar{\mu}_k, \bar{\Sigma}_k) \quad (4-1-2e)$$

Prinzipiell nimmt das KF-Verfahren zu jedem Zeitpunkt, zu dem eine Messung vorliegt, anhand des Bewegungsmodells eine Prädiktion des Systemzustands vor. Dabei basiert diese Prädiktion auf der letzten Schätzung des Systemzustands. Anhand der erfassten Messung wird dann diese A-priori-Zustandsschätzung aktualisiert, um die in der Messung enthaltene Information über den richtigen Systemzustand zu nutzen. Das Ergebnis ist dann ein A-posteriori-Schätzwert des Systemzustands, wobei die mit einem KF erstellten Schätzungen erwartungstreu sind und minimale Fehlervarianzen

<sup>69</sup> Die Zustandsbeschreibungen beziehen sich auf Parameter wie Position, Beschleunigung oder Geschwindigkeit.

aufweisen. Je geringer nun die Unsicherheit ist, desto höher ist die Gewichtung. Umgekehrt zeigen große Unsicherheiten zu grobe Schätzergebnisse im Vergleich zu den Messungen an [Köhler 05; Bar-Shalom 95, Stuker 04, Homberg 07].

Eine logische Erweiterung des Kalman-Filters stellt der Multi-Hypothesen-Tracker (MHT) dar, bei dem für jede Hypothese ein Kalman-Filter verwendet wird. Das Problem besteht darin zu entscheiden, wann eine Hypothese entsteht und wann mehrere wieder zu einer vereinigt werden [Pathan 09b, Pathan 09c]. Für diesen Zweck setzt man Heuristiken ein, die auch den Hauptaufwand des Verfahrens ausmachen, sowohl bei der Implementierung als auch bei der Laufzeit. Jede Hypothese wird in Abhängigkeit der aktuellen Messung mit  $\chi$  gewichtet, wobei  $n$  in der Gleichung 4-1-2f der Anzahl der Gaußverteilungen entspricht. Die Dichtefunktion des Objektzustandes  $p(\mathbf{z}_k | \mathbf{m}_k)$  zur Zeit  $k$  wird als Summe von Gaußfunktionen dargestellt. Jede Gaußfunktion wird durch einen Kalman-Filter aktualisiert. Aus der Größe der Übereinstimmung einer jeden Gaußfunktion  $N(\mu_k^{(n)}, \Sigma_k^{(n)})$  mit den Sensordaten berechnet sich ihre Gewichtung  $\chi_k^{(n)}$ . Zu jedem Zeitpunkt  $k$  wird für jede Hypothese  $n$  eine Gewichtung berechnet, die umso höher ausfällt, je besser die Hypothese in Übereinstimmung mit den Sensordaten steht. Durch ihre Fähigkeit, multimodale Verteilungen darstellen zu können, sind MHT-Algorithmen in einem breiteren Spektrum verwendbar und bilden eine logische Erweiterung der Kalman-Filter. Weil jede Hypothese auf Basis eines Kalman-Filters aktualisiert wird, gelten auch hier die Linearitätsannahmen:

- Das Prädiktionsmodell und das Messmodell müssen durch lineare Funktionen beschrieben werden.
- Die A-posteriori- und A-priori-Wahrscheinlichkeiten werden durch Gaußverteilungen beschrieben. Die Wahrscheinlichkeiten zum Zeitpunkt  $k$  sind somit durch den Mittelwert  $\mu_k$  und die Kovarianz  $\Sigma_k$  definiert.

$$p(\mathbf{z}_k | \mathbf{m}_k) \approx \sum_i \chi_k^{(n)} N(\mathbf{z}_k, \mu_k^{(n)}, \Sigma_k^{(n)}), \quad n = 1, \dots, N \quad (4-1-2f)$$

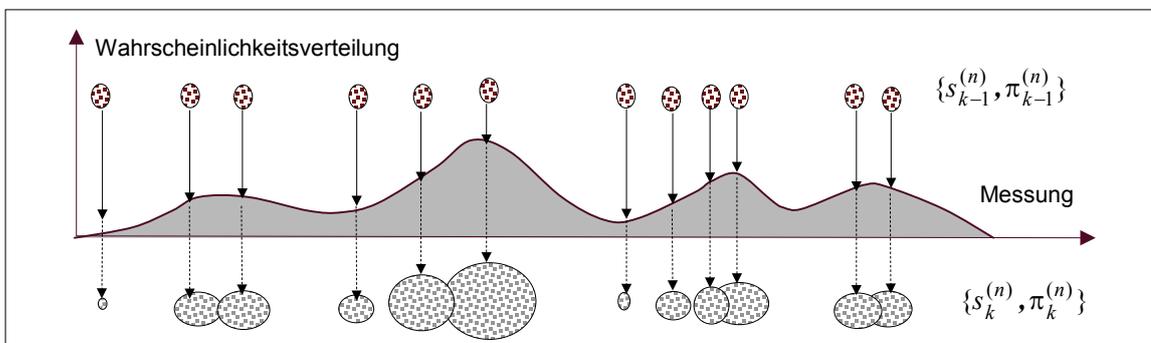
Auf diese Weise besteht die Möglichkeit, multimodale Wahrscheinlichkeiten zu modellieren und die Einsatzmöglichkeit im Gegensatz zum unflexiblen Kalman-Filter zu erweitern. Zur Umsetzung müssen hier zusätzliche Vorgehensweisen zur Handhabung der einzelnen Hypothesen in Betracht gezogen werden, die u.a. Entscheidungen über das Einfügen, Löschen und Verschmelzen von Hypothesen beinhalten können. Diese führen aber zur Einschränkung des ursprünglichen Effizienzvorteils des Kalman-Filters.

**Diskussion:** Der Kalman-Filter benötigt wenige Operationen bei den Messwertschätzungen und ist somit effizient im Rechenaufwand und für die Echtzeitanwendung gut geeignet. Allerdings setzen die bisherigen Betrachtungen des Kalman-Filters lineare Systeme voraus, die nicht ohne weiteres zu gewährleisten sind. Somit arbeiten die klassischen KF nicht zuverlässig für die Schätzung von Zuständen in Systemen, die mehrere unvorhersehbare Hypothesen zur Beschreibung benötigen. Allerdings bietet die Kalman-Theorie die Möglichkeit einer Erweiterung auf nichtlineare Systeme, die im Wesentlichen auf der Linearisierung der Systemgleichungen um den jeweiligen Arbeitspunkt beruhen. Dies kann z.B. durch eine Taylor-Reihenapproximation erfolgen, die nach dem zweiten Glied abgebrochen werden kann.

Durch die Mitführung der ersten Ableitung des Zustandsvektors kann dann die Zustandsbeschreibung  $A$  als Matrix (also linear) beschrieben werden. Der durch die Taylor-Approximation verbleibende Restterm wird dann von der Zufallsvariable  $\omega_k$  als "Unschärfe der Modellierung" übernommen. Ein KF, der ein auf solche Weise entstandenes System-Modell nutzt, heißt erweiterter Kalman-Filter (EKF) [Rath 01]. Infolgedessen liefert der EKF lediglich Näherungswerte für den Systemzustand, die darauf beruhen, dass erstens bei der Linearisierung der Terme<sup>70</sup> höherer Ordnung abgeschnitten wird (Truncation Error) und zweitens, dass die Linearisierung stets um den Schätzwert und nicht um den wahren Wert erfolgt (Base Point Error) [Stüker 04].

#### 4.1.2 Partikelfilter

Partikelfilter repräsentieren die zu schätzende Verteilung durch eine Menge von Samples. Die Samples werden auch als Partikel bezeichnet und geben den Filtern dieser Art ihren Namen. Im Gegensatz zum Kalman-Filter können bei dem Partikelfilter<sup>71</sup> die A-posteriori-Wahrscheinlichkeiten  $p(\mathbf{z}_k | \mathbf{m}_k)$  auch in nicht-linearen Systemen genauer bestimmt werden. Bei Partikelfiltern wird die Aufenthaltswahrscheinlichkeit des zu verfolgenden Objekts durch eine endliche Menge diskreter Einzelhypothesen beschrieben. Diese Menge wird auf der Grundlage eines Bewegungsmodells propagiert und durch die Anwendung eines Beobachtungsmodells neu bewertet (s. Abb. 4-1-1). Das Problem ist, dass es normalerweise unmöglich ist, effizient Samples aus der A-posteriori-Verteilung zu erhalten. Die Vermutung über den nächsten Zustand, ausgehend vom Bewegungsmodell und der alten A-posteriori-Wahrscheinlichkeit  $p(\mathbf{z}_{k-1} | \mathbf{m}_{k-1})$ , lässt sich mathematisch nicht ohne weiteres berechnen. Vor allem nicht, wenn es sich dabei um nichtlineare Systeme handelt.



**Abbildung 4-1-1:** Approximation einer beliebigen Wahrscheinlichkeitsverteilung durch Partikel. Die Kurve zeigt eine eindimensionale Wahrscheinlichkeitsverteilung, die von einer gewichteten Partikelmenge approximiert wird. Die Größe der Ellipsen repräsentiert die Größe des Partikelgewichtes.

Um eine gute Schätzung zu bestimmen, ist es entweder notwendig, dass die Partikelmenge den gesamten Zustandsraum ausreichend und gleichmäßig abdeckt oder aber man bedient sich einer Funktion, die angibt, wo sich die relevanten Teile der Wahrscheinlichkeitsverteilung im Zustandsraum befinden. So können die Partikel auf die wichtigen Teile im Zustandsraum verteilt und eine zuverlässige Schätzung erreicht

<sup>70</sup> Die Berücksichtigung dieser Terme wäre möglich, führt jedoch zu einem erhöhten Rechenaufwand [Stüker 04].

<sup>71</sup> Sequential Importance Sampling [Doucet 00], Bootstrap [Gordon 93] und Condensation gehören zu der Familie der Partikelfilter und werden auch als Sequentielle Monte-Carlo-Methode bezeichnet.

werden. Somit werden die zu schätzenden Verteilungen beim Partikelfilter durch eine Partikelmenge  $S_k$  verkörpert, die sich aus  $N$  Partikeln zusammensetzt, die die Zustandsvektoren  $\mathbf{z}_k$  und die zugehörigen Gewichte  $\pi_k$  (Gl. 4-1-3a) beschreiben. Die Gewichte summieren sich zu Eins auf und stellen jeweils die Likelihood-Funktion  $p(\mathbf{m}_k | \mathbf{z}_k)$ , des entsprechenden Zustandsvektors, dar. Die Aktualisierungsschritte erfolgen durch *Sampling-Methoden* auf den gewichteten Partikeln.

$$p(\mathbf{z}_k | \mathbf{m}_k) \approx S_k = \left\{ \langle s_k^{(n)}, \pi_k^{(n)} \rangle \mid n = 1, \dots, N \right\} \quad (4-1-3a)$$

Ein wesentliches Problem bei den Partikelfiltern ist das *Degeneracy-Problem*. Nach wenigen Iterationen werden, bis auf einen, alle restlichen Partikel vernachlässigbar, kleine Gewichte besitzen. In der Arbeit [Doucet 98] zeigte sich, dass die Varianz der Partikelgewichte mit der Zeit zunimmt, sich also dieses Problem nicht umgehen lässt. Aus dem Degeneracy-Problem ergibt sich der Effekt, dass ein großer Teil des Rechenaufwandes verwendet werden muss, um Partikel, deren Auftrittswahrscheinlichkeit gegen Null geht, zu aktualisieren. Eine Möglichkeit mit diesem Problem umzugehen ist der *Condensation-Algorithmus* [Isard 98]. Dadurch erfolgt eine Konzentration bei der Berechnung auf die Partikel mit den hohen Wahrscheinlichkeiten und nicht auf die ganze Umgebung oder gar den gesamten Zustandsraum. Dabei stellt der **Conditional Density Propagation-Algorithmus** eine iterative Form des Factored-Sampling-Verfahrens dar [Isard 98], das die Möglichkeit die A-posteriori-Wahrscheinlichkeitsdichte  $p(\mathbf{z}_k | \mathbf{m}_k)$  durch eine Menge von Partikel beschreibt.

Die Partikelmenge  $S_k$  zum Zeitpunkt  $k$  setzt sich aus  $N$  Partikeln zusammen. Jedes Partikel besitzt einen Zustand  $\mathbf{z}_k$  und ist mit einer Gewichtung  $\pi_k$  gepaart (Gl. (4-1-3a)). Zu jedem Zeitpunkt  $k$  existiert eine mit  $\pi_k = \{\pi_k^{(n)} \mid n = 1, \dots, N\}$  gewichtete Samplemenge  $S_k = \{s_k^{(n)} \mid n = 1, \dots, N\}$ , die die Wahrscheinlichkeitsdichte  $p(\mathbf{z}_k | \mathbf{m}_k)$  für den Aufenthaltsort eines Objektes wiedergibt. Der *Condensation-Algorithmus* gliedert sich in drei Schritte:

1. **BEWEGUNG:** Der Bewegungsschritt beinhaltet die Prädiktion des Bayes-Filters. Es wird der Zustand jedes Partikels durch das vom Anwender vorgegebene Bewegungsmodell prädiziert, dabei sind dem Bewegungsmodell keine Beschränkungen auferlegt. Die Entwicklung der Zustände  $\mathbf{z}_{k-1}$  in den Zustand  $\mathbf{z}_k$ , ist nicht zu verwechseln mit der Weiterentwicklung der Menge  $S_{k-1}$  in die Menge  $S_k$ . Dies geschieht erst in dem Sampling-Schritt.
2. **MESSUNG:** Die Messung bildet mit dem Sampling-Schritt die Aktualisierung des Bayes-Filters. Dazu werden die Partikelgewichte  $\pi_k^n$  berechnet. Diese bilden die Likelihood-Funktion  $p(\mathbf{m}_k | \mathbf{z}_k)$ . Der Zusammenhang zwischen einem Partikelgewicht und Likelihood-Funktion ist gegeben mit:

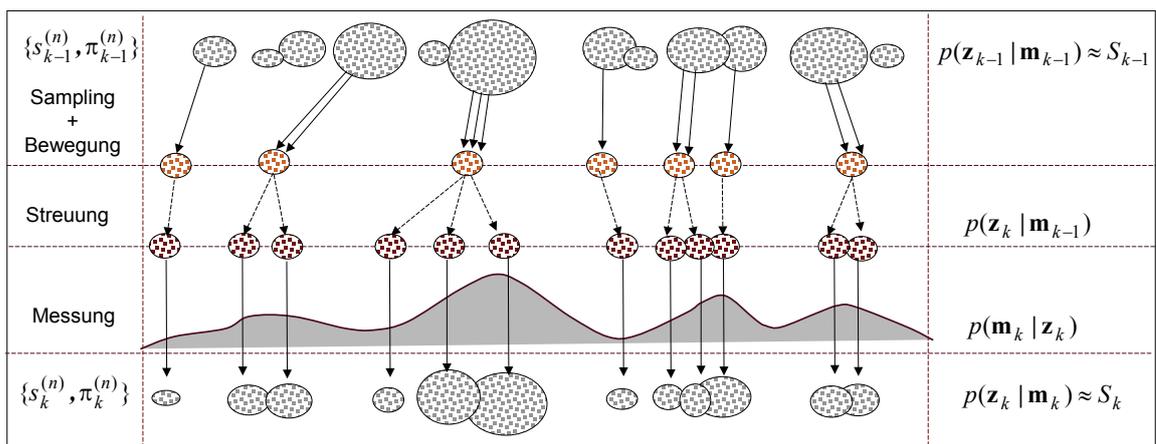
$$\pi_k^n = \frac{p(\mathbf{m}_k | z_k^n)}{\sum_{i=1}^N p(\mathbf{m}_k | z_k^i)} \quad (4-1-3b)$$

Der Nenner der Gleichung (4-1-3c) ist ein Normierungsfaktor, der Gewichte dermaßen normiert, dass sich eine Eins ergibt.

3. **SAMPLING:** Das Sampling erzeugt aus der A-priori-Wahrscheinlichkeit  $p(\mathbf{z}_k | \mathbf{z}_{k-1})$  und der durch die Kummulationskoeffizienten repräsentierten Likelihood  $p(\mathbf{m}_k | \mathbf{z}_k)$ , die neue Partikelmenge  $S_k$ , die der A-posteriori-Wahrscheinlichkeit  $p(\mathbf{z}_k | \mathbf{m}_k)$  entspricht. Dazu wird eine Menge von  $N$  gleichverteilten Zahlen  $r$  erzeugt [Isard 98]. Nach dem Prinzip des Ziehens und Zurücklegens, dessen Kumulationskoeffizienten am nächsten oberhalb der Zufallszahl ist, in die neue Partikelmenge aufgenommen<sup>72</sup>. Hoch gewichtete Partikel können also mehrmals ausgewählt werden, niedrig gewichtete auch überhaupt nicht. Dieser Schritt wird auch resampling genannt. Um das resampling im ersten Schritt der Partikelauswahl durch Ziehen und Zurücklegen optimal durchzuführen, wird jedes Partikel noch mit einer zusätzlichen Kumulationskoeffizient  $c_k^n$  erweitert. Die Partikel sind mit  $c_k$  der Größe nach sortiert, was es ermöglicht, das Ziehen mit Zurücklegen für ein Partikel in  $O(\log N)$  zu realisieren [Schwarzenberg 05].

$$c_k^n = \begin{cases} 0, & \text{wenn } n = 0 \\ c_k^{n-1} + \pi_k^n, & \text{sonst} \end{cases} \quad (4-1-3c)$$

Entsprechend der Gewichte  $\pi_{k-1}$  und der gegebenen Partikelmenge  $S_{k-1}$  des letzten Schritts wird eine neue Partikelmenge  $S_k$  für den aktuellen Zeitschritt  $k$  generiert (Sampling und Bewegung) (s. Abb. 4-1-2). Anschließend wird jedes Partikel  $s_k^{(n)}$  durch das Bewegungsmodell durch Überlagern von Rauschen variiert (Streuung). Anhand der aktuellen Messung werden die neuen Partikel mit der A-priori-Wahrscheinlichkeit  $p(\mathbf{z}_k | \mathbf{m}_{k-1})$  mit der Likelihood-Funktion  $p(\mathbf{m}_k | \mathbf{z}_k)$  gewichtet und schließlich das Gewicht durch das Beobachtungsmodell aktualisiert  $p(\mathbf{z}_k | \mathbf{m}_k)$ , um die neue Partikelmenge zu erhalten. Somit stellt der *Condensation* Algorithmus die fortlaufende Entwicklung einer Partikelmenge dar, die eine Verteilung annähert und repräsentieren kann. Hierbei besitzen die Partikelmengen die Möglichkeit, Wahrscheinlichkeitsverteilungen beliebiger Form zu approximieren.



**Abbildung 4-1-2:** Ablauf einer Iteration beim Condensation-Algorithmus. Die Kurve zeigt eine eindimensionale Wahrscheinlichkeitsverteilung, die von einer gewichteten Partikelmenge approximiert wird. Die Größe der Ellipsen repräsentiert die Größe des Partikelgewichtes.

<sup>72</sup> Dies entspricht der Monte-Carlo-Methode für die Erzeugung der neuen Partikelmenge mit der A-posteriori-Verteilung.

### 4.2 Einordnung und Ziele der eigenen Arbeiten

In diesem Abschnitt sollen die vielfältigen Einsatzgebiete des Partikel-Filters aufgezeigt und eigne Arbeiten in diesem Zusammenhang präsentiert werden. Es werden im Folgenden die einzelnen Varianten der Partikelfilter vorgestellt, die auf dem *Condensation*-Algorithmus basieren [Isard 98]. Der *Condensation*-Algorithmus ist in seiner Art problemunabhängig, da er sich lediglich um die Partikelverwaltung kümmert. Deterministische Variationen der Werte der Partikel erfolgen durch das Bewegungsmodell. Das Zuweisen von Wahrscheinlichkeitswerten an jedes Partikel übernimmt das Beobachtungsmodell.

Um die Entwicklung der Partikelmenge zu unterstützen, wurde der Kernel-Partikelfilter [Chang 03, 05] entwickelt, welcher eine durch Partikel repräsentierte Verteilung in eine kontinuierliche Approximierung überträgt. Eine darauf angewendete Mean-Shift-Methode dient der Gradientenschätzung und Maximumsuche, mit deren Hilfe die Partikel zu lokalen Maxima verschoben und neu gewichtet werden. Des Weiteren wurde ein Filtersystem aus Partikelfiltern und Kalman-Filtern zusammengestellt [Doucet 00, Murphy 01], in dem der nicht lineare Teil durch die Partikel des Partikelfilters angenähert und der lineare Teil durch den Kalman-Filter analytisch gelöst wird. Für die Effizienzsteigerung bei hochdimensionalen Zustandsräumen kann mit dieser so genannten Rao-Blackwellisations-Methode der Zustandsraum in einen linearen und nicht linearen Teil zur getrennten Behandlung zerlegt werden [Casella 96].

In weiteren Arbeiten [Sidenbladh 03] wurde die Bildszene mit mehreren Objekten in einem Partikelfilter erfasst, wobei jedes Partikel bzw. jeder Zustandsvektor dabei die Zustände aller Objekte enthält, die in der Szene enthalten sind. Somit beschreibt ein Partikel eine Variation aller systembeschreibenden Parameter, wodurch sich eine große Anzahl von Kombinationen ergibt. Auf diese Weise wird eine äußerst hohe Anzahl von Partikeln notwendig, um wahrscheinliche Zustände abzudecken und die Filterverteilung anzunähern. Des Weiteren ist eine Identifikation der einzelnen Objekte, was von Sidenbladh [Sidenbladh 03] ohnehin ausgeklammert wird, in dieser Aufstellung des Partikelfilters nur mit sehr großem Aufwand möglich.

Im Vergleich zum Standardpartikelfilter wurde der Mixture Particle Filter (MPF) [Vermaak 03] vorgestellt. Er wird aus mehreren Komponenten zusammengesetzt, wobei jede Komponente eine bestimmte Anzahl von Partikeln mit dem Zustand nur eines Objektes zugeordnet bekommt und in jedem Zeitschritt unabhängig entwickelt wird. Dieser Ansatz zeigt Anhaltspunkte zur Entwicklung eines stabileren Partikelfiltersystems zur Multi-Objektverfolgung. Zum einen ist eine geeignete Organisation der Partikel nötig, um diese im Filter zu unterscheiden und in Objektabhängigkeit differenziert zu entwickeln. Zum anderen ist für eine getrennte Analyse der Objekte eine Identifikation explizit erforderlich. Hierfür empfiehlt sich die Bewertung der Objekt-Messung-Beziehungen durch ein Assoziationsverfahren.

Ausgehend davon wurden verschiedene Arbeiten zur Objektverfolgung mit Partikelfiltern vorgestellt, die durch ihre Zusammensetzung Schwächen bei der unabhängigen Handhabung der Objekte aufweisen [Sidenbladh 03, Vermaak 03]. Dort beinhaltet jedes Partikel den Zustand der gesamten Bildsituation, wobei die Zahl der Kombinationen der Parameter, die durch das Bewegungsmodell variiert werden können, mit der Anzahl der Parameter und Objekte exponentiell steigt. Durch die vielen Kombinationen ist wiederum eine hohe Anzahl von Partikeln notwendig, um eine repräsentative Menge zu erstellen. Um eine aufwendige Assoziation zu umgehen, wählt beispielsweise

[Sidenbladh 03] diese Zustandsrepräsentation und begründet dies mit der nicht nötigen Objektidentifikation.

Immerhin lässt sich mit dieser Vorgehensweise keine sichere Unterscheidung der Objekte realisieren, da eine davon abhängige, differenzierte Bewegungsmodellierung somit auch nicht anwendbar ist. Auch bei der Repräsentierung des Zustands der Objekte im ganzen Bild [Vermaak 03] durch eine Filterverteilung ist die Unterscheidung von Objekten nicht möglich. Dort können in der Regel die Partikel zu Komponenten zusammengefasst werden. Jede Komponente bekommt eine bestimmte Anzahl von Partikeln mit dem Zustand nur eines Objekts zugeordnet. In jedem Zeitschritt  $k$  wird sie unabhängig bzw. individuell zeitlich entwickelt. Eine Folge davon, dass die Partikel einer Komponente immer nur mit einer Messung aktualisiert werden und dadurch werden die Partikel zu einer Partikelhäufung kollabieren. Die Objektidentifikation, vor allem nach Verdeckungen und unabhängige Bewegungsformulierung, ist auch hier nicht möglich.

Eine Erweiterung dieses vorgestellten Ansatzes um die Assoziation mit der Joint Probabilistic Data Association (JPDA) ist in [Vermaak 05] vorgeschlagen. Sie besagt, dass Assoziationswahrscheinlichkeiten aufgrund von Ähnlichkeitswerten zwischen Objekten und Messungen ermittelt werden und damit die Zuordnungsmöglichkeiten zwischen mehreren Objekten und Messungen bewertet werden. Mit dem Einsatz der herkömmlichen JPDA bleibt aber die Anwendung der Objektverfolgung auf Szenen mit fester und bekannter Anzahl von Objekten begrenzt.

Eine weitere Möglichkeit zur Datenassoziation kann durch ein Multi-Hypothesen-Verfahren realisiert werden [Cox 94]. Dabei werden nicht nur alle Zuordnungsmöglichkeiten zwischen den Objekten und Messungen im aktuellen Bild herangezogen, sondern auch alle Zuordnungsmöglichkeiten und -variationen über die vergangenen Zeitschritte betrachtet. Das ermöglicht die Handhabung einer variablen Objektanzahl, ist jedoch rechenaufwendig.

Fakt ist, dass der Einsatz von Partikelfiltern zur Multi-Objektverfolgung in verschiedenen Arbeiten untersucht wurde, in denen die Modellierung der Szene durch bloße Partikelfilteranwendung erfolgt. Allerdings bleiben die Identifizierung der einzelnen Objekte und die Ermittlung von deren Trajektorien jedoch aus [Sidenbladh 03, Vermaak 03]. In anderen Ansätzen wurde durch die Verwendung der Datenassoziation die Unterscheidung der Objekte mit simulierten Daten oder Radardaten untersucht. Allerdings wurden dort keine komplizierten Situationen behandelt [Kreucher 04, Karlsson 01]. Dabei bleibt die Anwendung jedoch auch auf Szenen mit konstanter und bekannter Objektanzahl begrenzt.

Ausgehend davon sind diese kurzen vorgestellten Verfahren objekt- und umgebungsabhängig. Schwierig gelten auch die automatische Initialisierung und die Erkennung von Konfliktsituationen, die zur Veränderung im Bild führen. Diese Veränderungen im Bild müssen in einem Filterschritt abgedeckt werden, auch wenn verschiedene und unabhängige Objekte enthalten sind. Zum einen sind unabhängige Bewegungsmodellierungen für die Objekte der Szene notwendig, die verschiedene Bewegungsmuster für Richtung und Geschwindigkeit aufweisen können. Zum anderen muss diese Bewegungsformulierung für jedes Objekt flexibel und dynamisch sein, damit die Vorhersage bei Veränderungen in der Bewegung angepasst werden kann. Somit erfüllen die herkömmlichen Verfahren auf der Basis von Bayes-Filter nur die Anforderung nach Robustheit und Flexibilität nur zum Teil.

Hierfür wird ein Verfolgungssystem auf der Basis von Partikelfiltern für die Analyse von Multi-Szenen entwickelt, in denen mehrere Objekte in mehrdeutige Situationen miteinander treten können. Die Herausforderung liegt in der Detektion und in der Verfolgung, vor allem aber in der Bestimmung der Trajektorie jedes einzelnen Objekts, selbst wenn die Objekte in verschiedener Hinsicht interagieren. Bisherige Schwächen vorhandener Verfahren, die durch partielle oder gar vollständige Verdeckungen oder durch stehen gebliebene Objekte entstehen, sollen deshalb untersucht und verbessert werden.

Eine geeignete Verarbeitung der umfangreichen Informationen soll in diesem System realisiert, um vor allem die Interaktion zwischen den Objekten auszuwerten und somit eine stabile Multi-Objektverfolgung zuzulassen. Dies führt dazu, dass das vorgeschlagene System ohne Voraussetzungen bzw. Wissen über die äußeren Bedingungen auskommt. Besonders wird hierbei auf die Zuverlässigkeit und Robustheit des Systems beim Auftreten bildspezifischer Störungen eingegangen.

### 4.3 Partikelfiltersystem mit Datenassoziation

Für die zu analysierenden Szenen mittels des vorgeschlagenen Partikelfiltersystems wird von einem statischen Aufnahmesystem, das Farbbildsequenzen liefert, ausgegangen. Die Szenen enthalten Objekte unterschiedlicher Art und unterschiedlichen Verhaltens, so genannte Multi-Objekte. Es wird kein Wissen zur Objektmodellierung eingebracht, da unterschiedliche Objekte mit unterschiedlichen Bewegungseigenschaften betrachtet werden können und es wird kein Umgebungsmodell verwendet, damit unterschiedliche Szenen betrachtet werden können. Ein wichtiges Grundelement des vorgeschlagenen Partikelfiltersystems ist die Verwendung des adaptiven Bewegungsmodells und des Beobachtungsmodells, die durch eine Datenassoziation verknüpft sind. Die Bedeutung der einzelnen Komponenten dieses Systems und deren Zusammenwirken lassen sich durch drei *Verarbeitungsschritte* beschreiben:

- Automatische Erfassung von Bildregionen
- Merkmalsextraktion und Zustandsbeschreibung
- Korrespondenzanalyse und Datenassoziation

Der erste Schritt realisiert die automatische Erfassung von bewegten Objekten (Initialisierungsschritt in Abbildung 4-3-1). In diesem Schritt können die Objekte unabhängig von ihrer Art und Form einfach durch ihre Bewegung erfasst und initialisiert werden. Dies beinhaltet die Farbbildaufnahme und die Objektselektion mittels des MD-Ansatzes in Verbindung mit der Nachverarbeitung von erfassten Regionen gemäß Kapitel 3. In der zweiten Verarbeitungsstufe des Systems wird die Merkmalsextraktion und Zustandsbeschreibung für die vorher selektierten Objekte aus der ersten Verarbeitungsstufe durchgeführt.

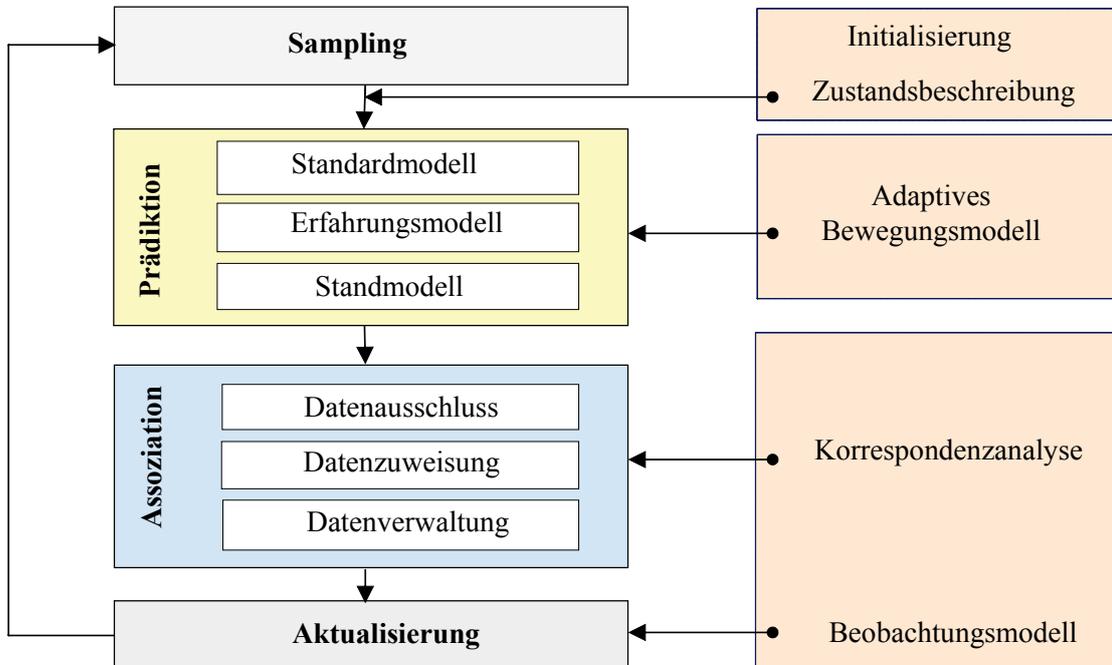
Die Korrespondenzanalyse und die Auswertung mittels Datenassoziation bilden die letzte Verarbeitungsstufe des Systems. Diese Verarbeitungsstufe (s. Abb. 4-3-1) beinhaltet die ursprünglichen Verarbeitungsschritte eines Partikelfilters<sup>73</sup> - Sampling, Prädiktion und Aktualisierung, die in dieser Arbeit so modifiziert werden, dass eine zuverlässige Bestimmung von Bewegungstrajektorien in Konfliktsituationen ermöglicht werden können. Diese Modifikationen beziehen sich erst auf die Prädiktion, die durch

---

<sup>73</sup> Als Spezialform von dem Condensation-Algorithmus

## 4. OBJEKTVERFOLGUNG AUF DER BASIS VON DATENASSOZIATIONEN

ein erweitertes *adaptives Bewegungsmodell* erfolgt, welches aus einem Standard-, Erfahrungs- und Standmodell besteht [Al-Hamadi 07a, 08a]. Diese drei Modelle werden in entsprechenden Konfliktsituationen durch eine Datenverwaltung gesteuert und eingesetzt. Zusätzlich wird der ursprüngliche Ablauf des Partikelfilters um einen 3-stufigen Assoziationsschritt mit *Datenausschluss*, *Datenzuweisung* und *Datenverwaltung* ergänzt, um die Problematik der Fluktuationserkennung und den Umgang mit Objektinteraktionen zu lösen [Al-Hamadi 07a, 08a].



**Abbildung 4-3-1:** Eine vereinfachte Darstellung der Einordnung der Verarbeitungsstufen im vorgeschlagenen Partikelfiltersystem. Die linke Hälfte zeigt den Ablauf des modifizierten Partikelfiltersystems mit seinen ursprünglichen Schritten Sampling, Prädiktion, Assoziation und Aktualisierung. Im rechten Teil sind die in dieser Arbeit bearbeiteten Schritte und Bestandteile des Partikelfilters in ihrer Reihenfolge eingeordnet.

Im rechten Teil der Abbildung 4-3-1 sind die einzelnen Schritte in ihren zeitlichen Einsatzpunkten im Partikelfilter gekennzeichnet. Nach der Initialisierung mittels der Bewegungssegmentierung erfolgt die Zustandsbeschreibung vor der Prädiktion. In der Prädiktion werden die Partikel gemäß eines Bewegungsmodells variiert. Um die Partikel im Aktualisierungsschritt zu bewerten, wird eine Ähnlichkeitsfunktion (so genannte Likelihood-Funktion) in Abhängigkeit von der Zustandsbeschreibung und die prädizierten Partikel definiert. Hier wird dem Aktualisierungsschritt noch ein Datenassoziationsschritt vorangestellt, der die Multi-Objektverfolgung in Konfliktsituationen ermöglicht. Die Initialisierung, das Bewegungs- und Beobachtungsmodell sind die Grundlage für das Voranschreiten eines Partikelfilters. Im nächsten Abschnitt wird die Bedeutung dieser drei *Verarbeitungsstufen* näher erläutert.

### 4.3.1 Automatische Erfassung von Bildregionen

In jedem Zeitschritt, also mit jedem neuen Bild, müssen Messdaten (bewegten Bildregionen) erfasst werden, die als Eingangsdaten für das Partikelfiltersystem dienen. Dies kann im Prinzip mit einem modellbasierten Verfahren z.B. Active Shape Modell

[Al-Hamadi 06g] bewerkstelligt werden. Allerdings geht aus der Ausgangsanalyse hervor, dass zum einen kein Vorwissen eingesetzt wird, mit dessen Hilfe Modelle formuliert werden können. Zum anderen ist für den Verfolgungsablauf auch eine automatische Initialisierung notwendig. Daher wird der MD-Ansatz für die Erfassung von bewegten Bildregionen angewendet [Al-Hamadi 03a, 05a, 06b], die von den Annahmen über ein statisches Aufnahmesystem ausgeht (s. Kapitel 3).

In Verbindung mit einem Nachverarbeitungsschritt werden Löcher durch morphologische Operationen geschlossen, die entstehen können, wenn innerhalb eines Objekts zu wenig Textur vorhanden ist und somit keine Bewegung an dieser Stelle gemessen wird. Auch zu kleine Regionen, die durch Kamerarauschen oder Luftverwirbelungen verursacht, werden durch Erosions- und Dilatationsvorgänge auf den Bewegungsmasken entfernt. Diese resultierenden Bildregionen bilden die so genannten Motion-Blobs (vgl. Kapitel 3). Dadurch werden die Informationen des gesamten Bildes auf die Bewegung reduziert und liefern dadurch schon feststehende Kandidaten für die Messung. Zur Vereinfachung der Initialisierung von Partikelfilter wird daher ein MB als ein Objekt angenommen und stellt am Ende die Grundlage zur Zustandsbeschreibung für die weiteren Verarbeitungsschritte des Trackingssystems dar.

### 4.3.2 Zustandsbeschreibung

Mit dem Ergebnis der automatischen Objekterfassung mittels MD-Ansatzes liegen Motion-Blobs als Träger für die Merkmalsextraktion und Zustandsbeschreibung vor. Diese beinhalten Objekte, die auch zur Unterscheidung und zeitlichen Zuordnung mit dementsprechenden Merkmalen charakterisiert werden können. Sie werden als Zustandsparameter zur Formulierung des Bewegungsmodells und des Beobachtungsmodells bei dem vorgeschlagenen Partikelfiltersystem angewendet. Der Zustandsvektor  $\mathbf{z}_k$  beinhaltet hier die Bildparameter  $\mathbf{z}_{b,k}$  und die Farbparameter  $\mathbf{z}_{co,k}$ , welche wie folgt beschrieben [Al-Hamadi 07a] werden.

**Bildparameter:** Es wird davon ausgegangen, dass ein erfasstes Objekt erstens eine kontinuierliche, gleichmäßige Bewegung aufweist, die dazu führt, dass keine komplizierte dynamische Formulierung des Zustands erforderlich ist. Daher sind für die Bewegungsvorhersage beispielsweise die Positionsbeschreibung, die Richtung und die Geschwindigkeit notwendig. Die Position  $(p_x, p_y)$  wird über den Schwerpunkt des jeweiligen erfassten MBs beschrieben. Zudem gehen aus der zeitlichen Zuordnung im Assoziationschritt ebenso Richtung und Geschwindigkeit mit dem Bewegungsvektor  $(d_x, d_y)$  hervor. Diese werden im ersten Teil des Zustandsvektors eines Objekts für die Bildparameter  $(\mathbf{z}_{b,k})$  zum Zeitpunkt  $(k)$  zusammengefasst:

$$\mathbf{z}_{b,k} = (p_{x,k}, p_{y,k}, d_{x,k}, d_{y,k})^T \quad (4-3-1a)$$

Die Bewegungseigenschaften jedes Objekts werden dadurch einfach erfasst und können ohne Rücksicht auf Objektart oder Objektverformung verwendet werden. Daher werden die Bildparameter in Gleichung 4-3-1a der Prädiktion dienen.

**Farbparameter:** Zur Unterscheidung der Objekte in und zwischen den Folgebildern werden Farbparameter herangezogen. Die Farbinformationen realer Szenen werden vom Sensor durch die atmosphärische Trübung lediglich als Farbtendenz erfasst, so dass Objekte mit unterschiedlichen Farbanteilen im HUV-Raum mit einer zusammenhän-

genden Punktwolke um die Grauwertachse<sup>74</sup> repräsentiert werden. Durch eine optimale Beschreibung dieser Punktwolke kann die typische Ausprägung der einzelnen Farbtendenzen für die Primitivenverbesserung ausgenutzt werden. Hier könnte der Einsatz des adaptiven Farbraumes<sup>75</sup> gemäß [Al-Hamadi 01a, 02a] eine optimale Farbbeschreibung und eine erhöhte Kontrastverstärkung in Richtung größerer Varianzen liefern (vgl. Kapitel 2). Durch die Vergrößerung der Adaptionsmechanismen gemäß [Al-Hamadi 01a] geht aber der adaptive Farbraum in den HUV-Farbraum über (vgl. Kapitel 2), der für die Aufgabenstellung in dieser Arbeit ausreichende Ergebnisse liefert. Aufgrund der Vielfalt der Zielobjekte ist eine geeignete Objektbeschreibung unabhängig von den räumlichen Anordnungen der Objektteile oder Farbsegmente im Bild notwendig. Durch eine Zerlegung und Neuordnung der UV-Ebene im HUV-Raum wird eine Trennung der einzelnen Farbanteile zur Objektcharakterisierung erreicht [Homberg 06], welches im Folgenden beschrieben wird (Abb. 4-3-2b). Die Neuordnung der Quadranten nach Vorschrift (4-3-1a/b) ermöglicht eine klare Trennung der Farbrichtungen und lässt eine Beschreibung charakteristischer Ausprägungen der Farbpunktwolken zu. Zudem liefert diese Trennung eine geeignete Grundlage für ein Clustering zur Bestimmung der Hauptfarbanteile.

$$U^+ = \begin{cases} U, & \text{falls } U \geq 0 \\ U + 2 \cdot U_{\max}, & \text{sonst} \end{cases} \quad (4-3-1b)$$

$$V^+ = \begin{cases} V, & \text{falls } V \geq 0 \\ V + 2 \cdot V_{\max}, & \text{sonst} \end{cases} \quad (4-3-1c)$$

Mit einer Teilung der UV-Ebene in ihre Quadranten und einer Verschiebung um das Doppelte von  $U_{\max}$  bzw.  $V_{\max}$ , der maximalen Ausdehnung der Farbebene in einem Quadranten, in U- bzw. V-Richtung werden die Quadranten an der Grau- bzw. Helligkeitsachse zerlegt und neu angeordnet. Die Objekte sind so als eine Zusammensetzung aus Farbwolken beschreibbar. Dabei liefern die Trennung und die U-V-Anteile der Wolken Merkmale für die Objektbeschreibung. Zur Beschreibung der Punktwolken werden zunächst die Punktwolken in Cluster<sup>76</sup> eingeteilt. Mit der Hauptkomponentenanalyse werden die einzelnen Clusterparameter wie Schwerpunkt  $s=(s_x, s_y)$ , stärkste Ausdehnung ( $ex$ ) und Ausrichtung  $r=(r_x, r_y, r_z)$  der stärksten Ausdehnung ermittelt und im zweiten Teil des Zustandsvektors  $z_{c,k}$  zu jedem Zeitpunkt k zusammengefasst [Al-Hamadi 07a, 08a].

$$z_{c,k} = (s^{(1)}, ex^{(1)}, r^{(1)}, \dots, s^{(cl)}, ex^{(cl)}, r^{(cl)})^T \quad (4-3-1d)$$

Dabei ist  $cl$  die Anzahl der gewählten Cluster. Die Anzahl der Cluster wird in Hinsicht auf die Objektbeschreibung durch dessen Hauptbestandteile bestimmt. Die größten Cluster repräsentieren die typischen größeren Objektteile. Kleine Cluster hingegen repräsentieren Ausreißer und können zeitlich differieren. Ursache dafür sind beispiels-

<sup>74</sup> Mittelwertbasierte Projektionskomponente mit 0,33 als Ladungsfaktoren für die einzelne Farbkanäle des RGB-Farbraumes.

<sup>75</sup> Inhaltsbasierte Projektionskomponenten in Abhängigkeit von der Streuung der Farbe im RGB-Farbraum.

<sup>76</sup> Hierfür wird ein bereits implementierter Algorithmus, der von [Niese 06] verwendet wurde, ausgeführt. Dieser ermittelt die Cluster basierend auf den Distanzen zwischen den Punkten durch eine den Regionenwachstums-Methoden ähnelnde Zuordnung.

## 4. OBJEKTVERFOLGUNG AUF DER BASIS VON DATENASSOZIATIONEN

weise kleine Farbanteile im Objekt, Störungen oder Randpixel im Segmentierungsergebnis. Um diese Unsicherheiten zu umgehen, werden die Cluster verwendet, die aus den größten Pixelanteilen des Objekts übertragen worden sind. Die experimentellen Untersuchungen in [Homberg 07] haben gezeigt, dass ein Objekt mit  $cl = 3$  Clustern typisch repräsentiert wird.

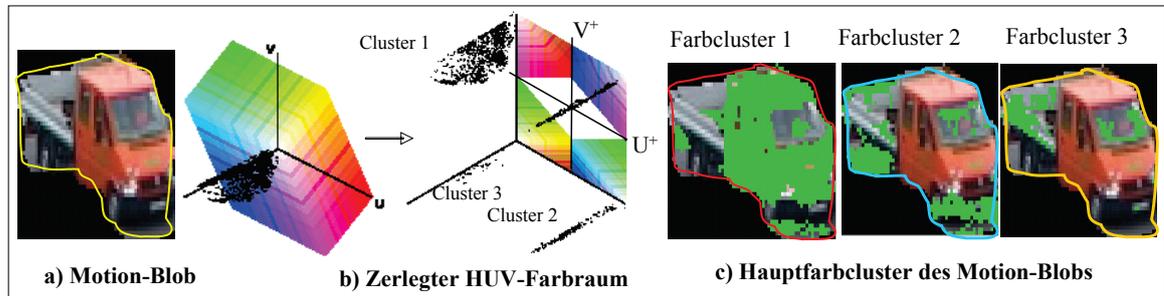


Abbildung 4-3-2: Die prinzipielle Darstellung der Einordnung der Farbcluster in dem zerlegten HUV-Farbraum.

Weitere, kleinere Punktwolken stellen wechselhafte Störungen, kleine instabile Regionen oder Randpixel dar und würden die grobe Objektbeschreibung dadurch destabilisieren. Im Ergebnis kann nun ein Objekt als Zusammensetzung aus drei Punktwolken durch deren Clusterparameter beschrieben werden. Auch zur Beschreibung der Farbparameter kann die Verwendung eines Farbsegmentierungsalgorithmus<sup>77</sup> (vgl. Kapitel 3) oder einer Farbhistogramme (vgl. Kapitel 6) ähnliche Ergebnisse liefern. Für die damit erzeugten Hauptregionen in Abhängigkeit von der Größe der Segmente könnten die Farbparameter (Gl. 4-3-1d) abgeleitet werden, indem jedes Segment durch eine Punktwolke darzustellen ist.

### 4.3.3 Korrespondenzanalyse und Datenassoziation

Die Verfolgung mehrerer Objekte in realen Bildsequenzen durch den Einsatz eines Partikelfiltersystems erfordert eine automatische Initialisierung, eine Zustandsbeschreibung und eine Definition von einem Bewegungs- und Beobachtungsmodell. Die Initialisierung wird mit dem MD-Ansatz gewährleistet, die die einzelnen MBs liefert, aus denen die Parameter für die Zustandsbeschreibung abgeleitet werden. Somit verläuft die Initialisierung der Zustandsvektoren<sup>78</sup> automatisch. Dazu stellt sich die Frage nach der Gestaltung der Partikelfilter, wie Objekte erfasst und die A-posteriori-Wahrscheinlichkeit  $p(\mathbf{z}_k | \mathbf{m}_k)$  berechnet werden. Für die Gestaltung des Partikelfiltersystems ist daher die *Differenzierung* zwischen den Objekten wichtig, die Anwendbarkeit der anschließenden Datenassoziation und die adaptive Bewegungsmodellierung für jedes einzelne Objekt. Dafür sind folgende Konstellationen zu berücksichtigen.

- Erstens werden die Objekte im frei fließenden Verkehr als unabhängig voneinander angenommen. Das bringt die Notwendigkeit mit sich, die Objekte in jedem Zeitschritt zu erfassen und den bisherigen Objekten und deren Bewegungstrajektorien

<sup>77</sup> Das Ziel eines Farbsegmentierungsverfahrens (z.B. das S&M- oder CSC-Verfahren) ist es, räumlich benachbarte und ähnliche Punktwolken zu Segmenten zusammenzufassen, die sich durch Parameter wie Schwerpunkt usw. mittels der Hauptkomponentenanalyse beschreiben lassen.

<sup>78</sup> Auf die gleiche Weise werden auch im weiteren Verlauf der Bewegungsanalyse neue Zustandsvektoren von neuen Objekten abgeleitet.

eindeutig zuzuordnen. Hier können die Partikel jeweils einen Objektzustand enthalten und werden zu den Komponenten  $K_{par}$  gruppiert, die einem Objekt zugeordnet werden, wodurch die Unterscheidung ermöglicht wird. Die Partikel werden nach einem einzigen Bewegungsmodell entwickelt, das für spezielle Aufnahmesituationen genutzt werden kann.

- Zweitens sind in speziellen Aufnahmesituationen, wie bei einem dichten Straßenverkehr oder Stau, die Objekte von einander abhängig und können sich in ihrer Bewegung beeinflussen. Ausgehend davon werden zwecks der Generalisierung der Szenen diese Situationen nicht getrennt betrachtet. Sie stellen nur einen weniger komplexen Spezialfall dar, der mit der Annahme über die Unabhängigkeit der Objekte sowie deren mögliche Zustände auch ausreichend abgedeckt werden kann.

Basierend darauf wird hier eine Zusammensetzung aus mehreren Partikelfiltern verwendet. Dabei wird ein Partikelfilter für ein Objekt eingesetzt, um eine Objektunterscheidung, eine Anwendbarkeit der Datenassoziation und Bewegungsmodellierung für jedes Objekt einzeln zu ermöglichen. Infolgedessen enthalten die Partikel (50 Partikel in unserem Fall) eines Filters lediglich den Zustand des zugehörigen Objekts. Beim Einsatz für unabhängige Objekte ist jedoch keine Variation zwischen den Partikelfiltern untereinander notwendig. Die Zustandsräume werden durch die Aufteilung der Zielobjekte auf die separaten Partikelfilter aufgeteilt, so dass jedes Objekt nur noch über einen sehr kleinen Zustandsraum geschätzt und repräsentiert wird. Dadurch verringern die Kombinationsmöglichkeiten und die Anzahl der Partikel für eine repräsentative Menge. Somit setzt sich die gesamte Filterverteilung des Partikelfiltersystems  $p(\mathbf{z}_k | \mathbf{m}_k)$  aus den einzelnen Verteilungen  $p^{(m)}(\mathbf{z}_k | \mathbf{m}_k)$  der  $m=1, \dots, M$  Partikelfilter zusammen. In Bezug auf die Gleichungen (4-1-1a) und (4-1-1b) ergeben sich hiermit für die Prädiktions- und Aktualisierungsschritte die Zusammensetzungen:

$$\mathbf{Prädiktion:} \quad p(\mathbf{z}_k | \mathbf{m}_{k-1}) = \sum_{m=1}^M \int p^{(m)}(\mathbf{z}_k | \mathbf{z}_{k-1}) \cdot p^{(m)}(\mathbf{z}_{k-1} | \mathbf{m}_{k-1}) d\mathbf{z}_{k-1} \quad (4-3-4a)$$

$$\mathbf{Aktualisierung:} \quad p(\mathbf{z}_k | \mathbf{m}_k) = \frac{\sum_{m=1}^M p^{(m)}(\mathbf{m}_k | \mathbf{z}_k) \cdot p^{(m)}(\mathbf{z}_k | \mathbf{m}_{k-1})}{\sum_{m=1}^M \int p^{(m)}(\mathbf{m}_k | \mathbf{z}_k) \cdot p^{(m)}(\mathbf{z}_k | \mathbf{m}_{k-1}) d\mathbf{z}_k} \quad (4-3-4b)$$

$$p(\mathbf{z}_k | \mathbf{m}_k) = \eta_k \cdot \sum_{m=1}^M p^{(m)}(\mathbf{m}_k | \mathbf{z}_k) \cdot p^{(m)}(\mathbf{z}_k | \mathbf{m}_{k-1}) \quad (4-3-4c)$$

Dabei ist  $\eta$  ein Normierungsfaktor. Im Gegensatz zu [Vermaak 03] wird hier keine Gewichtung zwischen den Komponenten bzw. Partikelfiltern vorgenommen. Zudem wird für jeden Partikelfilter ein eigenes Bewegungsmodell angewendet. Dadurch sind die Partikelfilter einzeln manipulierbar und somit an die jeweilige Gegebenheit in der Szene anpassungsfähig. Zusätzliche Anpassungsfähigkeit wird durch den Einsatz des zusammengesetzten Bewegungsmodells erreicht. So lassen sich unterschiedliche Bewegungsmodelle über der Zeit verändern und können die Partikelevolution jeweils entsprechend der Situation beeinflussen.

### 4.3.3.1 Adaptives Bewegungsmodell

Die *feststehenden* Bewegungsmodelle basieren auf eingebrachtem Wissen<sup>79</sup>, die aber bezüglich der Aufgabenstellung hier nicht ausreichend anwendbar sind. Zum einen enthalten die realen Bilder unterschiedliche Objekte, zum anderen können schon die Bewegungen eines Objekts vielfältig sein. Das zeigt den Bedarf, ein *adaptives Bewegungsmodell* zu formulieren. Das adaptive Bewegungsmodell gibt an, wie stark sich die Werte des Objektzustandes über die Zeit verändern können und hat die Aufgabe, die Zustandsvektoren der Partikel zu variieren. Dieser Vorgang entspricht der Prädiktion des Partikelfilters. Das Ergebnis der Prädiktion ist die effektive A-priori - Verteilung  $p(\mathbf{z}_k | \mathbf{m}_{k-1})$ . Demnach werden die Bewegungsvorschriften aufgestellt, die ausgehend vom letzten Zustand des Partikelfilters  $p(\mathbf{z}_{k-1} | \mathbf{m}_{k-1})$  mit dem Bewegungsmodell  $p(\mathbf{z}_k | \mathbf{z}_{k-1})$  die aktuellen Zustände der Partikel vorhersagen. Zu diesem Zweck wird zur Modellierung zunächst ein *Standardmodell* (MM<sub>N</sub>) aufgestellt, das an die gewöhnliche Bewegungsmodellierung bei Partikelfiltern angelehnt ist.

Hierfür wird der Objektzustand zur Zeit  $k$  mit einem Zustandsvektor  $\mathbf{z}_{b,k}$  beschrieben, der angibt, an welcher Position sich das Objekt im Bild befindet. Der Zustandsvektor  $\mathbf{z}_{b,k}$ , der die Position und die Geschwindigkeit in  $x$ - und  $y$ -Richtung im Bild kodiert, ist vierdimensional. Das  $n$ -te Partikel  $\mathbf{z}$  zum Zeitschritt  $k$  stellt sich folgendermaßen dar:  $\mathbf{z}_{b,k}^n = (p_{x,k}, p_{y,k}, d_{x,k}, d_{y,k})^T$ . Jedes Partikel entspricht einem spezifischen Objektzustand oder einer Hypothese davon.

Die Teilung der Zustandsparameter in Bild- und Farbparameter lässt hier eine einfache Erfassung der Änderungen des Objekts im Bild zu. Mit der Annahme über gleichmäßige Bewegungen und mögliche Bewegungsänderungen der erfassten Objekte in den Sequenzen wird jeweils die letzte Bewegung  $\mathbf{v}_{k-1} = (d_{x,k-1}, d_{y,k-1})^T$  verwendet. Damit wird aus dem letzten vierdimensionalen Zustand  $\mathbf{z}_{b,k-1} = (p_{x,k-1}, p_{y,k-1}, d_{x,k-1}, d_{y,k-1})^T$  der aktuelle Zustand  $\mathbf{z}_{b,k}$  geschätzt (Gl. 4-3-5a). Zusätzlich wird ein Rauschanteil  $\sigma = (\sigma_x, \sigma_x)^T$  addiert, der nach einer Normalverteilung  $N(\mu, \sigma)$  generiert wird, die Partikel variiert und somit einen größeren Bereich für die zugehörige Messung abdeckt. Dies entspricht einer linearen Transformation des Objektzustandes, die u.a. die Position<sup>80</sup> des Objektes verändert. In der Anwendung dieser Arbeit wird die Position und erfasste Geschwindigkeit in einfachen und eindeutigen Bewegungssituationen wie folgt ermittelt.

$$\text{MM}_N: \quad \mathbf{z}_{b,k} = f(\mathbf{z}_{b,k-1}, \mathbf{v}) + \sigma \quad (4-3-5a)$$

$$\begin{pmatrix} p_{x,k} \\ p_{y,k} \\ d_{x,k} \\ d_{y,k} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} p_{x,k-1} \\ p_{y,k-1} \\ d_{x,k-1} \\ d_{y,k-1} \end{pmatrix} + \begin{pmatrix} \sigma_x \\ \sigma_y \\ 0 \\ 0 \end{pmatrix} \quad (4-3-5b)$$

---

<sup>79</sup> In der Arbeit von Isard wurde das mittels eines 2-stufigen autoregressiven Prozesses modelliert. Dort wird zuerst das Bewegungsmodell per Hand initialisiert und anschließend iterativ mit immer schwierigeren, schnelleren Trainingssequenzen optimiert. Je nachdem wie komplex der Zustandsvektor des sich bewegenden Objektes ist, kann eine manuelle Parameterbestimmung des Dynamikmodells unbefriedigend sein.

<sup>80</sup> Die Position des Objektes wird damit in jedem Schritt mit dem normalverteilten Rauschen überlagert, um auf Positionsänderungen zu reagieren.

#### 4. OBJEKTVERFOLGUNG AUF DER BASIS VON DATENASSOZIATIONEN

Nach erfolgter Zuordnung der Objekte zu den Messungen werden auch die Parameter für die Bewegung  $\mathbf{v} = (d_{x,k}, d_{y,k})^T$  aktualisiert. Für dieses Bewegungsmodell würde im Grunde ein zweidimensionaler Objektzustand genügen, da die Positionsänderung und Bewegung des Objektes allein durch das Positionsrauschen eingeschätzt wird. Je nachdem wie schnell das betrachtete Objekt seine Geschwindigkeit verändern kann und wie hoch diese ist, muss man das Positionsrauschen vergleichsweise hoch ansetzen. Dadurch werden Partikel mit der falschen Position so lange propagiert, bis sie aufgrund schlechter Messwerte ein zu niedriges Gewicht erhalten und nicht mehr in die Partikelmenge übernommen werden.

Problematisch für dieses Modell sind insbesondere Bewegungen eines Objektes, bei denen es abrupt beschleunigt oder seine Richtung wechselt. In diesem Fall wird die bisherige Bewegung nämlich solange fortgesetzt, bis die Partikel mit falschem Geschwindigkeitsvektor“ absterben“ und Partikel in der näheren Umgebung von Messungen wieder stärkeren Einfluss auf die Schätzung haben. Um möglichst alle Bewegungen erfassen zu können, wird hier das gängige Standardmodell um zwei Modifikationen ergänzt. Damit wird eine adaptive Sucheigenschaft in der Partikelpropagierung erreicht, um die Partikelfilter in den Konfliktsituationen entsprechend der Situation des jeweiligen Objektes entwickeln zu können [Al-Hamadi 07, Al-Hamadi 08a].

- Die erste Modifikation, das *Erfahrungsmodell* (MM<sub>E</sub>), gilt bei Mehrdeutigkeiten in der Assoziation. Das heißt, die untersuchte Objektprädiktion erhält keine sichere Zuordnung zu einer Messung. In diesem Augenblick soll die Partikelmenge anhand der Erfahrung, aber ohne Aktualisierung durch eine Messung weiter entwickelt werden. Der Unterschied zum Standardmodell ist, dass in  $\bar{\mathbf{v}} = (d_{x,\bar{k}}, d_{y,\bar{k}})^T$  die Durchschnittsbewegungen mehrerer zurückgelegten Zeitschritte, also die Erfahrung über die Objektbewegung, verwendet wird. Da in diesen Fällen die Messungen nicht eindeutig zum Objekt zugeordnet werden können und somit keine Aktualisierungsgrundlage vorliegt, werden die erzeugten Partikel nicht vom Beobachtungsmodell gewichtet.

$$\text{MM}_E: \quad \mathbf{z}_{b,k} = f(\mathbf{z}_{b,k-1}, \bar{\mathbf{v}}) + \sigma \quad (4-3-5c)$$

$$\begin{pmatrix} p_{x,k} \\ p_{y,k} \\ d_{x,k} \\ d_{y,k} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} p_{x,k-1} \\ p_{y,k-1} \\ d_{x,\bar{k}} \\ d_{y,\bar{k}} \end{pmatrix} + \begin{pmatrix} \sigma_x \\ \sigma_y \\ 0 \\ 0 \end{pmatrix} \quad (4-3-5d)$$

Die Partikelevolution wird nun so gelenkt, dass die Partikel hauptsächlich in die gewohnte Richtung des Objekts entwickelt werden. Partikel, die sonst durch die Aktualisierung leicht gewichtet werden und absterben, bleiben nun aber durch die ausbleibende Aktualisierung bestehen und streuen so einen größeren Bereich ab, um eventuelle Bewegungsänderungen in den unsicheren Zeitschritten einzukalkulieren. Für die Anwendung der Arbeit wird die Position wie oben (Gl. 4-3-5c) ermittelt und außerdem die bekannte Bewegung wie Gleichung (4-3-5c) übernommen. Mit einer sicheren Zuordnung einer Messung wird wieder das Standardmodell verwendet.

- Die zweite Modifikation wird mit dem *Standmodell* ( $MM_S$ ) realisiert, das angewendet wird, wenn dem Objekt keine Messung gegenüber steht. Eine solche Situation tritt z.B. bei stoppenden Objekten aufgrund einer Ampel oder eines Staus im Straßenverkehr ein. Die automatische Objekterfassung kann dann an entsprechender Stelle keinen Kandidaten mehr finden. Dies ist auf den angewendeten MD-Ansatz zurückzuführen, der keine Änderung im Bild registrieren kann. Die automatische Aktivierung dieser Modelle wird ebenfalls in der Datenassoziation ermöglicht und organisiert, wo an der letzten bekannten Stelle nach den letzten bekannten Daten gesucht wird. Hier sorgt das Standmodell beispielsweise dann für einen Stopp in der Partikelevolution. Stattdessen wird die letzte Partikelmenge beibehalten und in jedem Zeitschritt von neuem gestreut bis wieder eine Messung zur Verfügung steht. In der Datenassoziation wird damit wieder das Standardmodell eingeschaltet. Für die Parameter wird nur von der letzten bekannten Position  $(p_{x,\hat{k}}, p_{y,\hat{k}})^T$  aus gestreut.

$$MM_S: \quad \mathbf{z}_{b,k} = f(\mathbf{z}_{b,\hat{k}}) + \sigma \quad (4-3-5e)$$

$$\begin{pmatrix} p_{x,k} \\ p_{y,k} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} p_{x,\hat{k}} \\ p_{y,\hat{k}} \end{pmatrix} + \begin{pmatrix} \sigma_x \\ \sigma_y \end{pmatrix} \quad (4-3-5f)$$

Diese drei Modelle setzen sich zu einem adaptiven Bewegungsmodell zusammen, das in der Datenassoziation (s. Abschnitt 4.3.4) gesteuert wird. Dabei ist immer nur ein Modell aktiv, das sich über das ausschließende Oder (XOR) gemäß Gleichung (4-3-5g) ergeben wird [Al-Hamadi 07,08]. Daher erweitert die Verwendung unterschiedlicher Bewegungsmodelle die Entwicklungsmöglichkeiten der Partikel in einem Filter. Mit den Bewegungsmodellen kann auf zusätzliche äußere Einflüsse reagiert und der Partikelfilter bzw. dessen Partikelevolution entsprechend angepasst werden. Die Entwicklung läuft bisher mit den Veränderungen des jeweiligen Objekts einher. Die Organisation der Modelleinsätze erfolgt in der Datenassoziation, die zur Erweiterung des Verfolgungssystems auf multiple Objekte betrachtet wird.

$$MM_A = MM_N \oplus MM_E \oplus MM_S \quad (4-3-5g)$$

Nachdem das Bewegungsmodell die Partikelmenge für den aktuellen Zustand prädiert hat, agiert das anschließende Beobachtungsmodell. Dort werden Ungenauigkeiten und Besonderheiten des angewendeten Bildsensors<sup>81</sup> modelliert, um die erfassten Messungen mit ihren Unsicherheiten besser in den Schätzprozess zu integrieren. Dies soll im Folgenden vorgestellt werden.

##### 4.3.3.2 Bestimmung des Beobachtungsmodells

Nach der Bewegungsschätzung erfolgt die Bewertung der prädierten Partikel, indem der prädierte Objektzustand  $\mathbf{z}_k$  anhand der aktuellen Messung  $\mathbf{m}_k$  durch das Beobachtungsmodell<sup>82</sup> nach ihrer Ähnlichkeit  $p(\mathbf{m}_k | \mathbf{z}_k)$  nach Gleichung (4-3-4c) gewichtet wird. Dafür liefern die effektive A-priori-Wahrscheinlichkeit  $p(\mathbf{z}_k | \mathbf{m}_{k-1})$  aus dem Prädiktions-

<sup>81</sup> Problematisch sind dabei das unumgängliche Rauschen des Bildsensors wie auch die Verzerrungen, die durch die verwendete Linse entstehen. Informationen darüber sind nicht gegeben und sollen hier auch nicht bestimmt werden, da höchste Präzision nicht gefordert ist.

<sup>82</sup> Das Beobachtungsmodell beschränkt sich für diese Arbeit auf die Analyse der Bildfolgen im Allgemeinen und die Objektmodellierung und zugehörige Merkmalsextraktion im Speziellen.

schritt in Abhängigkeit von dem Bewegungsmodell  $p(\mathbf{z}_k | \mathbf{z}_{k-1})$  und die Messungen im neuen Bild die Grundlage. Das ist auch die wesentliche Funktion der Aktualisierungsgleichung, die nach dem Bayes-Gesetz erfolgt. Dies besagt, dass die A-priori-Wahrscheinlichkeit  $p(\mathbf{z}_k | \mathbf{m}_{k-1})$  durch die Likelihood-Funktion  $p(\mathbf{m}_k | \mathbf{z}_k)$  gewichtet wird und somit die A-posteriori-Verteilung  $p(\mathbf{z}_k | \mathbf{m}_k)$  ergibt, die nun wiederum Grundlage für die neue A-priori-Verteilung ist. Zunächst sind die Identifikation und die Zuordnungen der verschiedenen Objekte und Messungen notwendig. Hierfür werden Merkmale herangezogen, die die Objekte möglichst eindeutig beschreiben. Dafür stehen in dieser Arbeit die zuvor diskutierten Bild- und Farbparameter zur Verfügung. Diese werden zur Formulierung von *Ähnlichkeitsmaßen* auf zwei Ebenen führen, die neben dem *Gewichtungsschritt* des Partikelfilters auch im Verarbeitungsschritt der Datenassoziation als Wahrscheinlichkeitswerte angewendet werden. Als Erstes wird ein Ähnlichkeitsmaß  $SM_{Obj}$  in der Zuordnungsebene verwendet, um die Objekte gegen die Messungen zu vergleichen. Diese besteht aus zwei Teilen (Gl. 4-3-6a/b).

- Der erste Teil ist die Ähnlichkeit  $SM_{Par}$ , die die Bildparameter des Objektes gegenüber der jeweils betrachteten Messung bewertet. Das Ähnlichkeitsmaß  $SM_{Par}$  setzt sich aus den gewichteten Ähnlichkeiten für die Position  $SM_{pos}$ , Geschwindigkeit  $SM_{spe}$  und Bewegungsrichtung  $SM_{dir}$  zusammen. Dadurch sollen Partikel, die von der Messung entfernt liegen oder andere Bewegungseigenschaften aufweisen, über die Bildparameter versetzt werden.
- Im zweiten Teil wird ein Vergleich der Farbclusterbeschreibung durchgeführt, um Mehrdeutigkeiten bei der Lage und der Bewegung zu beseitigen. Die Farbcluster der Objektprädiktionen und Messungen werden über die Ähnlichkeiten  $SM_{pos,i}$ ,  $SM_{ext,i}$  und  $SM_{dir,i}$  bewertet.

Das Ähnlichkeitsmaß<sup>83</sup> für die Partikelgewichtung  $SM_{Par}$  und die Objektvergleiche  $SM_{Obj}$  fassen sich wie folgt zusammen:

$$SM_{Par} = w_1 \cdot SM_{pos} + w_2 \cdot SM_{spe} + w_3 \cdot SM_{dir} \quad (4-3-6a)$$

$$SM_{Obj} = w_F \cdot SM_{Par} + w_C \cdot \sum_{i=1}^{cl} (w_5 \cdot SM_{pos,i} + w_6 \cdot SM_{dir,i}) \quad (4-3-6b)$$

In den Gleichungen (4-3-6a/b) wirken  $w_F$ ,  $w_C$  und  $w_1$  bis  $w_5$  als Normierungskonstante. Der Vergleich der Objekte mit den Messungen erfolgt wie die Organisation der Bewegungsmodelle im Schritt der Datenassoziation. Die einzelnen Ähnlichkeitsmaße werden im Intervall [0,1] angegeben, wobei die Bewertungsrichtung steigend ist. Der Wert 1 bedeutet die maximale Übereinstimmung. Für den Vergleich der geschätzten Position  $\mathbf{p}_k = (p_{x,k}, p_{y,k})^T$  und der entsprechenden Messung  $\mathbf{p}_{ref} = (p_{x,ref}, p_{y,ref})^T$  wird der euklidische Abstand verwendet. Mit der Distanz  $dist_{max}$ , die bei der Kandidatenzuordnung im Datenausschluss der Datenassoziation im Höchstfall erreicht werden darf, werden die Abstände normiert und durch Subtraktion in die vorgesehene Bewertungsrichtung gebracht (s. Gleichung 4-3-6c). Die Ähnlichkeitsmaße für den Vergleich der Ausdehnung der Farbcluster  $SM_{ext}$  sowie der Geschwindigkeiten der Partikel und Objekte  $SM_{spe}$  lassen sich analog ermitteln. Die Differenz der Geschwindigkeiten der

<sup>83</sup> Das Ähnlichkeitsmaß  $SM_{Obj}$  gehört nicht direkt zum Beobachtungsmodell. Da es jedoch der Findung der jeweiligen Messungen zur Partikelgewichtung dient und auf den Zustandsparametern basiert, wird es an dieser Stelle vorgestellt.

Prädiktion  $\mathbf{v}_k = (d_{x,k}, d_{y,k})^T$  und der Messung  $\mathbf{v}_{ref} = (d_{x,ref}, d_{y,ref})^T$  ist dafür ausreichend (s. Gleichung 4-3-6d). Diese Angabe wird ebenfalls normiert.

$$SM_{pos} = 1 - \frac{\|\mathbf{p}_k - \mathbf{p}_{ref}\|}{\text{dist}_{\max}} \quad (4-3-6c)$$

$$SM_{spe} = \frac{1}{1 + \|\mathbf{v}_k - \mathbf{v}_{ref}\|} \quad \text{oder} \quad SM_{spe} = 1 - \frac{\|\mathbf{v}_k - \mathbf{v}_{ref}\|}{\mathbf{v}_{\max}} \quad (4-3-6d)$$

Die Vektoren der Bewegungsrichtung  $\mathbf{v}_k$  der Partikel und Objekte lassen sich mit  $\mathbf{v}_{ref}$  der Referenzmessung über die Winkel zwischen den Vektoren vergleichen. Mit Hilfe des Skalarproduktes und den Beträgen der Vektoren wird dieser Winkel ermittelt. Das Winkelmaß liegt zwischen  $[-1, 1]$  und wird auch für die weitere Verwendung normiert.

$$SM_{dir} = \frac{1}{2} \left( \frac{\mathbf{v}_k \circ \mathbf{v}_{ref}}{\|\mathbf{v}_k\| \cdot \|\mathbf{v}_{ref}\|} + 1 \right) = \frac{1}{2} \cdot \left( \frac{(d_{x,k}, d_{y,k})^T \circ (d_{x,ref}, d_{y,ref})^T}{\|(d_{x,k}, d_{y,k})^T\| \cdot \|(d_{x,ref}, d_{y,ref})^T\|} + 1 \right) \quad (4-3-6e)$$

Da alle Ähnlichkeiten im Intervall  $[0, 1]$  liegen und die gleiche Bewertungsrichtung haben, lassen sie sich in den Ähnlichkeitsmaßen  $SM_{Obj}$  und  $SM_{Par}$  gleichrangig verwenden und gewichten.

#### 4.3.4 Datenassoziation

In den bisherigen Überlegungen zur Funktionsweise und Anwendung von Partikelfiltern wurde von einfachen Szenen ausgegangen. Reale Situationen, in denen mehrere Objekte auftreten und Konfliktsituationen<sup>84</sup> auslösen, wurden bisher außer Acht gelassen. Demzufolge besteht das Problem sowohl in der Datenzuordnung als auch in der Zustandsbeschreibung in Konfliktsituationen. Während eine bekannte und konstante Anzahl von Objekten die Datenordnung bei weitem erleichtert, stellt eine dynamische Formulierung für eine unkonstante Anzahl an unabhängigen Objekten eine komplexe Herausforderung dar.

Um eine optimale Datenzuordnung in Konfliktsituationen zu finden, wird in der Regel eine Datenassoziation angewendet. Für diese existieren verschiedene Methoden, die auf einem Nearest-Neighbor-Verfahren (NN-Verfahren) [Pavlina 03], einem PDA-Ansatz (Probabilistic Data Association) [Helmick 00], einem JPDA-Ansatz (Joint Probabilistic Data Association) [Vermaak 05] sowie auf der Multi-Hypothesen-Technik (MHT) [Cox 94] basieren können. Da sich die Betrachtung in diesem Abschnitt im Wesentlichen mit der Entwicklung einer Datenassoziation beschäftigt, sollen die oben genannten verwendeten Methoden näher betrachtet werden<sup>85</sup>:

In dem Nearest-Neighbor-Verfahren erfolgt die Zuordnung der Messungen zu Objektprädiktionen in Abhängigkeit ihrer Abstände zueinander schrittweise, bis alle Messungen zugewiesen sind. Eine Messung wird als Folge jenes prädizierten Objekts zugeordnet, zu dem sie am nächsten liegt. Demzufolge werden gebildete Zuordnungs-paare aus der Messung und der Prädiktion als richtig angenommen und somit mögliche

<sup>84</sup> Die Ursachen dafür sind auf nicht-kooperativen Messsituationen, wie beispielsweise Verschmelzen, Teilen, der Beginn oder das Ende einer Verdeckung, zurückzuführen.

<sup>85</sup> Die Betrachtung des Standes der Technik dient diesbezüglich der Zordenung eigener Arbeiten

falsche Zuordnungen toleriert. Allerdings wird mit der alleinigen Verwendung des Abstandes jedoch nicht notwendigerweise die optimale Lösung geliefert. Dies tritt meist dann ein, wenn sich Objekte dicht bei einander bewegen oder sich ihre Wege kreuzen. Im Gegensatz zum NN-Verfahren geht der PDA-*Ansatz* bei der Zuordnung von einem Suchbereich aus, der um ein einzelnes Objekt mit zuzuordnenden Messungen definiert wird. Diese Messungen werden entweder vom entsprechenden Objekt erzeugt oder durch Stördaten verursacht. Bei  $i$ -Messwerten innerhalb des Suchbereiches können im Prinzip  $i+1$  Hypothesen aufgestellt werden. Die Hypothesen  $\mathbf{H}^{(i)}$ , ( $i = 1, \dots, M$ ) deuten darauf hin, dass die Messung  $i$  vom entsprechenden Objekt erzeugt wurde. Für jedes betrachtete Objekt wird mit den  $M$  Messungen eine so genannte gewichtete Gesamtinnovation  $E_k$  innerhalb des Suchbereiches aus den  $i$ -Messwerten gemäß Gleichung (4-3-7a) [Bar-Shalom 95, Helmick 00] ermittelt.

$$E_k = \sum_{i=1}^M \beta_k^{(i)} E_k^{(i)} \quad (4-3-7a)$$

Hierbei ist  $E_k^{(i)}$  die Innovation<sup>86</sup> des  $i$ -ten Messwertes bezüglich des betrachteten Objekts und  $\beta^{(i)}$  die Wahrscheinlichkeit, dass der Messwert zu dem Objekt gehört. Die Summe der Wahrscheinlichkeiten  $\beta^{(i)}$  ist 1. Die Gesamtinnovation  $E_k$  geht mit  $\beta^{(0)}$ , der Wahrscheinlichkeit der Hypothese  $H^{(0)}$ , in den Aktualisierungsschritt als Gewichtung für die Zuordnungspaare ein. Diese Modellierung bezieht sich im Prinzip auf die Verwendung eines Kalman-Filters. Daher sei an dieser Stelle auf die Herleitung der Wahrscheinlichkeiten auf [Bar-Shalom 95, Homberg 07] verwiesen.

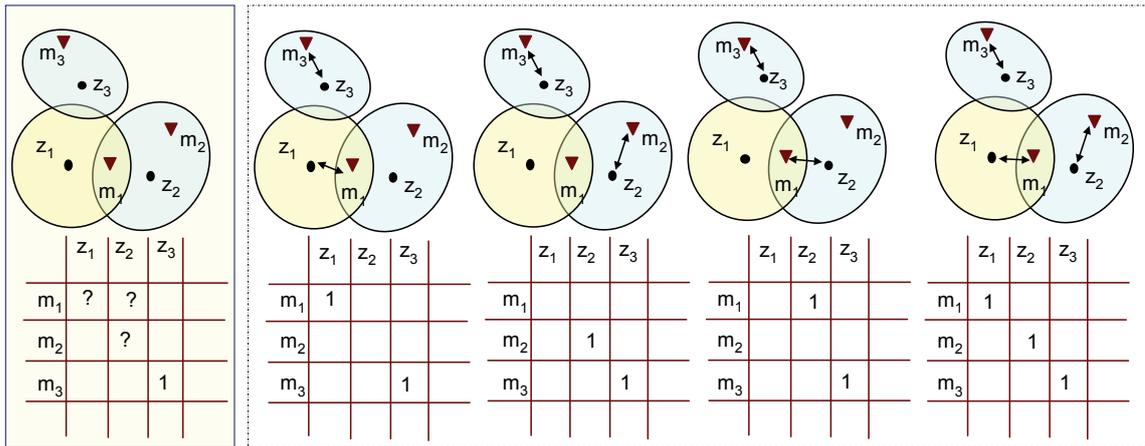
Der JPDA-*Ansatz* ist eine Ergänzung des PDA-Ansatzes und zieht zur Berechnung der Wahrscheinlichkeiten alle Beobachtungen und alle Objekte/Prädiktionen heran. Während bei dem PDA-Ansatz nur die Messungen im Suchbereich des jeweiligen Objekts betrachtet werden, wird auch die mögliche Zugehörigkeit zu anderen Objekten einbezogen. Die Assoziationswahrscheinlichkeit wird über alle Objektprädiktionen und Messungen berechnet. Dabei wird vorausgesetzt, dass die Anzahl der Objekte bekannt ist und jedes Objekt bestenfalls eine Messung verursacht. Analog zum PDA-Ansatz ist hier ebenfalls die Formulierung einer *Ähnlichkeit* entsprechend eines Beobachtungsmodells notwendig. Dadurch werden die Beziehungen zwischen Objekten, ihren Einzugsbereichen und den darin enthaltenen Messungen bewertet. Dabei setzt sich die Wahrscheinlichkeit  $p(H^{(i)})$  der  $i$ -ten Hypothese  $H^{(i)}$  nun aus den Ähnlichkeiten der zugehörigen Zuordnungen zusammen. Der Aktualisierungsschritt wird analog zum PDA mit dem gewichteten Innovationsvektor realisiert. Allerdings ist der JPDA-Ansatz sehr aufwändig, so dass verschiedene suboptimale Varianten existieren [Roecker 93].

Um dies zu verdeutlichen, wird die Konfliktsituation in der Abbildung (4-3-3) betrachtet, bei der zwei Messungen ( $m_1$  oder  $m_2$ ) einer Prädiktion ( $z_2$ ) zu zuordnen sind. Aufgrund der annähernd gleichen Abstände der beiden Messungen zur zweiten Prädiktion würden bei Anwendung der PDA beide Messwerte in etwa gleich stark in den gewichteten Innovationsvektor und damit in die Zustandsaktualisierung eingehen [Stüker 04]. Bei der JPDA wird hingegen das Gewicht in Abhängigkeit von der Wahrscheinlichkeit, dass diese zur ersten Prädiktion gehört, reduziert [Stüker 04].

---

<sup>86</sup> Die Innovationen stammen begrifflich von der Kalman-Filter Technik und geben die Diskrepanz zwischen Zuordnungen wieder.

#### 4. OBJEKTVERFOLGUNG AUF DER BASIS VON DATENASSOZIATIONEN



**Abbildung 4-3-3:** Beispiel einer Konfliktsituation (links), die durch den JPDA Ansatz aufgelöst wird. Die Messung  $m_1$  lässt sich anhand der Suchbereiche beider Prädiktionen  $z_1$  und  $z_2$  zuordnen. Die Messung  $m_3$  ist eindeutig. Die Beziehungen zwischen den Prädiktionen und Messungen können in einer Matrix erfasst werden.

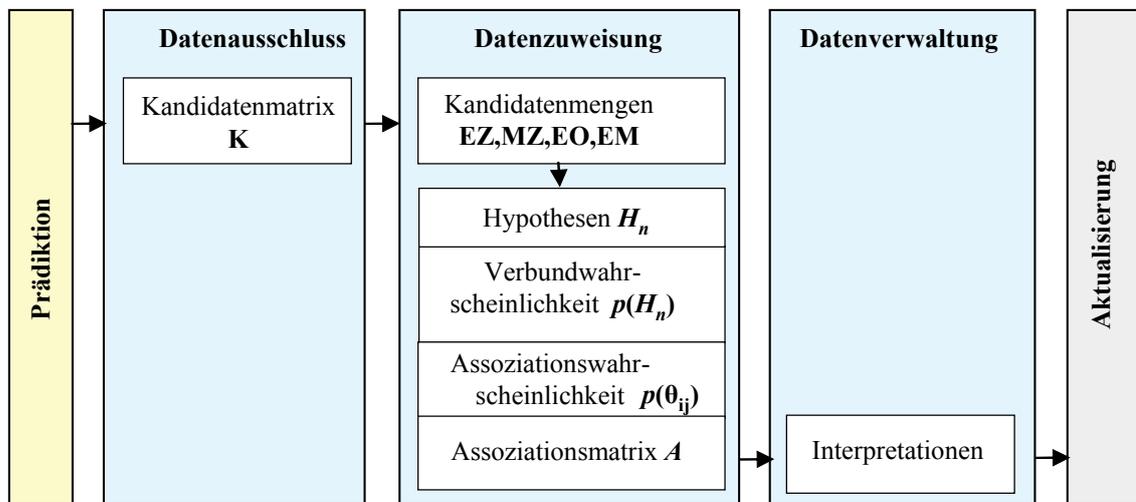
Abbildung 4-3-3 rechts zeigt die konfliktfreien Zuweisungen. Daraus ergibt sich, dass nicht alle Prädiktionen eine Messung zugeordnet bekommen und nicht alle Messungen Objekte zugeordnet werden (Geisterziele, Rauschen). Daraus erkennt man, dass mit dem Einsatz der JPDA die Anwendung der Objektverfolgung jedoch auf Szenen mit fester und bekannter Anzahl von Objekten begrenzt bleibt.

In einem anderen Ansatz verwenden Cox u. Hingorani [Cox 94] eine *Multi-Hypothesen-Technik* (MHT) zur Datenassoziation, die die Handhabung einer variablen Objektanzahl ermöglicht. Dabei wird nicht nur der aktuelle Zeitschritt verwendet, um die Assoziationsmöglichkeiten zu bewerten. Unter Einbeziehung der bereits vorhandenen Zeitschritte wird in der Zuordnung die Historie mit ausgewertet. Daher werden die Assoziationshypothesen baumartig aufgebaut. Aus den vorhandenen Objekten und den Kandidaten nach dem *Gating*<sup>87</sup> [Cox 94] werden alle kombinierbaren Hypothesen aufgestellt. Die Hypothesen sind analog wie bei dem JPDA-Ansatz alle durchführbaren Kombinationen aus sich ausschließenden Zuordnungspaaren. Mit jedem neuen Zeitschritt werden die Hypothesen erweitert und um neue ergänzt [Homberg 07, Blackman 99]. Damit wird eine stets größer werdende Menge von Hypothesen generiert, die alle mit der Zeit möglichen Assoziationen enthält und die wahrscheinlichste liefern kann. Hieraus ist der stark wachsende Aufwand mit jedem weiteren Zeitschritt und jedem hinzukommenden Objekt ersichtlich. Mit Schwellwerten und Beschränkungen wird in Weiterentwicklungen versucht, die Anzahl der Hypothesen nur begrenzt wachsen zu lassen. Dennoch steht dem Vorteil des zeitlichen Bezugs ein enormer Rechenaufwand gegenüber.

Ausgehend von dieser kurzen Beschreibung könnten für die Multi-Objektverfolgung bei dem JPDA-Ansatz und der Multi-Hypothesentechnik (MHT) interessante Eigenschaften festgestellt werden. Während der JPDA-Ansatz alle Zuordnungsmöglichkeiten der jeweiligen Konfliktsituation beachtet und vergleicht, bezieht die MHT Informationen aus den vergangenen Zeitschritten mit ein und verwendet alte Zuordnungen mit in der Bewertung der Hypothesen. Dadurch kann die Abschwächung der unwahrscheinlichen und Stärkung der wahrscheinlichen Hypothesen erzielt werden. Diese Gesichtspunkte werden bei den hier aufgebauten Datenassoziationen vereint, indem die gängige,

<sup>87</sup> Die Gatingmatrix stellt die Beziehungen zwischen den Prädiktionen und Messungen dar.

zweistufige Assoziation mit *Datenausschluss* und *Datenzuweisung* um eine dritte Stufe modifiziert und ergänzt wird: die *Datenverwaltung*<sup>88</sup> (s. Abbildung 4-3-1). Während im Datenausschluss erste Zuordnungspaare aufgrund ihrer Lage zueinander ausgeschlossen werden, werden in der Datenzuweisung die Objektkonstellationen einer Szene abstrakt in Mengen erfasst und in möglichst einfache Zuordnungsmöglichkeiten aufgelöst. Die Datenverwaltung kann anschließend die verbliebenen Zuordnungsmöglichkeiten interpretieren und kann die Entwicklungseigenschaften des Filtersystems wie Bewegungsmodelle oder Partikelfilter-Initialisierung einstellen. Dieser Schritt gibt jedoch Raum, die Fluktuation der Szenenobjekte zu organisieren und im Weiteren die Vorteile der beiden Assoziationsverfahren zu vereinen. Im Folgenden wird der gesamte Schritt der Assoziation in drei Stufen aufgebaut: Datenausschluss, Datenzuweisung und Datenverwaltung (s. Abb. 4-3-4).



**Abbildung 4-3-4:** Eine schematische Darstellung der vorgeschlagenen Datenassoziation, die sich aus Datenausschluss, Datenzuweisung und Datenverwaltung zusammensetzt. Im Datenausschluss werden mögliche Zuordnungspaare ermittelt. In der Datenzuweisung erfolgen eine Zerlegung der Szene in Kandidatenmengen und der weitere Ausschluss unwahrscheinlicher Zuordnungspaare. In der Datenverwaltung werden die verbliebenen Zuordnungspaare interpretiert und die repräsentierenden Partikelfilter auf das Ergebnis eingestellt.

##### 4.3.4.1 Datenausschluss

Im Prinzip werden im Assoziationsschritt Hypothesen bewertet, die alle gültigen Kombinationen aus allen Objekt-Messung-Zuordnungen wiedergeben. Allerdings wächst die Anzahl der Hypothesen mit jedem Zuordnungskandidaten exponentiell. Um die Anzahl und somit den Rechenaufwand einzuschränken, sollen im Voraus die Zuordnungsmöglichkeiten begrenzt werden.

Intuitiv ist ersichtlich, dass Messungen, die zu weit entfernt von einer Prädiktion liegen, nicht mehr zuzuordnen sind<sup>89</sup>. Dafür werden Konstellationen auf vereinfachende Repräsentationen und somit der Ausschluss von Zuordnungsmöglichkeiten untersucht. Dies wird in der Regel mit Hilfe von Einzugsbereichen, die um die Objektprädiktion platziert werden, erreicht. Dabei werden die Messungen, die im Bereich einer Prädiktion liegen, entweder als Zuordnungskandidat mit in den weiteren Assoziationsschritten

<sup>88</sup> Ein Verwaltungsschritt wurde in den bekannten Datenassoziationsansätzen vernachlässigt.

<sup>89</sup> Dieser Vorgang wird in der Literatur auch Gating genannt.

eingehen, andernfalls werden sie ausgeschlossen. Bei den ausgeschlossenen Messungen kann es sich um Ausreißer und Geisterziele handeln, die die Zustandsschätzung des betrachteten Objektes sogar verschlechtern können. Im Gegenteil dazu werden die gebildeten Zuordnungspaare aus der Messung und der Prädiktion als richtig angenommen. Somit wird möglicherweise eine falsche Zuordnung toleriert. Aus dem Verbleib bilden alle Prädiktionen und Messungen -für die diese Beziehung gilt- jeweils ein Zuordnungspaar. Hierfür werden zunächst alle Objektprädiktionen auf Messungen in ihrem Suchbereich<sup>90</sup> überprüft. Daher müssen die Suchbereiche einerseits groß genug definiert werden, um die korrekten Messwerte einzufangen und andererseits klein genug, um neue Objekte zu erfassen.

Grundsätzlich gibt es auch Suchbereiche, die statistische Abstandsmaße benutzen, um neben absoluten Abweichungen zwischen Prädiktion und Messung auch Unsicherheiten einzubeziehen. Dafür seien rechteckige und elliptische Suchbereiche genannt [Blackman 99]. Mit diesen Ansätzen können komplizierte Zustandsbeschreibungen verglichen werden. Für die hier wegen der Verallgemeinerung einfach gehaltene Zustandsbeschreibung, ist die Formulierung der Suchbereiche durch einfache Abstandsmaße ausreichend. Die einfachste Variante, einen solchen Suchbereich zu formulieren, ist die Verwendung von Abstandsmaßen wie die euklidische Distanz oder Mahalanobis-Distanz<sup>91</sup> [Aziz 99, Bar-Shalom 88].

Der Einfachheit halber wird hier die euklidische Distanz verwendet. Dadurch wird für jede Objektprädiktion  $z$  ein Suchbereich  $R$  um die durchschnittliche Position  $p$  der prädierten Partikel aufgespannt. Die Grenze wird auf die größte Abweichung der Partikel zu  $p$  gesetzt plus einen Sicherheitsrand von 5 Pixel, um zum einen Mindestbereich zu garantieren, falls die Partikel einen zu kleinen Bereich abdecken, und zum anderen Bewegungsänderungen zu erfassen. Liegt eine Messung  $m$  im Suchbereich eines Objekts, ist sie Kandidat für die jeweilige Zuordnung der  $O$ -Objekte. Für die Weiterverarbeitung werden die Ergebnisse in die Kandidatenmatrix  $K=(k_{ij})$  der Größe  $O \times M$  eingetragen:

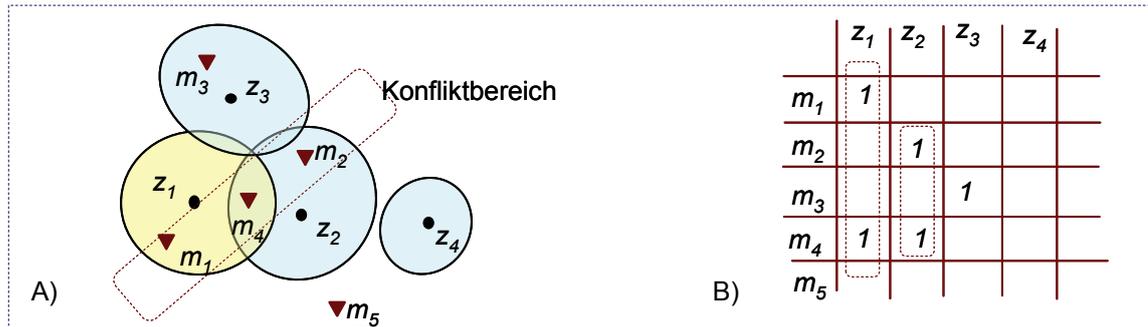
$$k_{ij} = \begin{cases} 1, & \text{falls } m_{(j)} \in R(z_{(i)}), i = 1, \dots, O \text{ und } j = 1, \dots, M \\ 0, & \text{sonst} \end{cases} \quad (4-3-7b)$$

In der Kandidatenmatrix sind somit alle Beziehungen und Grenzen zwischen allen Objektprädiktionen  $z_{(i)}$  und Messungen  $m_{(j)}$  enthalten und für weitere Schritte auslesbar. Beziehungen zwischen Objekten und Messungen werden dabei mit den Einträgen  $k_{ij} = 1$  gekennzeichnet. Objekte und Messungen ohne gemeinsame Einträge bilden kein Kandidatenpaar (Abb. 4-3-5). Liegt aber zwischen  $z_{(i)}$  mit  $m_{(j)}$  eine Konfliktsituation vor, die über den Datenausschluss nicht aufgelöst werden konnte, werden solche Situationen in der Datenzuweisung weiter analysiert und vereinfacht. Abbildung 4-3-5 zeigt ein Beispiel für diese Beziehungen zwischen Prädiktionen und Messungen. Dort ist zu erkennen, dass die Beziehung von  $z_{(3)}$  und  $m_{(3)}$  eindeutig ist und deshalb auf eine konfliktfreie Situation reduziert werden kann. Im Gegenteil dazu liegen mehrere

<sup>90</sup> Der Suchbereich bezieht sich auf die Position

<sup>91</sup> Die Mahalanobis-Distanz stellt die Summe der  $n$  quadrierten Komponenten des Residuenvektors dar, die mit den Diagonalelementen der Kovarianzmatrix gewichtet werden. Bei dem Ansatz in [Bar-Shalom 88] wird der Messvektor  $m$  als zuverlässig betrachtet, wenn die korrespondierende Mahalanobis-Distanz innerhalb eines Toleranzbereiches liegt und ansonsten verworfen. Die Festlegung des Schwellwerts dafür erfolgt ausgehend von der  $\chi^2$ -Verteilung der Mahalanobis-Distanz [Mecke 99].

Konfliktsituationen von  $z_{(1)}$  und  $z_{(2)}$  vor, die über die Datenzuweisung gelöst werden könnten.



**Abbildung 4-3-5:** Ermittlung der Kandidatenmengen. Die Beziehungen zwischen den Prädiktionen  $z_1 - z_4$  und Messungen  $m_1 - m_5$  (A) können in einer Matrix (B) erfasst werden. Auf der abstrakten Ebene der Matrix sind zusammenhängende Zuordnungspaare in unabhängige Paarmengen zerlegbar. Zusammenhänge sind durch Mehrfachbelegung von Spalten oder Zeilen feststellbar und durch unterschiedliche gestrichelte Einfärbung gekennzeichnet.

##### 4.3.4.2 Datenzuweisungen

In der Datenzuweisung erfolgt eine weitere Vereinfachung der Konfliktsituationen durch die Anwendung des Prinzips des JPDA-Ansatzes, das aber nicht als determiniertes Ergebnis verwendet wird. Das resultierende Ergebnis des JPDA-Ansatzes wäre eine weitere Gewichtung, die für die genaue Weiterentwicklung der Partikelfilter anhand der Messungen notwendig ist. Dieser Schritt bindet also die Betrachtung und Bewertung von miteinander konkurrierenden Zuordnungsmöglichkeiten und durchläuft dazu mehrere Verarbeitungsschritte (Abb. 4-3-4). Hierbei wird erstens die aufgestellte Kandidatenmatrix in Kandidatenmengen gemäß ihrer räumlichen Abhängigkeit gemäß des definierten Suchbereiches  $R$  (s. Datenausschluss) geteilt. Anschließend werden aus den Kandidatenmengen Zuordnungshypothesen aufgestellt, die durch die JPDA bewertet werden. Mit den resultierenden Assoziationswahrscheinlichkeiten aus der JPDA wird die Assoziationsmatrix als Ergebnisse der Datenzuweisung erzeugt, die die Grundlage für die Interpretation im Verwaltungsschritt ist (Abb.4-3-4).

### 1). Kandidatenmengen

Die Eingabe für die folgenden Berechnungen ist die resultierende Kandidatenmatrix  $K$  aus dem Datenausschlusssschritt, welche die Zusammenhänge zwischen den Objekten und den Messungen durch gemeinsame Kandidaten gemäß der festgelegten Suchbereiche beinhaltet, ebenso wie die Grenzen zwischen Kandidatenpaaren. Diese Zusammenhänge werden in der Matrix  $K$  durch gemeinsame Spalten und Zeilen repräsentiert.

In einer Konfliktsituation können für eine Objektprädiktion mehrere Messungen aus den überlappten Suchbereichen in der entsprechenden Spalte eingetragen werden. Diese Konstellation verbindet die beiden Objektprädiktionen/Messungen (Abb. 4-3-5), die in der Matrix durch die gemeinsam besetzte Zeile der Messung deutlich herauszufinden ist. Die anderen Gruppen hingegen weisen mit dieser keine gemeinsam besetzte Zeile oder Spalte auf. Diese Eigenschaften sollen verwendet werden, um die Szene zu teilen und die Einträge der Kandidatenmatrix in kleinere Kandidatenmengen abzubilden. Mit der Unterteilung von zusammenhängenden Zuordnungspaaren  $\theta_{ij}$  zwischen Objektprädiktion  $z_{(i)}$  und Messung  $m_{(j)}$  mit  $k_{ij}=1$  wird eine klare Trennung zwischen unabhängigen

#### 4. OBJEKTVERFOLGUNG AUF DER BASIS VON DATENASSOZIATIONEN

Konstellationen erreicht. Dabei sind die einzelnen Objekte oder Messungen in der Kandidatenmatrix durch leere Spalten bzw. leere Zeilen gekennzeichnet. Bei Vernachlässigung der einzelnen Elemente würden diese verloren gehen und die angestrebte automatische Organisation der Objektfluktuation in der Szene ausgeschlossen. Daher werden diese auch als Zuordnungspaar ausgewiesen:  $\theta_{i0}$  und  $\theta_{0j}$  kennzeichnen, dass das  $i$ -te Objekt bzw. die  $j$ -te Messung ohne Zuordnungskandidaten ist. Die Zusammenhänge der Bildszene in der Kandidatenmatrix werden durch Nachbarschaftsbeziehungen der Einträge repräsentiert und können in vier verschiedene Mengen abgebildet werden.

- **Eindeutige Zuordnungen (EZ)** zwischen einem Objekt und einer Messung sind durch einen einzelnen Eintrag in sich kreuzenden Zeilen und Spalten gekennzeichnet, wobei die Zeilen und Spalten jeweils keine weiteren Einträge beinhalten. Diese eindeutigen Zuordnungen werden jeweils in eine 1-elementige Menge  $EZ_l$  abgebildet.

$$EZ_l = \left\{ (\theta_{ij}) \mid \sum_{j=1}^M k_{ij} = 1 \wedge \sum_{i=1}^O k_{ij} = 1 \right\} \quad (4-3-8a)$$

- Leere Spalten deuten auf **einzelne Objekte (EO)** hin, denen im Datenausschluss keine Messung zugeordnet werden konnte. Bei der Vernachlässigung der einzelnen Elemente würden diese verloren gehen und die angestrebte automatische Organisation der Objektfluktuation in der Szene ausgeschlossen. Daher werden auch einzelne Objekte als Zuordnungspaar ausgewiesen:  $\theta_{ij}$  kennzeichnet das  $i$ -te Objekt ohne Zuordnungskandidaten. Einzelne Objekte werden jeweils in eine 1-elementige Menge  $EO_l$  abgebildet.

$$EO_l = \left\{ (\theta_{ij}) \mid \sum_{i=1}^O k_{ij} = 0 \right\} \quad (4-3-8b)$$

- **Einzelne Messungen (EM)** sind in der Kandidatenmatrix über leere Zeilen auslesbar. Diese einzelnen Elemente dürfen wie die einzelnen Objekte nicht vernachlässigt werden und sind folgend mit  $\theta_{0j}$  für die  $j$ -te Messung ohne Zuordnungskandidaten aufgeführt. Die einzelnen Messungen werden ebenfalls jeweils in eine 1-elementige Menge  $EM_l$  abgebildet.

$$EM_l = \left\{ (\theta_{ij}) \mid \sum_{j=1}^M k_{ij} = 0 \right\} \quad (4-3-8c)$$

- **Mehrdeutige Zuordnungen (MZ)** weisen mehrfach und gemeinsam besetzte Zeilen und Spalten auf. Die Zuordnungspaare einer mehrdeutigen Situation werden jeweils in einer Menge  $MZ_l$  zusammengefasst.

$$MZ_l = \left\{ \theta_{ij} \mid \sum_{j=1}^M k_{ij} > 1 \vee \sum_{i=1}^O k_{ij} > 1 \right\} \quad (4-3-8d)$$

Für die Mengen  $MZ_l$  gilt außerdem: Die Zuordnungspaare  $\theta_{ij}$  und  $\theta_{mn}$  gehören zu einer Kandidatenmenge, wenn für ihre Einträge  $k_{ij}$  und  $k_{mn}$  in der Kandidatenmatrix gilt:

1.  $k_{ij}$  und  $k_{mn}$  sind direkte Nachbarn:  $(i \neq m \wedge j = n) \vee (i = m \wedge j \neq n)$  oder
2.  $k_{ij}$  und  $k_{mn}$  sind über Folge direkter Nachbarn miteinander verbunden.

$l$  indiziert jeweils nach der Anzahl der entsprechenden Mengen. Im Ergebnis repräsentiert nun jede Kandidatenmenge eine Teilszene der gesamten Bildszenen und ist im Weiteren separat analysierbar. Die Zusammenhänge und Grenzen zwischen den Objekten und den Messungen durch gemeinsame Kandidaten und Suchbereiche sind dabei erhalten geblieben. Die Mengen  $MZ_l$  repräsentieren mehrdeutige Teilszenen. Im Folgenden werden diese weiter analysiert und nach dem Prinzip der JPDA vereinfacht.

### 2). Hypothesengenerierung

Mit der Zerlegung der Zuordnungspaare  $\theta_{ij}$  in die Kandidatenmengen ( $EZ_l$ ,  $EO_l$ ,  $EM_l$  und  $MZ_l$ ) liegen die Zusammenhänge der Teilszenen eindeutig vor. Sie haben das Ziel, die Zuordnungsmöglichkeiten innerhalb einer Teilszene weiter einzugrenzen und zu vereinfachen. Dies erfolgt über die Hypothesenaufstellung und deren Bewertung anhand des JPDA-Ansatzes, wodurch die Verhältnisse zwischen wahrscheinlichen und unwahrscheinlichen Zuordnungspaaren verstärkt werden. Dabei beschreiben die Hypothesen alle Zuweisungskombinationen, bei denen eine Beobachtung maximal einer Prädiktion zugeordnet wird. Sie führt dazu, dass sie konfliktfrei sind. Ausgegangen wird von den zuvor aufgestellten Kandidatenmengen. Hierfür werden die Zuordnungspaare innerhalb der Kandidatenmengen anhand ihrer Ähnlichkeiten  $SM_{Obj}$  (Gl. 4-3-6c) bewertet. Da diejenigen Mengen aus einzelnen Objekten ( $EO_l$ ), einzelnen Messungen ( $EM_l$ ) und eindeutiger Zuordnung ( $EZ_l$ ) konfliktfreie Zuordnungspaare darstellen und somit eindeutig sind, werden nur noch die Kandidatenmengen  $MZ_l$  der Konfliktsituationen in Betracht gezogen und im folgenden analysiert.

Zunächst werden die Hypothesen  $H_n$  aus den Zuordnungspaaren einer Menge aufgestellt. Hierfür werden für alle Objekte Nullzuordnungen eingefügt, damit die Hypothesenbildung gewährleistet ist. Das heißt, dass für jedes Objekt  $i$  ein Zuordnungspaar  $\theta_{i0}$  in die betrachtete Kandidatenmenge eingefügt wird, um in den Mehrdeutigkeiten eine mögliche Nichtzuordnung zuzulassen, falls das Objekt von keiner der Messungen verursacht wurde. Mit den Nullzuordnungen und den Beschränkungen durch die Suchbereiche ergeben sich  $N$ -Kombinationsmöglichkeiten. Diese entsprechen den Hypothesen. Jede Hypothese enthält demnach eine zulässige Kombination aus den Zuordnungspaaren. Dabei werden keine Mehrfachzuweisungen pro Objekt vorgenommen. Für die Messungen hingegen sind Mehrfachzuweisungen möglich. So ergeben sich  $N$ -Hypothesen  $H_n$  nach der Vorschrift:

$$H_n = \{ \theta_{ij} \mid \forall \theta_{ij} \wedge \theta_{kl} \text{ mit } i \neq k \} \quad (4-3-9a)$$

Mit dieser Vorschrift wird davon ausgegangen, dass ein Objekt höchstens eine Messung verursacht, eine Messung aber von mehr als einem Objekt herrühren kann. Dies ist in Situationen von Verschmelzungen der Blobs der Fall. Aus der Abbildung 4-3-5 stellen die Objekte ( $z_1$  und  $z_2$ ) und die Messungen ( $m_1$ ,  $m_4$  und  $m_2$ ) eine mehrdeutige Kandidatenmenge, für die die Hypothesen aufgestellt werden. Mit den eingefügten Nullzuordnungen  $\theta_{10}$  und  $\theta_{20}$  ist die Menge  $\{ \theta_{10}, \theta_{11}, \theta_{14}, \theta_{20}, \theta_{22}, \theta_{24} \}$  die Kandidatenmengen der Konfliktszene. Nach Vorschrift (4-3-9a) ergeben sich die in Tabelle 4-3-1 möglichen aufgelisteten Hypothesen.

#### 4. OBJEKTVERFOLGUNG AUF DER BASIS VON DATENASSOZIATIONEN

**Tabelle 4-3-1:** Hypothesen  $H_n$  aller N Zuordnungskombinationen der Kandidatenmenge aus dem Beispiel der Abbildung 4-3-5 nach der Vorschrift (4-3-9a).

	$H_1$	$H_2$	$H_3$	$H_4$	$H_5$	$H_6$	$H_7$	$H_8$	$H_9$
$z_1$	$\theta_{10}$	$\theta_{10}$	$\theta_{10}$	$\theta_{11}$	$\theta_{11}$	$\theta_{11}$	$\theta_{14}$	$\theta_{14}$	$\theta_{14}$
$z_2$	$\theta_{20}$	$\theta_{22}$	$\theta_{24}$	$\theta_{20}$	$\theta_{24}$	$\theta_{22}$	$\theta_{22}$	$\theta_{24}$	$\theta_{20}$

### 3). Assoziationsmatrix

Die Hypothesen jeder Kandidatenmenge ( $H_n$ ) sind nun die Grundlage für die Berechnung der Verbundwahrscheinlichkeit, die wiederum die Grundlage für die Bestimmung der Assoziationswahrscheinlichkeiten darstellt, die durch die Konkurrenz der Zuordnungsmöglichkeiten in einer Menge verstärkt oder abgeschwächt werden. Hierfür sind die Einzelwahrscheinlichkeiten der Zuordnungs-paare erforderlich, die sich aus der Berechnung der Ähnlichkeiten  $SM_{Obj}$  (Gleichung 4-3-6c) zwischen den Objekten und Messungen der Zuordnungs-paare  $\theta_{ij}$  ergeben. Diese Ähnlichkeit liefert Werte zwischen  $[0,1]$ , so dass diese hier als Einzelwahrscheinlichkeiten  $L(\theta_{ij}) = SM_{Obj}$  angewendet werden können.

Aus den Einzelwahrscheinlichkeiten ergibt sich die gemeinsame Wahrscheinlichkeit einer Hypothese. Daher lassen sich die Einzelwahrscheinlichkeiten der Nullzuordnungen  $\theta_{i,0}$  für jedes Objekt aus dem Komplement der durchschnittlichen Einzelwahrscheinlichkeiten des jeweiligen Objekts herleiten. Folgende Berechnungen folgen nun gemäß dem Prinzip der JPDA [Homberg 07, Helmick 00]:

- Es werden zunächst die Verbundwahrscheinlichkeiten  $p(H_n)$  für alle Hypothesen  $H_n$  ermittelt. Diese setzen sich jeweils aus den Einzelwahrscheinlichkeiten  $L(\theta_{ij})$  der Zuordnungen  $\theta_{ij}$  zusammen:

$$p(H_n) = \prod_{\theta_{ij} \in H_n} SM_{Obj}(\theta_{ij}) = \prod_{\theta_{ij} \in H_n} L(\theta_{ij}) \quad (4-3-9b)$$

In der Verbundwahrscheinlichkeit  $p(H_n)$  ist damit jede Zuordnungskombination, die in der Hypothese  $H_n$  enthalten ist, bewertet worden.

- Die Assoziationswahrscheinlichkeiten der Zuordnungen  $\theta_{ij}$  ergeben sich nun aus der Summe der Verbundwahrscheinlichkeiten aller Hypothesen, die diese Zuordnung unterstützen.

$$p(\theta_{ij}) = \sum_{\forall H_n | \theta_{ij} \in H_n} p(H_n) \quad (4-3-9c)$$

Für Herleitungen und Details sei an dieser Stelle auf [Helmick 00] verwiesen. Die Assoziationswahrscheinlichkeiten der Zuordnungs-paare weisen nun stärker differenzierbare Verhältnisse auf als die der Einzelwahrscheinlichkeiten. Hypothesen mit wahrscheinlichen Zuordnungskombinationen erreichen eine starke Verbundwahrscheinlichkeit, während die mit unwahrscheinlichen Zuordnungskombinationen kleine Werte erhalten. In der Hypothesensumme wird dann die Assoziationswahrscheinlichkeit einer unwahrscheinlichen Zuordnung erheblich reduziert. Im Gegenzug wird dadurch eine konkurrierende Zuordnungsmöglichkeit wahrscheinlicher, die zuvor eine ähnliche Einzelwahrscheinlichkeit aufwies, die in ihren Kombinationen aber andere starke Zuordnungsmöglichkeiten zulässt.

#### 4. OBJEKTVERFOLGUNG AUF DER BASIS VON DATENASSOZIATIONEN

Allgemein liegen nach den vorangegangenen Schritten nun alle Kandidatenmengen ( $EZ_l$ ,  $EO_l$ ,  $EM_l$  und  $MZ_l$ ) und die Assoziationswahrscheinlichkeiten aller Zuordnungs-paare aus den Mengen  $MZ_l$  vor. Für diese gegebenen Beobachtungen und die bekannten Prädiktionen wird eine Assoziationsmatrix  $A$  aufgestellt mit dem Reihenindex  $i$  und Spaltenindex  $j$ , deren Elemente eine skalare Assoziationsmaßzahl darstellen, die die Ähnlichkeiten zwischen der  $i$ -ten Beobachtung und dem  $j$ -ten Objekt quantifizieren. Mit dieser Assoziationsmatrix  $A$  sollen alle Zuordnungsmöglichkeiten an den Verwaltungsschritt weitergegeben werden. Hierzu werden alle Zuordnungs-paare, auch die der eindeutigen Mengen, in die Matrix  $A=(a_{ij})$  eingetragen. Die Einträge repräsentieren hier wieder die Objektkonstellationen mit ihren Zusammenhängen, wie sie in der Kandidatenmatrix vorlagen (s. Gl. 4-3-9d).

Die mehrdeutigen Zuordnungsmöglichkeiten sind hierbei mit ihren Assoziationswahrscheinlichkeiten eingetragen, aber noch nicht weiter aufgelöst. Da für die Partikelentwicklung die Messungen als Grundlage für die Aktualisierung dienen, ist für die angestrebte Stabilität jedoch eine eindeutige, nicht anteilige Zuordnung erforderlich. Der Inhalt der Matrix wird wie folgt vereinfacht.

$$a_{ij} = \begin{cases} 1, & \text{falls } \theta_{ij} \in \{EZ_l, EO_l, EM_l\} \\ p(\theta_{ij}), & \text{falls } \theta_{ij} \in \{MZ_l\} \\ 0, & \text{sonst} \end{cases} \quad (4-3-9d)$$

$i = 1, \dots, O \quad \text{und} \quad j = 1, \dots, M$

Im Gegensatz zur ursprünglichen Idee des JPDA-Ansatzes, bei der die Werte der Assoziationswahrscheinlichkeiten als eine Gewichtung für den Aktualisierungsschritt gedacht sind, werden hier die Werte als Empfehlung aufgefasst. Die Ergebnisse des JPDA-Ansatzes liefern für alle Zuordnungskandidaten Wahrscheinlichkeitswerte, die durch konkurrierende Zuordnungs-paare verstärkt oder abgeschwächt wurden. Hier werden nun die Zuordnungs-paare beibehalten, die für das jeweilige Objekt die größte Assoziationswahrscheinlichkeit aufweisen. Die Zuordnungs-paare mit den geringeren Wahrscheinlichkeitswerten werden verworfen und im Gegensatz zu ursprünglichen JPDA-Anwendung nicht weiter betrachtet. Dies wird erreicht, indem für jedes Objekt der größte Eintrag seiner Spalte mit Eins und die anderen mit Null ersetzt werden [Al-Hamadi 07, Homberg 07, Al-Hamadi 08]:

$$a_{ij} = \begin{cases} 1, & \text{falls } a_{ij} = \max(a_{i1:M}) \\ 0, & \text{sonst} \end{cases} \quad (4-3-9e)$$

$i = 1, \dots, O \quad \text{und} \quad j = 1, \dots, M$

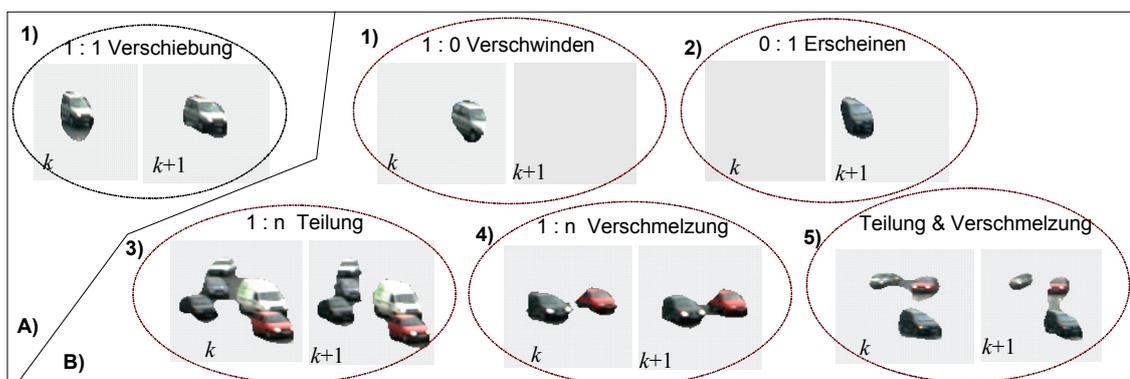
So können weitere Zuordnungsmöglichkeiten zwischen den  $O$  Objekten und den  $M$  Messungen aufgelöst werden. Eine Garantie für ein vollständiges Auflösen von Mehrdeutigkeiten ist jedoch nicht gegeben. Insbesondere bei unterschiedlicher Anzahl von Objekten und Messungen, wie im Fall verschmelzender Objekte, wird eine solche Auflösung nicht erfolgen. Diese Konstellationen werden im folgenden Schritt der Datenverwaltung interpretiert und die zugehörigen Partikelfilter daran angepasst. Mit der Assoziationsmatrix  $A$  werden dafür alle verbleibenden Zuordnungsmöglichkeiten an den Datenverwaltungsschritt weitergegeben.

### 4.3.4.3 Datenverwaltung

Ziel des Datenverwaltungsschritts ist die Adaption des Partikelfiltersystems an die Bildsituation, so dass die Partikelfilter ihre Partikel aktualisieren und für den nächsten Zeitschritt propagieren können. Die Anpassung an die Bildsituation benötigt eine weitere Vorgehensweise bezüglich der Organisation der Objektfluktuation, die die Initialisierung und den Verwurf von Partikelfiltern gemäß den Bildgegebenheiten ausführt. Zudem sollen den Objekten und ihren Partikelfiltern Messungen - wenn vorhanden - eindeutig zugewiesen werden, damit eine Bewertungsgrundlage für die Aktualisierung des Partikelfilters vorliegt. Entsprechend müssen Verhaltensvorschriften für Partikelfilter ohne eindeutige Zuordnung wie beispielsweise in Konfliktsituationen gegeben werden, um die Partikel weiter entwickeln zu können [Al-Hamadi 07a, 08a].

Als Ausgangspunkt dafür steht hier die berechnete Assoziationsmatrix  $A$ , die alle nach der Datenzuweisung verbliebenen Zuordnungspaare enthält. Zusammenhänge und Grenzen zwischen den Zuordnungspaaren, die aus der Gesamtszene stammen, sind wie auch zuvor in der Kandidatenmatrix erhalten geblieben. Das heißt, wie bei der Zerlegung in die Mengen  $EZ_l$ ,  $EO_l$ ,  $EM_l$  und  $MZ_l$  nach den Vorschriften (4-3-8/a,b,c,d) ist der Assoziationsmatrix zu entnehmen, ob und mit welchen Objekten und Messungen ein Objekt in Konflikt liegt. Mit dieser Auswertung der Assoziationsmatrix können die enthaltenen Zuordnungspaare jeweils einer der vier beschriebenen Gruppen zugeordnet werden. Diese Differenzierung der Zuordnungspaare stellt nun die Grundlage für den Interpretations- und Verwaltungsschritt dar. Zunächst seien die Überlegungen aus Abschnitt 4.3.4.2 und der Abbildung 4-3-5a über die Interpretierbarkeit von den Objektkonstellationen der *vier Gruppen* ohne zusätzliche Informationen aufgegriffen:

- 1). *Eindeutige Zuordnung* umfasst alle Situationen, in denen keine Konfliktsituationen entstehen (s. Abb. 4-3-5a/A).
- 2). *Mehrdeutige Zuordnung* beinhaltet Situationen, bei denen die Zuweisung von Objekten und Messungen in eindeutiger Weise nicht möglich ist. Ursachen dafür sind die Verschmelzung und Teilung von Objekten aufgrund einer Teilverdeckung.
- 3). *Einzelnes Objekt* erfasst Objekte, die keine Messung in der Nähe ihrer Prädiktion wie bspw. beim Verlassen der Szene oder stehen bleibenden Objekten aufgrund Stau haben.
- 4). *Einzelne Messung* wird aufgrund einer Bewegung (z.B. Erscheinen von einem Objekt in Abbildung 4-3-5a/B) oder auch Störungen verursacht, ohne ein Objekt vorliegt.



**Abbildung 4-3-5a:** Die Objektkonstellationen bei der eindeutigen Zuordnung A) und Objektinteraktionen in Konfliktsituationen B) in realen Verkehrsszenen.

#### 4. OBJEKTVERFOLGUNG AUF DER BASIS VON DATENASSOZIATIONEN

Für die Analyse solcher Objektkonstellationen verwendet die menschliche Wahrnehmung dafür Kontextinformationen wie beispielsweise die Ausrichtung und Texturen von Objekten, die nach der Merkmalsextraktion und Zustandsbeschreibung aufgrund der automatischen Bewegungssegmentierung hier nicht zur Verfügung stehen. Stattdessen kann die Idee und gleichzeitig der Vorteil der Multi-Hypothesen-Technik ausgenutzt werden: das Verwenden zusätzlicher *zeitlicher* Informationen. Daher zieht diese Technik alle Zuordnungsmöglichkeiten der vorangegangenen Zeitschritte in die Hypothesenbildung mit ein, was allerdings den Aufwand exponentiell wachsen lässt. Um dieses Prinzip optimal auszunutzen, wird hier für jedes Objekt nur der Status über die vergangenen Bewegungs- und Konfliktverhältnisse einbezogen. Dafür werden die folgenden Status angewendet:  $\{NEU, AKTIV, MULTI, SUCHEND, STEHEND, TOT\}$ . Damit werden die Beziehungen zwischen den Objekten und Messungen in den Teilbildsituationen nun interpretierbar. Diese Interpretation ermöglicht wiederum das Treffen von Entscheidungen über Zuordnungen und die Aktivierung des entsprechenden Bewegungsmodells. Diese Status werden den Partikelfiltern in jedem Schritt entsprechend seiner Situation zugeordnet. Sie dienen der Unterscheidung bei der Anwendung der Bewegungsmodelle und werden im nächsten Assoziationsschritt wieder verwendet und aktualisiert.

**Eindeutige Zuordnungen:** Einträge der Assoziationsmatrix ohne unmittelbaren Nachbarn liegen in keiner Konfliktsituation. Diese Zuordnungen umfassen alle Situationen, in denen eine eindeutige Zuordnung zwischen einem Objekt und einer Messung erfolgen kann und somit keine Konflikte entstehen und entsprechen den Mengen  $EZ_l$ .

- Da eine Messung zugeordnet werden konnte und die Entwicklung der jeweiligen Partikelfilter in den nächsten Schritten standardmäßig abläuft, erhalten die Partikelfilter zu ihren Zustandsparametern einen Status AKTIV, um die möglichen Situationen innerhalb dieser Mengen unterscheiden zu können. Dieser Status soll Aufschluss über die letzten Bewegungs- und Konfliktverhältnisse eines Objektes geben. Zusätzlich werden die jeweils zugeordneten Messungen den Partikelfiltern im Aktualisierungsschritt als Grundlage dienen. Die Bewegungsschätzung für den nächsten Zeitschritt erfolgt für diese Paare mit dem *Standardmodell* (Abb. 4-3-5b).

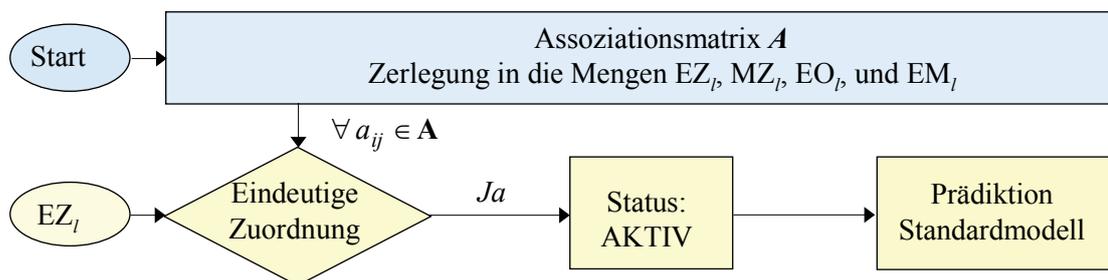


Abbildung 4-3-5b: Die vereinfachte Interpretation der Assoziationsmatrix  $A$  im Fall einer eindeutigen Zuordnung.

**Mehrdeutige Zuordnungen:** Diese Situationen lassen sich aus den Einträgen in der Assoziationsmatrix ableiten, die einen oder mehrere Nachbarn aufweisen. Die Ursachen dafür sind auf nicht-kooperative Messsituationen, wie beispielsweise Verschmelzen, Teilen, der Beginn oder das Ende einer Verdeckung, zurückzuführen. Stehen gebliebene oder verlassende Objekte in Verbindung mit anderen Objekten können ebenfalls solche

#### 4. OBJEKTVERFOLGUNG AUF DER BASIS VON DATENASSOZIATIONEN

Konfliktsituationen herbeiführen, die sich aus der berechneten Assoziationsmatrix ableiten lassen. Diese Zuordnungsmöglichkeiten entsprechen den Mengen  $MZ_l$  (s. Abb. 4-3-5c).

- An dieser Stelle werden auch Störungen entfernt, die beispielsweise durch Segmentierungsfehler auftreten können und Mehrdeutigkeiten verursachen. Aus Störungen resultierende, erfasste Objekte zeichnen sich durch ihre Unbeständigkeit aus. Wurde im letzten Schritt ein Partikelfilter für ein Objekt initialisiert, trägt es den Status *NEU* und gilt noch nicht als etablierter Partikelfilter, das heißt, dass er noch keine Historie und somit auch keine Stabilität aufweist. Da dem Objekt keine aktuelle Messung entspricht, wird es als Störung interpretiert und erhält den Status *TOT*. Dies veranlasst das Partikelfiltersystem diesen Partikelfilter zu verwerfen.
- Handelt es sich um etablierte Partikelfilter, ist davon auszugehen, dass sich mehrere Objekte einen Blob teilen und dadurch keine eindeutige Zuordnung erhalten können. Die Partikelfilter erhalten den Status *MULTI* und schätzen im nächsten Schritt mit dem *Erfahrungsmodell*.

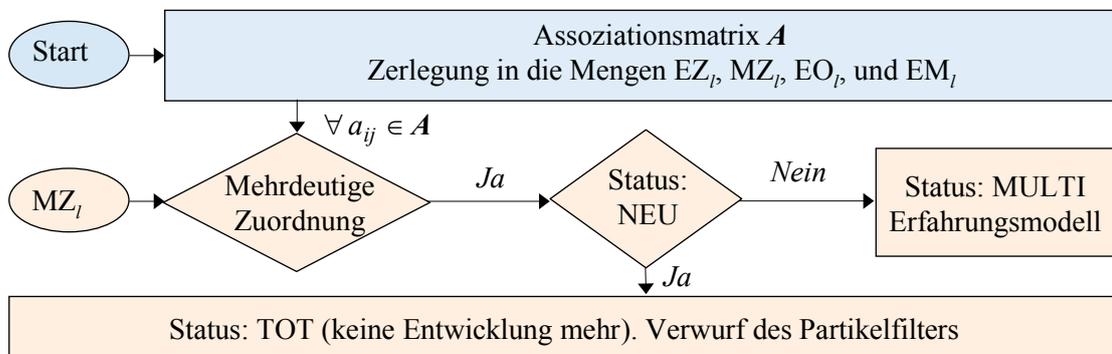


Abbildung 4-3-5c: Interpretation der Assoziationsmatrix  $A$  im Fall einer mehrdeutigen Zuordnung.

**Einzelne Objekte:** Hier werden die Objekte bzw. Partikelfilter betrachtet, deren Assoziationsmatrix keiner Messung zugewiesen werden konnte und somit leere Spalten in der Assoziationsmatrix aufweist. Sie entsprechen den Kandidatenmengen  $EO_l$ . Dies kann mehrere Ursachen haben, die wie folgt unterschieden werden.

- Ein stehen gebliebenes Objekt wird in der Bewegungssegmentierung nicht mehr erkannt, obwohl es jedoch noch in der Szene vorhanden ist. Mit einer Messung an der letzten bekannten Stelle sind Daten aus dieser Region zu extrahieren und mit den Eigenschaften des Objektes anhand von  $SM_{Obj}$  zu vergleichen (Stopp-Test in Abbildung 4-3-5d). Bei Übereinstimmung wird das Objekt als *STEHEND* interpretiert. Für den Partikelfilter wird das *Standmodell* eingeschaltet.
- Eine weitere Ursache kann das Verlassen der Szene sein. In diesem Fall zeichnet sich das Objekt durch seine randnahe Position aus. Unter der Annahme gleichmäßiger Bewegungen liegt auch die Bewegungsschätzung außerhalb der betrachteten Szene (Rand-Test in Abbildung 4-3-5d). Das Objekt erhält den Status *TOT*. Dies veranlasst ebenfalls das Partikelfiltersystem diesen Partikelfilter zu verwerfen. Des Weiteren werden Partikelfilter als *SUCHEND* interpretiert, da diese keinem der bisherigen Fälle entsprechen. Dieser Fall kann aufgrund von Verdeckung verursacht werden. Für diesen Partikelfilter wird das *Erfahrungsmodell* eingeschaltet.

#### 4. OBJEKTVERFOLGUNG AUF DER BASIS VON DATENASSOZIATIONEN

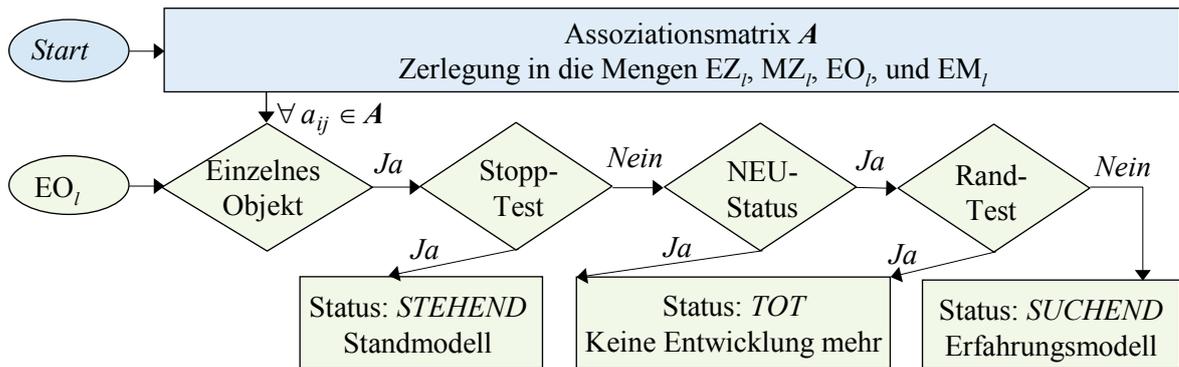


Abbildung 4-3-5d: Interpretation der Assoziationsmatrix im Fall eines einzelnen Objekts.

**Einzelne Messungen:** Die verbleibenden Zuordnungspaare weisen in der Assoziationsmatrix leere Zeilen auf, da für sie im Datenausschluss kein entsprechendes Objekt gefunden worden ist. Ursache dafür können aber auch Störungen sein, die für eine stabile Analyse als solche erkannt werden müssen.

- Tritt im Datenausschluss der Fall auf, dass einem Objekt wegen eines zu kleinen Suchbereichs keine entsprechenden Messungen zugeordnet wurde, ist es zuvor als einzelnes Objekt behandelt und mit dem Status *SUCHEM* versehen worden. An dieser Stelle wird es mit der freien Messung über  $SM_{Obj}$  verglichen (Suchend-Test in Abbildung 4-3-5e). Bei Übereinstimmung erfolgt die Objekt-Messung-Zuordnung. Der entsprechende Partikelfilter erhält den Status *AKTIV* und schätzt im nächsten Schritt mit dem *Standardmodell*.
- Meist liegen allerdings freie Messungen zu Objekten vor, die noch nicht in der Szene erfasst wurden – die also die Szene betreten haben. In diesem Fall wird eine Initialisierung eines neuen Partikelfilters veranlasst, der den Status *NEU* erhält.

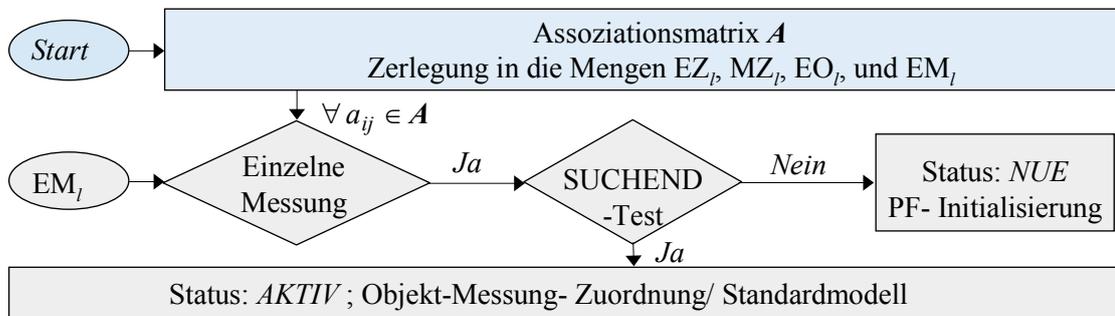


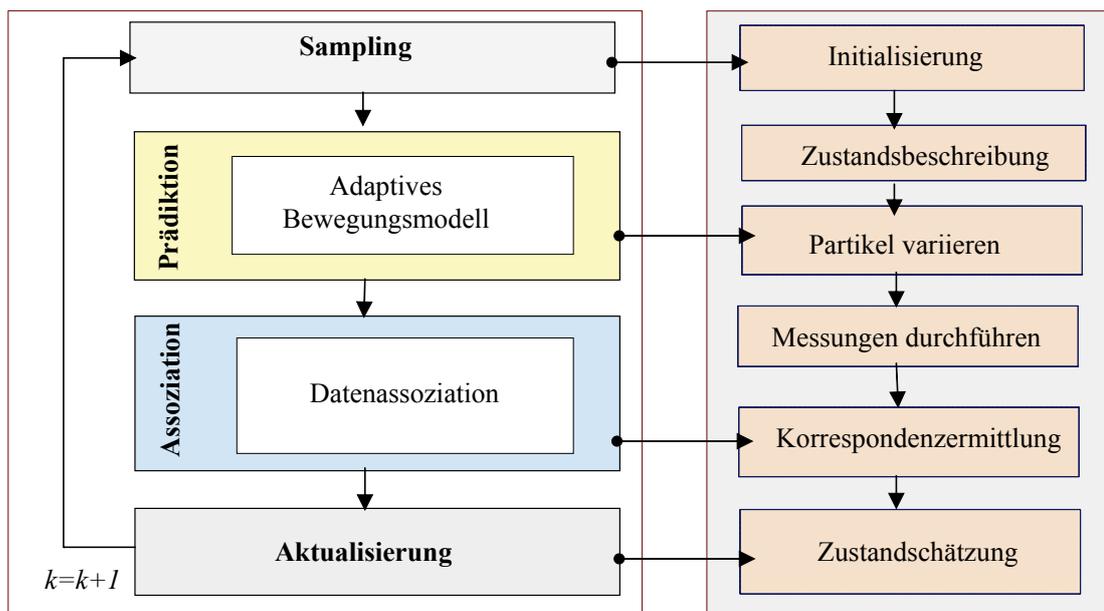
Abbildung 4-3-5e: Interpretation der Assoziationsmatrix im Fall einer einzelnen Messung.

Mit der hier durchgeführten Interpretation aller Einträge der Assoziationsmatrix ist das Partikelfiltersystem der Bildsituation angepasst. Es wurden Partikelfilter entsprechend der Objektfluktuation initialisiert und abgewiesen. Daher erhalten die erstellten Partikelfilter bei der Initialisierung den Status *NEU*. In jedem weiteren Zeitschritt werden die Status in der Datenassoziation aktualisiert und dienen der Analyse im nächsten Zeitschritt. Die Partikelfilter besitzen ihre Zuordnungen, wenn die entsprechende Messung vorliegt. Dabei erhalten Partikelfilter mit eindeutigen Zuordnungen den Status *AKTIV*, mit mehrdeutigen Zuordnungen *MULTI* und ohne Zuordnungen *SUCHEM*.

Partikelfilter, die als stehend identifiziert werden, bekommen den Status *STEHEND* und zu verwerfende Partikelfilter den Status *TOT*. Partikelfilter von Objekten in Konfliktsituationen sind durch das vorgeschlagene adaptive Bewegungsmodell in ihrem Entwicklungsverhalten angepasst. Daher erhält das Partikelfiltersystem alle Vorschriften, um den Assoziationschritt zu verlassen und die Aktualisierungsschritte der Partikelfilter durchzuführen.

#### 4.3.5 Verlauf einer Iteration des Partikelfiltersystems

Der für diese Arbeit angewendete Partikelfilter ist dem *Condensation* Algorithmus (s. Abschnitt 4.1) nachempfunden. Der Objektzustand ist ein vierdimensionaler Vektor, der die Position und die Geschwindigkeit jedes Motion-Blobs im Bild kodiert. Jedes Partikel entspricht einem Objektzustand. Wie der Partikelfilter mit dem adaptiven Bewegungs- und Beobachtungsmodell die Objektverfolgung für jedes Bild durchführt, wird schematisch in Abbildung 4-3-6 dargestellt. Für die fortlaufende Entwicklung der geschätzten Zustände werden die Prädiktion und die Aktualisierung um einen Sampling-Schritt ergänzt, welcher aus der gewichteten Partikelmenge nach der Aktualisierung eine neue Partikelmenge für die nächste Zeititeration generiert. Stark gewichtete Partikel werden vermehrt in die neue Menge aufgenommen, dagegen schwach gewichtete Partikel seltener.



**Abbildung 4-3-6:** Schematische Darstellung für den Verlauf einer Iteration des Partikelfiltersystems. Durch die automatische Initialisierung mittels MD-Ansatzes werden alle Partikel mit dem Zustandsparameter initialisiert. Mittels des adaptiven Bewegungsmodells werden aus dem vorhergehenden Partikelzustand den aktuellen unbekanntem Objektzustand vorhergesagt und mittels eines Beobachtungsmodells bewertet. Das weitere Schema entspricht der Abbildung 4-1-1.

**Sampling:** Im Sampling-Schritt des Zeitschritts  $k$  werden aus den gewichteten Partikelmengen  $S_{k-1}$  der  $M$  Partikelfilter die neuen Partikelmengen  $S_k$  generiert. Diese enthalten die  $N$  Partikel mit den prädierten Zuständen der jeweiligen Objekte, wobei die zu einer Messung ähnlichen Prädiktionen vermehrt und unähnliche seltener in die neue Menge übernommen wurden. Partikel mit großen Gewichten werden über einen

größeren Bereich repräsentiert als Partikel mit kleinen Gewichten. So wird gewährleistet, dass stark gewichtete Partikel, also Partikel mit typischem Objektzustand, mehrfach in die neue Menge  $S_k$  eingehen und schwach gewichtete dagegen seltener oder gar nicht. Auf diese Weise werden sinnvolle Objektzustände weiterentwickelt und nicht brauchbare verworfen. Dieser Vorgang erinnert an die Selektion in der Evolution und wird auch als Eigenschaft der Partikelevolution bezeichnet. Die Partikelevolution ermöglicht es, Objektzustände abzufangen, die durch unvorhersehbare Veränderungen des Objekts auftreten können. Es ermöglicht aber auch die Zustände weiter zu schätzen, wenn Mängel in den Messdaten auftauchen. Die Gewichte enthalten nun noch gleiche Werte, das heißt die Partikel sind noch ungewichtet/gleichgewichtet und die Partikelmenge stellt somit die ungewichtete Apriori-Verteilung dar (s. Abb. 4-1-2).

**Prädiktion:** Die Apriori-Menge ist die Grundlage für die Voraussagen der Partikelmengen mit dem adaptiven Bewegungsmodell. Hier wird nun jede Partikelmenge entsprechend der Objektsituation prädiziert, indem je nach aktuellem Status das passende Modell angewendet wird. *AKTIVE* Partikelfilter entsprechen eindeutigen Zuordnungen zwischen Objekt und Messung und werden daher mit dem Standardmodell entwickelt. Die Objektsituationen von Partikelfiltern mit dem Status *MULTI* und *SUCHEND* wurden als mehrdeutig bzw. verdeckt interpretiert und werden nun mit dem Erfahrungsmodell prädiziert. Für die Partikelmengen der *NEUEN* und *STEHENDEN* Partikelfilter wird das Standardmodell angewendet.

**Messung:** Wie in den Abschnitten 4.3.2 und 4.3.3 die Funktionalität des Partikelfiltersystems beschrieben wird, werden durch die Bewegungssegmentierung und Merkmalsextraktion die objektbeschreibenden Messungen für den Assoziations- und den Aktualisierungsschritt zur Verfügung gestellt.

**Assoziation:** In der 3-stufigen Assoziation wird nun festgestellt, welche Messung zu welchem Objekt zuzuordnen ist bzw. welcher Partikelfilter anhand welcher Messung aktualisiert werden muss. Darüber hinaus erfolgt die Analyse der Objektsituationen und somit die Zuordnung zum zugehörigen Status, die die Interpretation der Objektsituationen beschreiben und die Entwicklung der Partikelfilter lenken.

**Aktualisierung:** Partikelfilter mit dem Status *TOT* sind nicht mehr in der Szene enthalten und werden somit nicht mehr in den Entwicklungsprozess einbezogen. Dem Partikelfilter mit den Status *MULTI* und *SUCHEND* konnten keiner Messung zugeordnet werden. Somit liegt keine Grundlage zur Aktualisierung vor. Diese wird daher ausgesetzt, sodass die prädizierte Partikelmenge ungewichtet in den Sampling-Schritt eingeht und die Partikel als Folge einen größeren Suchbereich abtasten. Hingegen weisen *AKTIVE*, *NEUE* und *STEHENDE* Partikelfilter zugeordnete Messungen auf, mit denen sie über das Beobachtungsmodell aktualisiert werden. Hiermit stehen die Partikelmengen der Filter für den nächsten Zeitschritt bereit. Auf diese Weise entsteht eine Art evolutionäre Entwicklung der Partikelmenge entsprechend der Gegebenheiten im Bild, wobei eine Konzentration bei den Berechnungen auf die Partikel mit den höchsten Wahrscheinlichkeiten, nicht auf die ganze Umgebung oder gar den ganzen Zustandsraum, erreicht wird.

### 4.4 Analyse

Im Folgenden werden Untersuchungsergebnisse vorgestellt, die durch Anwendung des entwickelten Partikelfiltersystems gewonnen wurden. Dazu werden Resultate aus einigen realen Bildsequenzen und gestörte Datensätze von Verkehrsaufnahmen dargestellt, um die Leistungsfähigkeit und Effizienz des entwickelten Systems bei der Analyse von realen Bildsequenzen bei dem Einfluss von nicht-kooperativen Messsituationen (wie Deformationen, Verschmelzung und Verdeckung) zu demonstrieren.

#### 4.4.1 Bildsequenzen ohne Konfliktsituationen

Die erste Bildsequenz entstand durch die Aufnahme einer Verkehrsszene, die zunächst der Erklärung der Verarbeitungskette des entwickelten Verfahrens dient. Vorteilhafte Eigenschaften für eine Analyse sind hingegen die hohe Bildhelligkeit, kleine perspektivische Größenverzerrung und Teilverdeckung (durch Pfeiler) des einzelnen Objektes (Abb. 4-4-1). Um das bewegte Objekt erfassen zu können, wird ausgehend von einem statischen Aufnahmesystem der vorgestellte MD-Ansatz angewendet. Herausforderungen sind die Schattenregionen und die schwachen Texturen auf den Objektoberflächen, die zur Verformung, Löchern und zerklüfteten Regionen innerhalb einer Objektregion führen, die durch eine Nachverarbeitung mit einer Verarbeitungskette nach [Al-Hamadi 03, 05] beseitigt werden können.

Durch die Verwendung des MD-Ansatzes mit morphologischen und konturbasierten Trennungsoperatoren wird der Großteil der Störungen entfernt (vgl. Kapitel 3). Außerdem wird der Bewegungsschatten entfernt und Löcher innerhalb der Region geschlossen. Darüber hinaus werden kleine Verdeckungen (z.B. durch einen Pfeiler) aufgrund der morphologischen Operationen die Segmentierungsergebnisse nicht beeinflussen (Abb. 4-4-1).

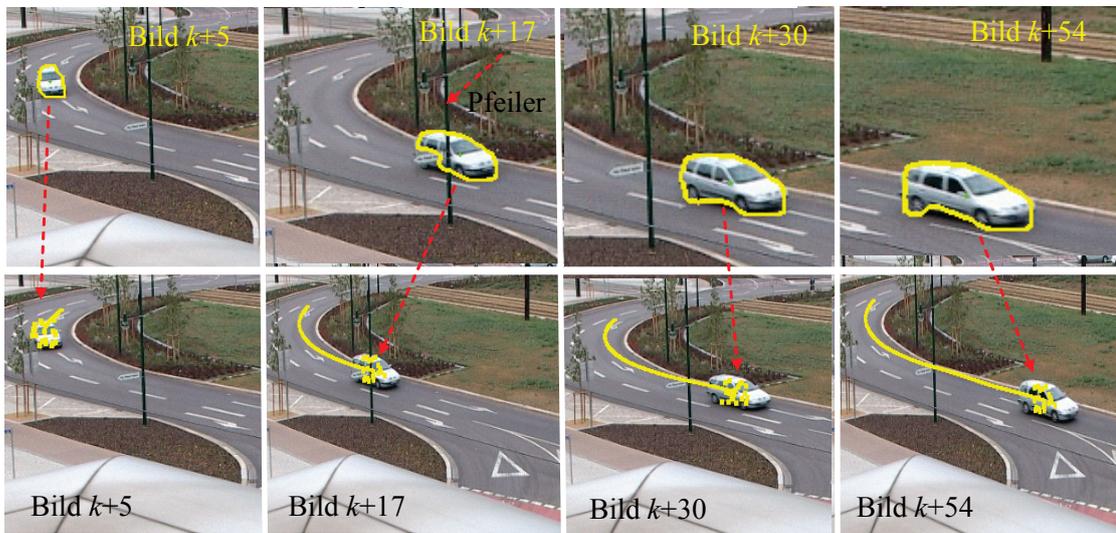
Die Segmentierung verläuft vollständig automatisch und liefert somit in jedem Zeitschritt die MBs (erste Reihe in Abbildung 4-4-1), die als Merkmalsträger dienen. Auf diese Weise wird in objektreicheren Bildszenen sowohl die geforderte automatische Initialisierung möglich als auch die Erkennung neuer Objekte realisiert.

Der Blob bildet nun die Grundlage für die Extraktion der Merkmale, durch die ein Objekt beschrieben und verfolgt wird. Mit der vorgestellten Farbraumzerlegung ist der *Blob* über seine Hauptfarben beschreibbar. Hierfür werden Schwerpunkt, Ausdehnung und Ausrichtung der Farbcluster im zerlegten Farbraum herangezogen (s. Abschnitt 4.3.2). Das hier betrachtete Objekt enthält hauptsächlich nur zwei Farbrichtungen, sodass sich der dritte Cluster auf Randpixel bezieht. Dies besteht allerdings durch die Sequenz hinweg und wirkt somit auch charakteristisch für diesen Blob. Grenzen dieser Farbcluster werden im Verlauf des Kapitels weiter betrachtet. Die Bild- und Farbparameter dienen nun dem Partikelfilter als Eingabe. Mit dem Bewegungsmodell wird zunächst die Bewegung des Objekts geschätzt, indem die Partikel nach der Vorschrift des Bewegungsmodells transformiert und variiert werden.

Im Initialschritt steht jedoch im vorgestellten Bewegungsmodell noch keine Bewegungsvorschrift zur Verfügung. So wird in der ersten Iteration durch bloßes Variieren der Partikel geschätzt (Abbildung 4-4-1; Bild  $k+5$ ). Im verwendeten Material betragen die bekannten größten Bewegungsdistanzen etwa zehn Pixel pro Zeitschritt. Mit zusätzlichem Sicherheitsrand werden die Partikel initial im Intervall von [15, 15] gleichverteilt gestreut. Die einfache Szene in Abbildung 4-4-1 weist keine Konfliktsi-

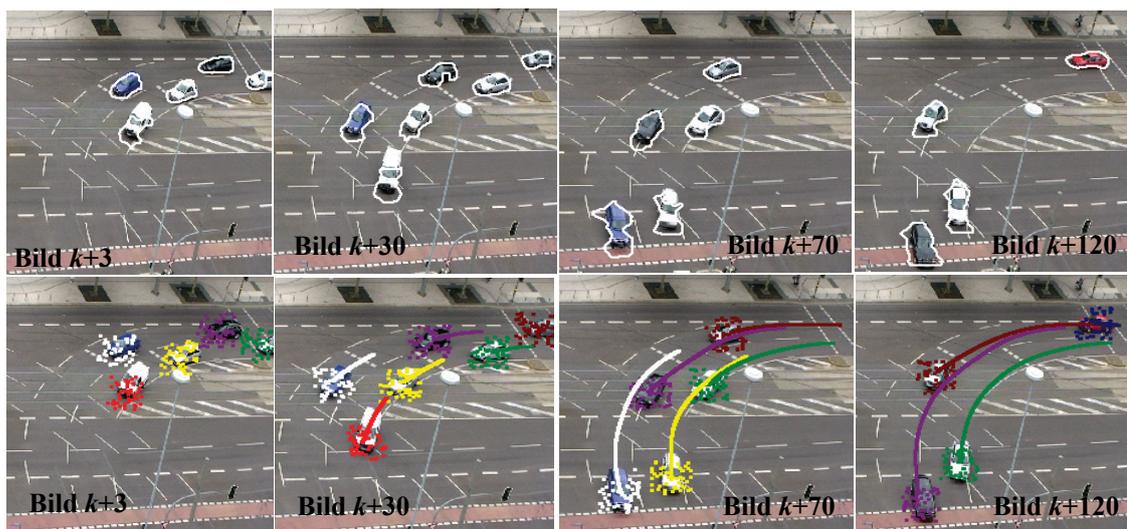
#### 4. OBJEKTVERFOLGUNG AUF DER BASIS VON DATENASSOZIATIONEN

tuationen auf. Eine eindeutige Zuordnung und Bewertung ist relativ einfach. Somit kann die Objektbewegung in den Folgeschritten mit dem Standardmodell geschätzt werden. Die Bilder der Zeile B kennzeichnen die platzierten Partikel durch gelbe Quadrate.



**Abbildung 4-4-1:** Objektverfolgung mit dem vorgeschlagenen Partikelfiltersystem. Durch automatische Initialisierung werden alle Partikel mit dieser Position und Geschwindigkeit 0 initialisiert. Die Bewegungstrajektorien beschreiben die tatsächliche Bewegung des Objektes.

In der neuen Messung, in Reihe B (Abb. 4-4-1) bewertet das Beobachtungsmodell nun die Partikel. Dabei wird die Ähnlichkeit über Position, Bewegungsrichtung und Bewegungsgeschwindigkeit ermittelt. Der gewichtete Durchschnitt der Partikel stellt nun die Werte für den neuen Zustand des Objekts und den Verlauf der Objekttrajektorie dar. Reihe B zeigt den jeweiligen Stand der Objekttrajektorie mit den Partikeln. Daraus wird die neue Iteration eingeleitet.



**Abbildung 4-4-2:** Verkehrsaufnahme dicht fahrender PKWs. Die erste Reihe beinhaltet die automatische Segmentierung und somit die Initialisierung des Partikelfiltersystems. Die zweite Reihe beschreibt die ermittelten Bewegungstrajektorien durch das vorgeschlagene Partikelfiltersystem.

## 4. OBJEKTVERFOLGUNG AUF DER BASIS VON DATENASSOZIATIONEN

---

In Abbildung 4-4-2 ist die Analyse einer Verkehrsaufnahme von fahrenden PKWs dargestellt. Dabei werden in frei fließendem Straßenverkehr keine Konfliktsituationen auslöst. Die erste Reihe beinhaltet die automatische Segmentierung und somit die Initialisierung des Partikelfiltersystems. Die zweite Reihe beschreibt die ermittelten Bewegungstrajektorien durch das vorgeschlagene Partikelfiltersystem.

### 4.4.2 Bildsequenzen mit Konfliktsituationen

Reale Bildszenen enthalten im relativ dichten Straßenverkehr mehrere Objekte mit Konfliktsituationen, die in der Regel ein Korrespondenzproblem bei der Datenzuweisung verursachen. Solche problematischen Situationen treten nicht nur durch Verdeckungen, Verschmelzung, Zerfallen in der Szene oder andere Objekte auf, sondern auch bei Beleuchtungsänderungen, Objektsfluktuationen und Interaktionen von Objekten (*aufgrund des dichten Verkehrs*). Bei Ansätzen mit bewegungsorientierter Objekterfassung erzeugen aber auch stehen gebliebene Objekte Konfliktsituationen, da keine dem Objekt entsprechende Messungen mehr geliefert werden kann. Um dieses Korrespondenzproblem zu lösen, wurden der Assoziationsschritt und ein adaptives Bewegungsmodell entwickelt, die die Handhabung solcher außerordentlichen Situationen ermöglichen. Hierzu werden folgend entsprechende Konfliktsituationen anhand der realen Bildszenen betrachtet.

Abbildung 4-4-3A zeigt eine Szene, in der ein Objekt stehen bleibt. Der Personenkraftwagen (PKW) im unteren Bildabschnitt hält an einem Fußgängerüberweg an und beim völligen Stillstand wird der PKW vom System aufgrund des angewandten MD-Segmentierungsansatzes nicht mehr erfasst. Erst beim Weiterfahren wird er wieder erkannt und als neues Objekt initialisiert werden. Die Herausforderung besteht nun darin, ihn als Teil der Bildszene und die Trajektorie als seine Historie zu bewahren, die für die Lösung des Korrespondenzproblems und der weiteren Klassifikation angewendet werden kann. Hierfür sind in dem vorgeschlagenen System das Bewegungsmodell und der Assoziationsschritt zuständig. Ausgehend davon wird im ersten Zeitschritt des Stillstands nach dem Datenausschluss und der Datenzuweisung mindestens ein Objekt bzw. ein Partikelfilter ohne Messkandidaten verbleiben, der in der weiteren Datenverwaltung als einzelnes Objekt getestet wird. Handelt es sich nun um den stehen gebliebenen PKW, so registriert der Stopp-Test die letzten bekannten Daten an der letzten bekannten Stelle und interpretiert den PKW als stehen geblieben. Daher wird der Partikelfilter des PKWs als STEHEND eingeschätzt und bleibt somit in der Szene erhalten. Durch den Status STEHEND wird für diesen Partikelfilter das Bewegungsmodell auf das Standmodell geschaltet und im nächsten Schritt nach dessen Vorschrift prädiziert. Erst wenn sich der PKW weiterbewegt, wird er in der Bewegungssegmentierung wieder erkannt, dem entsprechenden Partikelfilter eindeutig zugeordnet und zur Bewertung der Partikel herangezogen. Für den nächsten Zeitschritt erhält er wieder den Status AKTIV und wird entsprechend mit dem Standardmodell geschätzt.

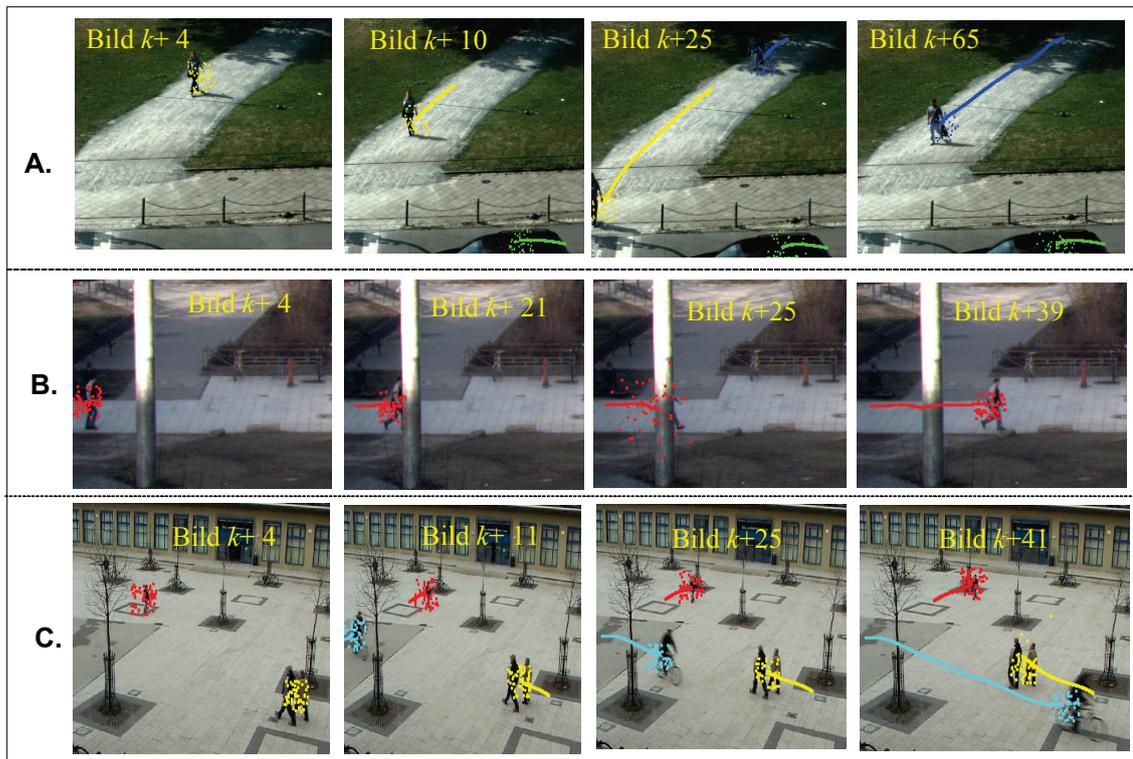
In Abbildung 4-4-3B ist eine weitere Bildszene dargestellt, die eine weitere nicht-kooperative Messsituation beinhaltet, die zur uneindeutigen Zuordnung von Korrespondenzen und somit zur falschen Einschätzung der Bewegungstrajektorie führen wird. Dort bewegt sich die Person hinter einem Pfeiler entlang und wird dabei kurzfristig vollständig verdeckt. Dem Betrachter der Situation erschließt sich durch die Kontextinformationen der Szene, dass das Objekt kein stehen gebliebenes, sondern ein verdecktes ist. Ähnliche Interpretation erfolgt in der Datenverwaltung, in der das Objekt gleichfalls

#### 4. OBJEKTVERFOLGUNG AUF DER BASIS VON DATENASSOZIATIONEN

als einzelnes Objekt getestet wird (Abb. 4-3-4c). Allerdings sind die Daten des Objekts nun weder an der letzten bekannten Stelle nachmessbar – ist also nicht stehen geblieben – noch befindet sich das Objekt am Szenenrand und würde mit der geschätzten Bewegung die Bildszene verlassen. Außerdem handelt es sich um ein etabliertes Objekt, das heißt, es hat den Status AKTIV, so dass es nicht als kurzfristige Störung interpretiert werden könnte.

Dem Flussdiagramm in Abbildung 4-3-4c ist zu entnehmen, dass der Partikelfilter des Objekts nach den Bewertungen nun in den Status SUCHEND übergeht. Somit wird für den nächsten Zeitschritt das Erfahrungsmodell aktiviert. Da keine Messung zugeordnet werden konnte, bleiben die Partikel des Filters ungewichtet und gehen deshalb jeweils mit gleichen Chancen in den Sampling-Schritt ein. Dadurch werden auch unwahrscheinliche Partikel in die neue Partikelmenge aufgenommen. Das Erfahrungsmodell schätzt und variiert nun anhand der durchschnittlichen Bewegungen der letzten Iterationen die laufende, aber verdeckte Bewegung des Objekts. Durch die nicht erfolgte Bewertung und die Variation im Erfahrungsmodell streuen die Partikel einen weiteren Bereich, als im gewichteten Zustand.

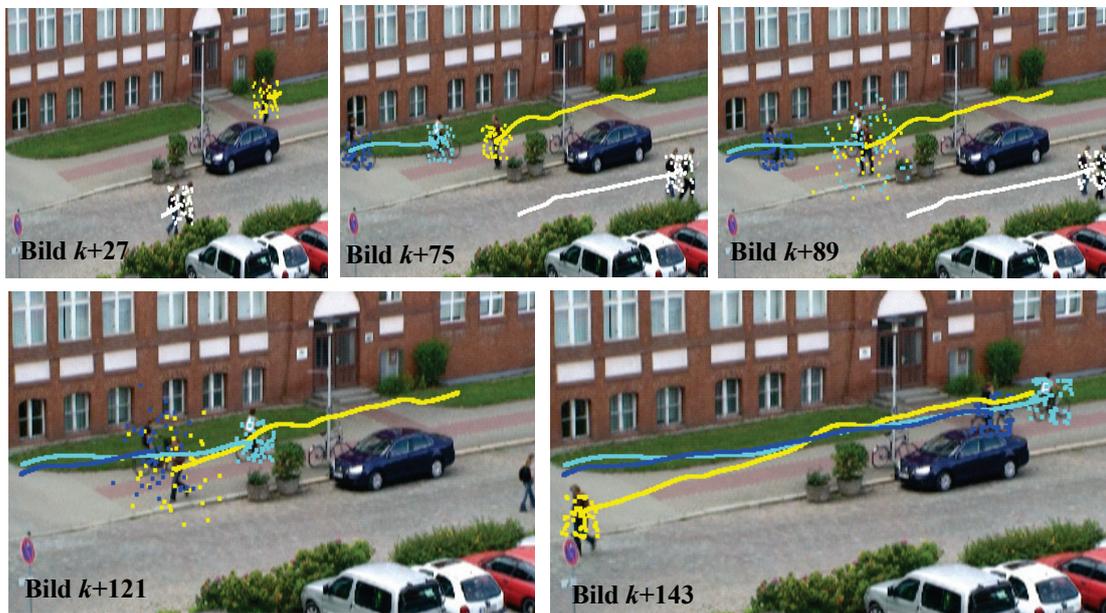
Abbildung 4-5-2B zeigt den vergrößerten Bereich, den die Partikel abdecken, wenn die Person sich hinter dem Pfeiler befindet. Erst mit dem Hervortreten wird die Person wieder erfasst und dem Partikelfilter zugeordnet. Die Bewertung der Partikel erfolgt wieder anhand der zugeordneten Messung. Somit konzentrieren sich die Partikel wieder um die Position des Objekts.



**Abbildung 4-4-3:** Die erste Reihe beinhaltet die ermittelten Bewegungstrajektorien durch das vorgeschlagene Partikelfiltersystem. Obwohl der PKW im Bild  $k+25$  bis  $k+65$  stehen geblieben ist, bleibt das Objekt Teil der Szene und die Trajektorie erhalten. Die zweite und dritte Reihe zeigen Resultate des hier entwickelten Verfahrens bei Teilverdeckung und Interaktionen. Die Punkte geben die Partikel an, die durch die Information aus der Datenassoziation positioniert wurden. Dadurch wird es problemlos möglich, die Partikelmenge an dem Störmerkmal vorbeizubewegen.

#### 4. OBJEKTVERFOLGUNG AUF DER BASIS VON DATENASSOZIATIONEN

Abbildungen 4-4-3C und Abbildung 4-4-4 zeigen Situationen, in denen deformierbare Objekte sich im Laufe ihrer Bewegungen verdecken und somit Uneindeutigkeiten in der Zuordnung auslösen. Für den Aktualisierungsschritt wird demzufolge keine Messung und zur nächsten Bewegungsschätzung keine aktuelle Vorschrift vorliegen. Zur Lösung solcher Objektsituationen wirkt der Assoziationsschritt wie folgt: Zunächst bewegen sich die Objekte ohne Konflikte in der Szene. Sie werden vom Partikelfiltersystem als eindeutige Zuordnungen verarbeitet. Mit dem Status AKTIV erfolgt ihre Entwicklung mit dem Standardbewegungsmodell. Nähern sich die Objekte an, verschmelzen sie in der Bewegungssegmentierung zu einem Blob. Diese Teilszene wird demnach vorerst mit einem angenommenen Objekt erfasst. Die Partikelfilter der Objekte sind hingegen noch im Partikelfiltersystem enthalten. In den Schritten des Datenausschlusses und der Datenzuweisung (Abschnitt 4-3-4) wird diese Situation nun als Konfliktsituation erkannt.



**Abbildung 4-4-4:** Die Bilder zeigen die Resultate des hier entwickelten Partikelfiltersystems. Mit der Datenassoziation bleiben die Objekte Teil der Szene. Die ermittelten Bewegungstrajektorien durch das adaptive Bewegungsmodell beschreiben die tatsächlichen Bewegungen der unterschiedlichen Objekte trotz der Teilverdeckung. Die Punkte geben die Partikel an, die durch die Information aus der Datenassoziation positioniert wurden. Dadurch wird es problemlos möglich, die Partikelmenge an dem Störmerkmal vorbeizubewegen.

Die einzelne Person im Hintergrund  $z_1$  erhält eine eindeutige Zuordnung zur Messung  $m_1$ , während der Radfahrer  $z_2$  und die Personen  $z_3$  die einzelne Messung  $m_2$  in ihren überlappenden Suchbereichen (gemäß des Datenausschlusses) erhalten. In der Datenzuweisung mit der Bildung der Kandidatenmengen, Hypothesengenerierung und JPDA konnte diese Situation nicht weiter aufgelöst werden, so dass die Konfliktszene mit Radfahrer und Paar in die Datenverwaltung eingeht. Da beide Objekte etablierte Partikelfilter im Partikelfiltersystem sind – bisher also den Status Aktiv tragen – wird diese Szene als Multi-Szene interpretiert. Beide Partikelfilter erhalten den Status Multi. Somit erfolgt wie im vorangegangenen Beispiel die nächste Bewegungsschätzung mit dem Erfahrungsmodell. Die weiter gestreuten Partikel im Bild kennzeichnen die Entwicklung der ungewichteten und weiter entwickelten Partikel während der

Konfliktphase. Erst mit dem Ende der Verdeckung werden die Objekte wieder einzeln erfasst. Die Bewertung der Partikel erfolgt wieder anhand der zugeordneten Messung. Somit konzentrieren sich die Partikel wieder um die Position des Objekts.

### 4.5 Schlussfolgerung

Die oben beschriebenen Beispielsequenzen zeigen, dass das in dieser Arbeit vorgeschlagene System auf der Basis Partikelfilter und Datenassoziation die in Abschnitte 4.2 und 1.1 gestellten Anforderungen erfüllt. Die Grundlage für die Erfüllung der Anforderungen liegt jeweils in den vorgeschlagenen Ansätzen der Bewegungssegmentierung, der Zustandsbeschreibung sowie der Multi-Objektverfolgung.

- Für die automatische Objekterfassung wurde ein regionenbasierter Ansatz entwickelt, der auf Bewegungssegmentierung in Verbindung mit morphologischen Operationen und Separationsalgorithmen basiert. Dieser liefert automatisch die Motion-Blobs und ermöglicht dadurch die automatische Initialisierung der Partikelfilter. Darüber hinaus erfordert dieser Ansatz keine zusätzlichen Informationen aus der Bildszene, weder über Objekte noch Umgebung, und erlaubt so den Einsatz für verschiedenste reale Bildszenen.
- Ausgehend von der Tatsache, dass sich in den Bildszenen verschiedenste Objekte aufhalten, sowohl starre als auch nicht starre Objekte, kann hier kein Modellwissen verwendet werden, das die Analyse auf bestimmte Objekte beschränkt. Daher wird eine effiziente Beschreibung über die Farbe der Hauptbestandteile der Objekte vorgeschlagen, die durch eine Farbraumzerlegung und Beschreibung der Haupt-Cluster erfolgt. Somit wird eine ausreichende Objektbeschreibung ohne bildbezogene Anordnungen erreicht, die auch bei Deformationen nicht starrer Objekte erhalten bleibt.
- Zur Multi-Objektverfolgung wurde ein System aus unabhängigen Partikelfiltern, die je ein Objekt repräsentieren angewendet. Dadurch ist eine getrennte und individuelle Bewegungsschätzung der unterschiedlichen Objekte in der Bildszene möglich. Die Partikelfilter erfassen jeweils die Bewegung pro Objekt und erreichen somit eine zerlegte und vereinfachte Beschreibung der multimodalen Bewegung der Bildszene. In der Datenassoziation erfolgen die Interpretation von Konfliktsituationen und die Organisation deren unterschiedlicher Handhabung. Daher basiert die Fluktuationserkennung sowohl auf den Segmentierungsergebnissen als auch auf der Datenassoziation, die das Partikelfiltersystem veranlasst, neue Partikelfilter zu initialisieren und alte zu verwerfen. Zudem wird die Entwicklung der Partikelfilter von partiell oder vollständig verdeckten und auch stehen gebliebenen Objekten durch die Einschaltung der entsprechenden Bewegungsmodelle (Erfahrungs- und Standmodell) in der Datenassoziation erreicht. Mit der Datenassoziation in ihren drei Schritten wurde eine Ebene in das Verfahren integriert, die es ermöglicht, die Objekte der Bildszene losgelöst von der Bildebene zu erfassen und dadurch die Sucheigenschaften der Partikelfilter auszunutzen, um auch Konfliktsituationen zu bewerkstelligen.

Allerdings bergen die zusammengesetzten Verarbeitungsschritte des vorgeschlagenen Systems auch Schwierigkeiten in der Analyse. Anhand der komplizierten Szene mit Schattenwurf entstehen abwechselnd zusammenhängende und getrennte Blobs aus

#### 4. OBJEKTVERFOLGUNG AUF DER BASIS VON DATENASSOZIATIONEN

---

mehreren Objekten. Diese Unbeständigkeiten in den Segmentierungsergebnissen führen zu vielen initialisierten Partikelfiltern, die wieder verworfen oder falsch zugeordnet werden. Ähnliche Effekte entstehen auch bei Lichtkegeln auf der Straße durch Fahrzeugbeleuchtung. Für das entwickelte Verfahren zeigt sich eine hohe Abhängigkeit von der Qualität der Segmentierungsergebnisse. Daher ist für die Analyse solcher Situationen eine Integration von Methoden, die diese Effekte eliminieren, erforderlich.

Mit der Verwendung regionenbasierter Bewegungssegmentierung wurde das Verfahren mit der Annahme entwickelt, dass ein Motion-Blob ein Objekt repräsentiert. Dies erschwert nicht nur die Zuordnung im oben genannten Schattenproblem, sondern erzwingt auch die Erfassung von eng auftretenden Objekten als ein Objekt, wie das Fußgängerpaar. Damit einher geht die Erkennung beim erstmaligen Trennen eines solchen Blobs, wenn sich die darin enthaltenen Objekte weiter von einander bewegen. Das entwickelte System würde für diese Objekte neue Partikelfilter initialisieren, diese aber nicht dem alten zuordnen können. Für die Lösung dieses Problems sind weitere Analysen der Blobs mit bildbezogenen Eigenschaften notwendig, die aber genaueres Wissen über die vorhandenen Objekte voraussetzen.

Die umfangreiche Zusammensetzung des entwickelten Systems aus den verschiedensten Methoden enthält Stellen, die in optimierter Weise zur Verbesserung des Systems beitragen können. Der Schwerpunkt dabei ist die Verbesserung der Segmentierungsergebnisse.

- Die Untersuchungen und aufgezeigten Grenzen verlangen die Integration einer Schatteneeliminierung. Ebenso sind Methoden zur Entfernung anderer Störeffekte wie reflektierte Fahrzeugbeleuchtung, die zur Verschmelzung der Blobs mehrerer Objekte führen, erforderlich. Für die Analyse von Szenen mit direktem, aber auch indirektem Schatten, ist eine weitere Ausnutzung des adaptiven Farbraumes [Al-Hamadi 01a] denkbar. Die Untersuchungen der Arbeit beschränkten sich gegebenenmaßen auf gleichmäßig beleuchtete Szenen. Die verwendete Farbraumzerlegung enthält noch den Helligkeitsanteil. Mit der Abkopplung dieser Komponente können erweiterte Objektbeschreibungen aufgestellt werden, die invariant gegen Beleuchtungsschwankungen sind und somit den Einsatzbereich des Verfahrens erweitern. Hier kann die Verwendung des CSC-Algorithmus weitere Verbesserungen liefern.
- Auch am adaptiven Bewegungsmodell wäre eine weitere Verbesserung möglich, da das hier benutzte Bewegungsmodell Schwächen besitzt, insbesondere wenn die Objekte spontan die Bewegungsrichtung wechseln. Hierzu wäre es denkbar, den Bewegungszustand der Objekte mit einem Fuzzy-System (s. Kapitel 5) mit den Zuständen für unterschiedliche Bewegungsrichtungen zu verwenden, bei dem die Partikelmenge dann je nach Zustand unterschiedlich propagieren werden kann.
- In dieser Arbeit wurden im Fall der Auswertung von merkmalspezifischen Ähnlichkeitsmaßen eines Objektes feste Gewichte für deren Integration in den Schätzprozess gewählt. Diese Ähnlichkeitsmaßen könnte durch eine adaptive Gewichtung ähnlich wie in Kapitel 3 abhängig vom aktuellen Bildinhalt optimiert werden. Dadurch könnte eine stabilere Verfolgung auch bei sich ändernden Gegebenheiten in der Szene ermöglicht werden.
- Im Datenausschluss wird bisher die euklidische Distanz verwendet, die je nach Ausdehnung und Verteilung der Partikel auch ein großes unwahrscheinliches Gebiet in den Suchbereich einschließt und dadurch auch unwahrscheinliche Zuordnungspaare aufstellt. Die Partikel jeder Prädiktion beschreiben ein zusammenhän-

gendes Gebiet, in dem das Objekt erwartet wird. Mit der Mahalanobis-Distanz ist die Beschreibung dieses Gebiets kompakter möglich und somit auch eine genauere Aufstellung der Zuordnungspaare denkbar, da der Unsicherheitsbereich nicht mehr von den Koordinatenachsen aufgespannt, sondern von den Eigenvektoren der Innovationskovarianz [Stüker 2004].

- In der Datenzuweisung werden bei der Hypothesenbildung Mehrfachzuweisungen pro Objekt derzeit ausgeschlossen. Von Kirubarajan [Kirubarajan 00] wird eine JPDA mit Mehrfachzuweisung vorgeschlagen. Mit Mehrfachzuweisungen pro Objekt kann beispielsweise die Zuordnung von gesplitteten Blobs durch fehlerhafte Segmentierung ermöglicht und somit die Verfolgung von Störungen als Objekte vermieden werden. Das entwickelte Verfahren wurde unter allgemein gehaltenen Annahmen entworfen, so dass für spezielle Einsätze zusätzliche, unterstützende Informationen herangezogen werden können. Beispielsweise kann durch die Verwendung entsprechender Umgebungsmodelle, die die Szene in Bereiche für Straßen, Fußwege, Überwege, Ampeln etc. teilen und so eine weitere Grundlage für Annahmen über die Objekte und deren Verhalten bieten. Ebenso ist die Verarbeitung speziellerer Informationen auf der Ebene der Partikelfilter denkbar.

Der hier realisierte Einsatz von Partikelfiltern schöpft die Möglichkeiten und Vorteile eines Partikelfilters noch nicht vollständig aus. In dieser Arbeit wurden die Partikelfilter in einer einfachen Modellierung verwendet. Für speziellere Anwendungen sind komplexere Eingaben als die Blobs möglich. Zur Differenzierung von Zielobjekten sind dabei beispielsweise verschiedene Objektmodelle vollstellbar, die dann auf der Assoziationsebene organisiert werden. Dies stellt auch eine Möglichkeit dar, die Trennung verschmolzener Objekte zu erreichen, die in diesem Verfahren durch die Differenzbildtechnik als ein Objekt angenommen werden. Welche dieser Ansätze auch weiterverfolgt werden, hängt von der gewünschten Anwendung ab. Diese Arbeit stellt dafür eine Grundlage dar.

### 4.6 Zusammenfassung

In diesem Kapitel wurde schrittweise ein System zur Multi-Objektverfolgung auf der Basis von Partikelfiltern entwickelt. Dafür wurden eingangs die Zielszenen verallgemeinert und Anforderungen analysiert. Der Anspruch, Verformungs-, Objekt- und Umgebungsunabhängigkeit zu erreichen, fordert eine Verarbeitung der Szenen auf abstrakterer Ebene, die Objekte aller Art in jeder Form repräsentiert. Außerdem sind Mechanismen notwendig, die die Verfolgung interagierender Objekte organisiert und ermöglicht.

Das Verfahren zur Multi-Objektverfolgung ist aus verschiedenen Schritten zusammengesetzt: Bildaufnahme, Bewegungssegmentierung, Merkmalsanalyse sowie Multi-Objektverfolgung. Im Schritt der Bewegungsanalyse wird eine Verarbeitungskette auf der Basis der Differenzbildtechnik eingesetzt, die objektunabhängig und verformungsrobust die Bewegungen im Bild liefert. Zur Formulierung und Repräsentation der Objekte werden Bild- und Farbparameter verwendet. Auf Bildebene sind dies lediglich ortsrelevante Daten wie Position oder Richtung. Zusätzlich ist ein Farbfingerabdruck für die Objekte entwickelt worden, mit der Anordnungsverhältnisse auf Bildebene ebenfalls umgangen werden.

#### 4. OBJEKTVERFOLGUNG AUF DER BASIS VON DATENASSOZIATIONEN

---

Die Bewegungsanalyse und die Merkmalsextraktion sind die Basis für die Verwendung von Partikelfiltern. Die Bild- und Farbparameter dienen der Formulierung der Bewegungs- und der Beobachtungsmodelle. Zur unterschiedlichen Modellierung der zu schätzenden Objektzustände in unterschiedlichen Situationen werden drei Modelle formuliert und angewendet: Standardmodell, Erfahrungsmodell, Standmodell. Dadurch entsteht ein Bewegungsmodell, das sowohl dynamisch auf die Bewegungsänderungen des jeweiligen Objekts reagiert als auch die Partikelevolution den äußeren Umständen anpasst.

Für die Multi-Objektverfolgung ist ein Partikelfiltersystem aufgestellt worden, das aus unabhängigen Partikelfiltern, die jeweils ein Objekt repräsentieren, besteht. Durch die Unabhängigkeit wird die individuelle Bewegungsschätzung mit dem adaptiven Bewegungsmodell möglich. Zudem wird durch die Zusammensetzung aus Partikelfiltern der Zustandsraum, somit Variationsmöglichkeiten und auch die Anzahl der notwendigen Partikel verringert. Die Verfolgung multipler Objekte und deren Trajektorien setzt räumliche und zeitliche Identifizierung voraus. Die Zuordnungen erfolgen im Assoziationsschritt in drei Stufen Datenausschluss, Datenzuweisung und Datenverwaltung. Durch Datenausschluss und Datenzuweisung wird die erfasste Szene zerlegt. Die Konfliktsituationen werden zwischen Objekten und Messungen soweit wie möglich aufgelöst und durch die JPDA Empfehlungen für die Objekt-Messung-Zuordnungen gegeben. Die Verwendung von Statusvariablen  $\{NEU, AKTIV, MULTI, SUCHEND, STEHEND, TOT\}$  gibt Aufschluss über die Bewegungs- bzw. Konflikthistorie der Objekte und unterstützt im Verwaltungsschritt die Wahl des jeweiligen Bewegungsmodells und die endgültige Objekt-Messung-Zuordnung. Damit ist die Grundlage für den Aktualisierungsschritt des Partikelfiltersystems und somit die Schätzung des nächsten Zeitschritts geschaffen. Die Organisation der einzelnen Partikelfilter und der Objekte mit ihren Trajektorien wird mit einer verbundwahrscheinlichkeitsbasierten Methode bewerkstelligt und in einem Verwaltungsschritt assoziiert und organisiert. Die Partikelevolution wird in unterschiedlichen Situationen mit unterschiedlichen Bewegungsmodellen gelenkt.

### 5. Datenfusion auf der Basis von Partikelfiltern

Die Idee der Analyse von bewegten Bildern dient dem grundlegenden Detektieren, Erkennen und Verfolgen von bewegten Objekten in Bildfolgen sowie zur umfassenden Beschreibung des Objektverhaltens. Dafür wurden bereits verschiedene leistungsfähige Verfahren entwickelt, die allerdings noch weit entfernt von der Vielseitigkeit und Zuverlässigkeit des visuellen Systems des Menschen sind. Dies ist vor allem darauf zurückzuführen, dass der Mensch es mit seinem visuellen System schafft, Bildinformation parallel, sowohl mit intensitäts- als auch mit merkmalsorientierten Verfahren zu extrahieren und die gewonnenen Daten intelligent in verschiedenen Abstraktionen miteinander zu fusionieren. Der Wunsch, die Objekte mit Hilfe technischer Systeme ähnlich wie beim biologischen System zu erkennen und zu verfolgen, hat mit der Leistungssteigerung der Bildaufnahme- und Rechentechnik in den letzten Jahren stetig an Bedeutung gewonnen. Allerdings liegt die Herausforderung insbesondere bei der Objektidentifikation, vor allem im Erhalt und in der Zuordnung der Trajektorien der Objekte, in Situationen mit gegenseitiger Interaktion und variablen Situationen des Umfeldes, wie es häufig bei Szenen mit dichtem Verkehr vorkommt. Somit konzentriert sich die Forschungsaufgabe in diesem Kapitel auf die Berechnung der Bewegungstrajektorien durch eine Datenfusion von intensitäts- und merkmalsbasierten Bilddaten, die durch eine geeignete Abstraktion des Multi-Objektverfolgungsproblems durch einen Partikelfilter erfolgen wird.

#### 5.1 Intuitive Einführung und Einordnung der eigenen Arbeiten

Im Bereich der Bildfolgenanalyse gibt es zahlreiche verschiedene Verfahren zum maschinellen Tracking von Objekten, die sich in intensitäts-, merkmals-, und modellbasierte Verfahren unterteilen lassen (s. Kapitel 2). Bei den intensitätsbasierten Verfahren werden ausschließlich zeitliche Intensitätswertänderungen in den Bildfolgen ausgewertet und sind somit für lokale, kontinuierliche Bewegungsschätzungen geeignet. Probleme verursachen nicht-kooperative Messsituationen, die zu starken Modifikationen der Intensitätswerte und damit zu einem Korrespondenzproblem führen können. Dazu kommt, dass diese Verfahren rechenaufwendig und für Echtzeitanwendung nicht geeignet sind (s. Kapitel 2).

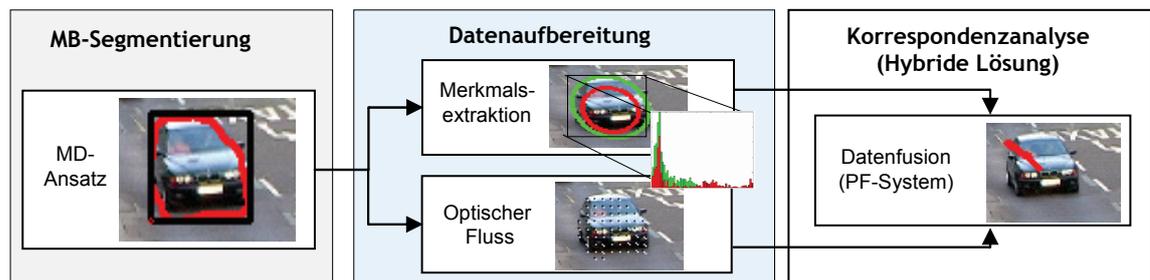
Im Gegensatz dazu weisen die merkmalsbasierten Ansätze (s. Kapitel 3) eine höhere Stabilität und Robustheit bei gestörten Bildsituationen oder größerer Pixelverschiebung auf. Dies ist zum einen auf die saubere Segmentierung der Objekte und die damit verbundene Reduzierung der Bildmerkmale auf die Bewegung, zum anderen auf ein sogenanntes Multi-Matching-Prinzip zurückzuführen. Immerhin gelten die Segmentierung, die Merkmalerfassung, die Merkmalszuweisung und die Ermittlung von optimalen Gewichten für die angewendeten merkmalspezifischen Ähnlichkeitsmaße in bildspezifischen Störsituationen (s. Kapitel 3) noch als schwierig. Dies führt dazu, dass es bei ähnlichen Objekten oft zu objektübergreifenden, falschen Merkmalszuordnungen kommt.

Ausgehend von diesen Überlegungen besteht die Herausforderung beim Multi-Objekt-Tracking mit herkömmlichen intensitäts- und merkmalsbasierten Verfahren demzufolge sowohl in der Datenzuordnung, Datenzuweisung als auch in der Zustandsschätzung

während der Konfliktsituationen und Objektinteraktionen. Um diese Defizite beseitigen zu können, wird in diesem Kapitel ein neuer Lösungsweg auf Grundlage der Datenfusion beschrieben, der eine Zustandsschätzung und eine Datenzuweisung von Objekten in Konfliktsituationen ermöglicht. Dafür nutzt die entwickelte Systemstruktur die Stärken der intensitäts- und merkmalsbasierten Verfahren so aus, dass die Schwächen beider Verfahren kompensiert werden können. Die Grundlage für die Zustandsschätzung und Datenzuweisung ist die bekannte Bayes-Schätzung (vgl. Kapitel 4), die Fuzzy-Logik und die bis jetzt vorgestellten Bewegungsanalyseverfahren (vgl. Kapitel 2, 3 und 4). Es wurde daraufhin eine Datenfusionsstruktur auf der Grundlage des Bayesfilters zur Multiobjektverfolgung entwickelt. Dafür wurden die Objektkonstellationen in realen Bildsequenzen untersucht und für die Datenfusion klassifiziert.

### 5.2 Hybride Systemstruktur zur Multi-Objektverfolgung

Für die hier entwickelte Systemstruktur zur Datenfusion wird von einem statischen Aufnahmesystem ausgegangen, das Farbbildsequenzen liefert (Abb. 5-1-1). Die zu analysierenden Szenen enthalten Objekte unterschiedlicher Art und unterschiedlichen Verhaltens, die sich größtenteils unabhängig von einander bewegen (vgl. Kapitel 4). Aufgrund dessen arbeitet diese Struktur objekt- und umgebungsunabhängig. Da starre, aber auch nicht starre Objekte auftreten können, ist eine Stabilität gegen Objektverformungen erforderlich.



**Abbildung 5-1-1:** Die vereinfachte Darstellung der Hybriden Systemstruktur zur Multi-Objektverfolgung in Farbbildsequenzen.

Um das Verständnis der Funktionalität und des Zusammenwirkens einzelner Komponenten zu erleichtern, wird die Systemstruktur zur Multi-Objektverfolgung (Abb. 5-1-1), die sich aus drei Schritten zusammensetzt, wie folgt beschrieben.

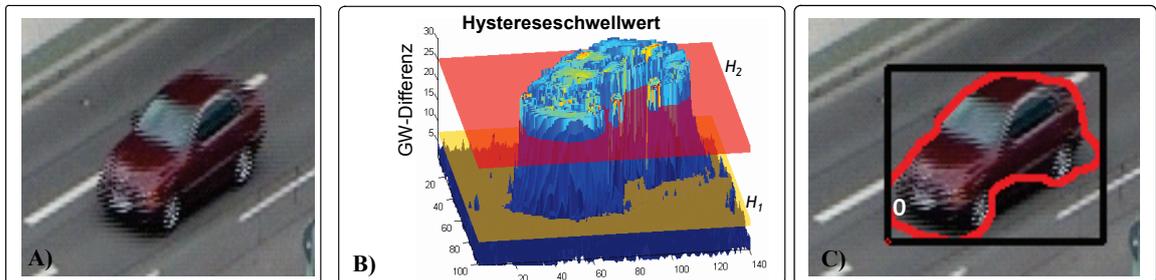
- 1). In dem Segmentierungsschritt werden alle Motion-Blobs (MBs) unabhängig von ihrer Art oder Form allein durch die aus der Bewegung resultierende Intensitätsänderung mittels des vorgestellten MD-Ansatzes in Verbindung mit dem Nachverarbeitungsalgorithmus erfasst und initialisiert (vgl. Kapitel 3).
- 2). Im Datenaufbereitungsschritt werden farb-, form- und bewegungsbasierte Merkmale aus den resultierenden MBs, die der Objektbeschreibung dienen, extrahiert.
- 3). Im letzten Schritt erfolgt die Korrespondenzanalyse, die die gewonnenen Merkmalsdaten verarbeitet und mittels Partikelfiltersystem fusioniert, um vor allem die Interaktion zwischen den Objekten zu interpretieren und so eine stabile Multi-Objektverfolgung zu gewährleisten.

Der letzte Analyseschritt berücksichtigt die Konfliktsituationen und stellt daher das Hauptaugenmerk in diesem Kapitel dar, in dem die Korrespondenzanalyse durch die Fusion der aus der Datenaufbereitung resultierenden Bilddaten erfolgt.

### 5.2.1 MB-Segmentierung

Für die automatische Erfassung bewegter Bildregionen wird der MD-Ansatz nach [Al-Hamadi 04c, 05d, vgl. Kapitel 3] eingesetzt, um die gesamten Informationen des Bildes optimal auf die Bewegung reduzieren zu können. Anstatt einfacher Schwellen für den MD-Ansatz zu verwenden, wurde hier ein Hystereseschwellwert angewendet. Dafür werden zwei Schwellwerte  $H_1$  und  $H_2$ , mit  $H_1 < H_2$  definiert. Die Grundidee dabei basiert auf ein Regionswachstum, das von den Bereichen mit maximaler Differenz  $H_2$  ausgeht, bis eine minimale Differenz  $H_1$  unterschritten ist. Der Schwellwert  $H_2$  wird so gewählt, dass er die Objekte erfasst und größer als sämtliche durch das Rauschen verursachte Differenzen ist (Abb. 5-2-1A).

Eine einfache Schwellwertbinarisierung mit  $H_2$  würde dann das Rauschen unterdrücken. Es würden jedoch große Löcher hervorgerufen werden, die durch texturarme Regionen in den Objekten entstehen. Der Schwellwert  $H_1$  soll so gewählt werden, dass er diese texturarmen Regionen erfasst. Eine einfache Schwellwertbinarisierung mit  $H_1$  würde dann alle wichtigen Regionen erfassen, allerdings würde Rauschen die nachverarbeitenden Schritte stören und so z.B. Ränder ausfransen lassen oder Objekte miteinander verbinden.



**Abbildung 5-2-1:** MD-Ansatz mit den unterschiedlichen Schwellwerten  $H_1$  und  $H_2$  als Schrittaben eingetragen.

Durch den MD-Ansatz wird die Bildinformation durch eine Differenzmaske reduziert und der Großteil der Störungen entfernt. Die Differenzmaske wird durch die Hintergrundmodellierung und morphologische Operatoren nachbearbeitet, so dass Löcher und zerklüftete Regionen, die sich durch schwache Texturen gebildet haben, geschlossen werden (vgl. Kapitel 3). Die Zuweisung von getrennten Regionen in der resultierenden Bewegungsmaske an einer Datenstruktur wird als Labeling bezeichnet. Die daraus resultierenden Regionen werden Motion-Blobs genannt und liegen als Träger für die Merkmalsextrahierung und Zustandbeschreibung in dem Datenaufbereitungsschritt vor und stellen die Knoten in dem Zustandsmodell dar.

### 5.2.2 Datenaufbereitung

Durch die Datenaufbereitung sollen die in den zu erfassenden MBs vorhandenen relevanten Informationen auf einen Merkmalsraum mit niedriger Dimension abgebildet werden. Infolgedessen werden einfache Merkmale verwendet, die keine explizite

Repräsentation der Region benötigen und sich direkt aus dem MB (auch Knoten genannt) berechnen lassen. Zu ihnen gehören das Bewegungsvektorfeld  $\mathbf{V}$ , das Farbhistogramm  $\mathbf{h}$ , die umgebende Ellipse  $f$ , die prädizierte Ellipse  $f'$  und der Flächeninhalt  $A$ , die sich kurz wie folgt beschreiben lassen:

**Bewegungsvektorfeld:** Für die Berechnung des Bewegungsvektorfeldes  $\mathbf{V}$  wird der im Kapitel 2 vorgestellte pyramidale Lucas-Kanade-Algorithmus (s. PLK-Algorithmus) mit der Adaption gemäß Kapitel 2 angewendet (Abb. 5-2-2A). Verglichen mit dem HS-Verfahren im Kapitel 2 zeichnet sich der adaptierte PLK-Algorithmus durch eine hohe Robustheit gegenüber Rauschen und größere Pixelverschiebungen aus. Darüber hinaus wird der Vektorfeldausbreitungseffekt durch den MD-Ansatz minimiert (s. Kapitel 2). Die Parameter des Algorithmus bestehen aus der Größe des Abtastfensters und der Anzahl der Level der Bildpyramide. Um den Aufwand für die Berechnung des Bewegungsvektorfeldes  $\mathbf{V}$  über den PLK-Algorithmus zu reduzieren, wird die MB-Maske gitterförmig, mit einer Größe von  $d_{grid}$  abgetastet und diejenigen Vektoren berechnet, die auf den Abtastpunkten des Gitters liegen. Um auch für kleine Objekte eine robuste Verfolgung gewährleisten zu können, darf das Vektorfeld nicht weniger als 5 Verschiebungsvektoren beinhalten. Dies erfordert eine adaptive Anpassung<sup>92</sup> der Gittergröße  $d_{grid}$  auf die Fläche  $A$  des MBs (Abb. 5-2-2A).

**Farbhistogramm:** Als weiteres wichtiges Merkmal dient der Vergleich der Farbinformation der erfassten MBs, damit Objekte einander zugeordnet werden können, die zeitlich länger als ein Bild auseinander liegen. Das Farbmerkmal, welches benutzt wird, ist das Farbhistogramm  $\mathbf{h}$ , das aus der RGB-Maske eines MBs berechnet wird. Dabei besteht ein Histogramm aus einem Vektor, der für jede Quantisierungsstufe ein Element besitzt. Jedes Element enthält die Anzahl der quantisierten Farbwerte, die mit dem Index des Elementes übereinstimmen. Ein Element in dem Histogramm wird als *Urne*  $U$  bezeichnet, da es bildlich gesehen, die Menge der Pixel, die in diese Quantisierungsstufe fallen, beinhaltet. Um Zusammenhänge zwischen Histogrammen herausfinden zu können, existieren verschiedene Ähnlichkeitsmaße<sup>93</sup> in unterschiedlichen Farbräumen. Der einfachste Weg ist die Interpretation der Histogrammdata, als Vektoren (Abb. 5-2-2B). Der Abstand dieser Vektoren ist dann durch die  $L_\theta$ -Norm oder auch *Minkowski-Abstand* gegeben und berechnet sich aus den relativen Häufigkeiten  $h_{rel1}$  und  $h_{rel2}$  durch Gleichung (5-2-1a).

$$MIN_\theta(h_{rel1}, h_{rel2}) = \left( \sum_{u=0}^{U-1} |h_{rel1}(u) - h_{rel2}(u)|^\theta \right)^{\frac{1}{\theta}} \quad (5-2-1a)$$

Die Norm  $L_2$  ergibt sich für  $\theta=2$  und entspricht dem Euklidischem Abstand von  $h_{rel1}$  und  $h_{rel2}$  im  $U$ -dimensionalem Raum. Ein weiteres Abstandsmaß ist der *Histogrammschnitt*, bei dem die relativen Häufigkeiten als Flächen betrachtet werden. Dabei wird die Ähnlichkeit über die gemeinsamen Flächen der Histogramme berechnet.

$$HS(h_{rel1}, h_{rel2}) = 1 - \left( \sum_{u=0}^{U-1} \min(h_{rel1}(u), h_{rel2}(u)) \right) \cdot \left( \sum_{u=0}^{U-1} h_{rel2}(u) \right)^{-1} \quad (5-2-1b)$$

<sup>92</sup> Gittergröße wird in Abhängigkeit von der Fläche eingestellt. Wenn die Fläche bspw. kleiner als 25 Pixel sind, ergibt sich die Gittergröße aus der Wurzel der Fläche durch maximal geforderten Pixel (z.B.5).

<sup>93</sup> Diese hier getroffene Auswahl orientiert sich an den Arbeiten von [Paulus 01, Knauer 05].

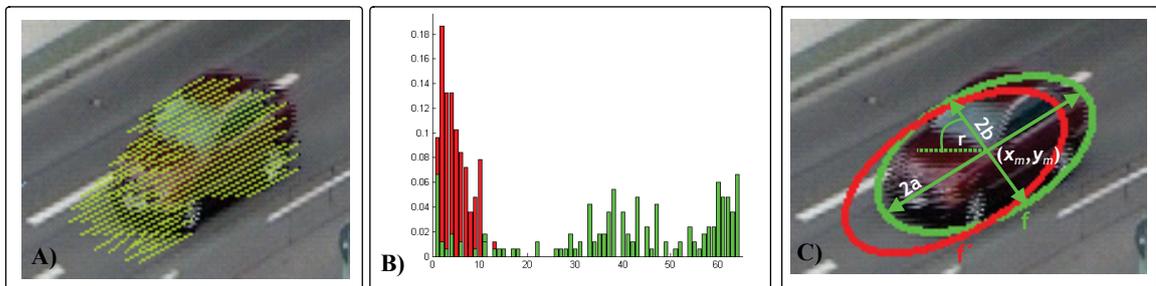
## 5. DATENFUSION AUF DER BASIS VON PARTIKELFILTERN

Da man mehrere Farbkanäle verwendet, werden die Histogramme für die entsprechenden Kanäle erzeugt und einzeln durch Histogrammdistanzen verglichen. Die Kombination der Histogrammdistanzen der einzelnen Kanäle wird in der Regel durch Euklidische Distanzen erfolgen.

**Formmerkmale:** Die *Formmerkmale* bilden den zweiten Teil der merkmalsbasierten Daten in der Gleichung (5-2-2). Der Anspruch, der an die Formmerkmale gestellt wird, ist, dass sie die grundlegenden Positions- und Formeigenschaften der Knoten wiedergeben und nicht sensitiv auf ein Pixelrauschen oder geringe Segmentierungsfehler der Bewegungsmaske reagieren sollen. Aus diesen Gründen wurde eine elliptische Darstellung für die Motion-Blobs gewählt (Abb. 5-2-2C). Das Fitting der Ellipse auf die Bewegungsmaske eines MBs erfolgt wie üblich in zwei Schritten. Zunächst erfolgt ein grobes Fitting durch die Bestimmung der Boundingbox und der Orientierung/Ausrichtung eines resultierenden MBs. Anschließend findet eine Optimierung/feines Fitting in Anlehnung an [Halir 98] durch die Methode im Sinne der kleinsten Quadrate statt. Eine Möglichkeit zur Bestimmung der Orientierung ist die Hauptkomponentenanalyse, bei der die Komponenten in Richtung größter Varianzen ausgerichtet werden. Eingangsdaten sind dafür die Positionen der erfassten Bildregion. Die Parameter, die die Form eines MBs beschreiben, werden durch den Vektor  $f$  zusammengefasst:

$$f = (x_m \quad y_m \quad a \quad b \quad r)^T \quad (5-2-2)$$

wobei  $a$ ,  $b$  die Halbachsen,  $x_m$ ,  $y_m$  der Mittelpunkt und  $r$  die Ausrichtung der Ellipse sind.



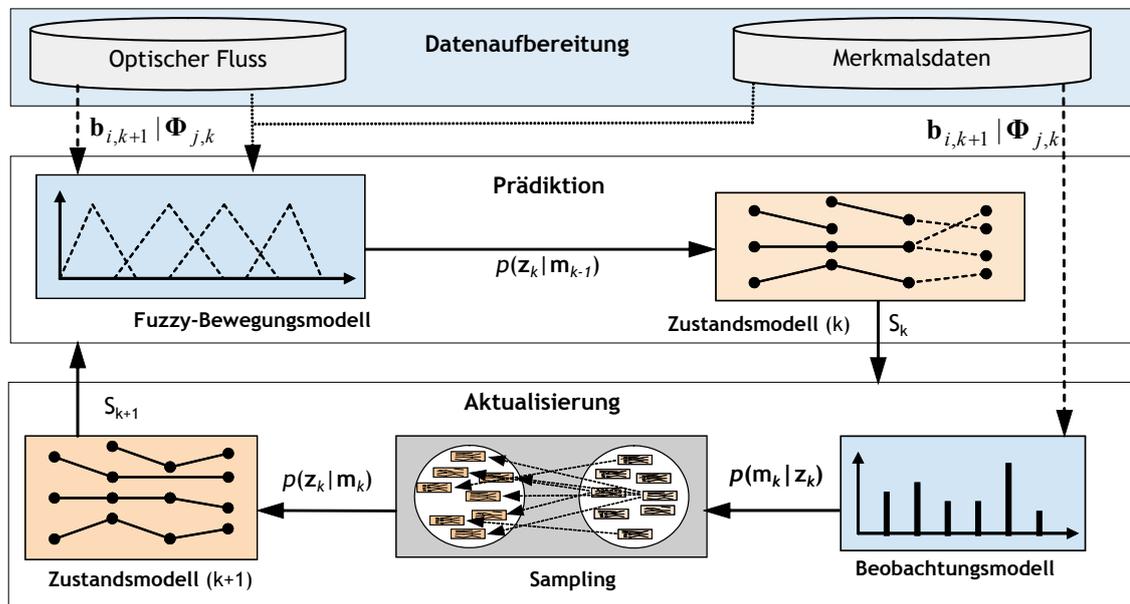
**Abbildung 5-2-2:** Darstellung des berechneten Verschiebungsvektorfeldes (A), die Farbhistogramme (B) und Formmerkmale des MBs durch Ellipsenparameter der Ellipse  $f$  und der prädizierte Ellipse  $f'$  durch das Verschiebungsvektorfeld (C).

Die prädizierte Verschiebung der MB-Fläche lässt sich als prädizierte Ellipse  $f'$  aus dem eingeschätzten Verschiebungsvektorfeld  $V$  eines MBs ermitteln, indem die Optimierung der Ellipsengleichung für die durch das Verschiebungsfeld bestimmten verschobenen Punkte berechnet wird. Die prädizierte Ellipse für jeden Knoten wird dann analog zur Gleichung 5-2-2 beschrieben. Die Abbildung 5-2-2C zeigt an einem Beispiel die Ellipsen  $f$  und  $f'$ , durch die die Form des MBs und die prädizierte Form des MBs in einer realen Bildszene repräsentiert werden. Diese beschriebenen Merkmale stellen nun die Ausgangsdaten für das Zustandsmodell dar.

### 5.3 Korrespondenzanalyse

Der letzte Schritt in Abbildung 5-1-1 beinhaltet die Korrespondenzanalyse und stellt daher das Hauptaugenmerk in diesem Kapitel dar. Die Korrespondenzanalyse erfolgt durch die Fusion der bewegungs- und merkmalsbasierten Daten auf der Grundlage vom Partikelfiltersystem.

In der Datenfusion liegt die Überlegung auf die korrekte Verfolgung von Objekten durch die Lösung von Konfliktsituationen im Sinne der Zuordnung, wie sie öfter durch Teilverdeckung, den Zerfall und die Verschmelzung von MBs entstehen können. Deshalb wird das MB-Matching als fuzzy-stochastischer Prozess modelliert, in dem der Zustand als stochastische Variable definiert wird, der die Trajektorien enthält. Dies führt dazu, dass die zeitliche Entwicklung dieser Zustandsvariablen durch einen Markov-Prozess erster Ordnung aufgrund der Verknüpfung mit dem letzten Zeitschritt beschrieben werden kann, wobei die Übergangsfunktion eines Zustandes in den nächsten als nichtlinear und nichtstetig aufgrund des Matchingsprozesses beschrieben werden muss. Dafür eignet sich der Partikelfilter, der durch seine einfache Struktur (Prädiktion und Aktualisierung) die Möglichkeit bietet, die Merkmale so zu fusionieren, dass sie sich gegenseitig überwachen und voneinander teilweise unabhängig sind. Dabei gehen die intensitätsbasierten Daten in Form eines optischen Flussfeldes durch das Bewegungsmodell und die merkmalsbasierten Daten in Form eines Farbhistogramms und Ellipsen durch das Beobachtungsmodell getrennt in die Objektverfolgung ein (Abb. 5-3-1). Dies hat den Vorteil, dass die Fehler der Merkmalsdaten voneinander *relativ unabhängig* sind, d.h. Fehler bei den intensitätsbasierten Berechnungen des optischen Flusses haben geringfügigen Einfluss auf die Fehler bei der Merkmalsextraktion.



**Abbildung 5-3-1:** Vereinfacht schematische Darstellung des Partikelfiltermodells für die Datenfusion aus der Datenaufbereitung, welches hier ein Zustand-, Bewegungs- und Beobachtungsmodell beinhaltet.

Die Abbildung 5-3-1 zeigt die schematische Darstellung des Partikelfiltermodells für die Datenfusion, welches ein Zustands-, Bewegungs- und Beobachtungsmodell beinhaltet. Im Unterschied zu dem in Kapitel 4 vorgestellten Partikelfiltersystem und

herkömmlichen Systemen [Schwarzenberg 05, Sidenbladh 03, Vermaak 03, Al-Hamadi 07a, Al-Hamadi 08a] wird hier ein Bewegungsmodell auf der Grundlage eines Fuzzy-Systems angewendet, um die Begrenztheit eines linearen Bewegungsmodells zu überwinden. Dies wird anschließend als Fuzzy-Bewegungsmodell bezeichnet. Motiviert wird diese Art der Fusion folgendermaßen:

- Im Zustandsmodell wird ein MB in jedem Bild durch einen Knoten repräsentiert, der durch die zeitliche Verbindung in der Bildsequenz eine Bewegungstrajektorie bildet. Diese Trajektorie ist nun Bestandteil des Zustandes<sup>94</sup>  $\mathbf{z}_k$ , der sämtliche Verbindungsmöglichkeiten der MBs, die sich aus dem Bewegungsmodell ergeben, zu einem bestimmten Zeitpunkt enthält. Somit beinhaltet der Zustand  $\mathbf{z}_k$  die Objekttrajektorien mit den möglichen Variationen ( $n$ ), die als Zufallsvektor modelliert werden. Der Zufallsvektor wird durch eine Partikelmenge  $S_k$  repräsentiert, wobei ein Partikel  $s$  aus der Partikelmenge  $S_k$  einen möglichen Zustand aus dem Zustandsvektor beinhaltet. Die Beobachtung des Systems liefert die Messung  $m$  mit dem Messvektor  $\mathbf{m}=(\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_n)^T$ .
- Die Aufgabe des Partikelfilters ist nun die Schätzung der A-posteriori-Verteilungsdichte  $p(\mathbf{z}_k|\mathbf{m}_k)$  des Systemzustandes  $\mathbf{z}$  aus den beobachteten Systemmessungen  $\mathbf{m}_k$ . Hierfür wird in der *Prädiktion* die A-priori-Wahrscheinlichkeit  $p(\mathbf{z}_k|\mathbf{m}_{k-1})$  rekursiv aus dem Bewegungsmodell  $p(\mathbf{z}_k|\mathbf{z}_{k-1})$  und dem vorherigen Zeitschritt  $p(\mathbf{z}_{k-1}|\mathbf{m}_{k-1})$  ermittelt. Die Entwicklung des Zustands  $\mathbf{z}_k$  aus dem vorhergehenden Zustand  $\mathbf{z}_{k-1}$  erfolgt nun durch das Fuzzy-Bewegungsmodell (Abb. 5-3-1), wobei die Übergangsfunktion, die aus dem alten Zustand den neuen Zustand berechnet, in Abhängigkeit von der Bewegungsschätzung durch den optischen Fluss steht. Zusätzlich wird auf Merkmalsdaten zurückgegriffen in den Situationen, in denen eine direkte Bild-zu-Bild Zuordnung nicht möglich ist, wie z.B. bei der Trennung von verschmolzenen Objekten (Abb. 5-3-1). Die Übergangsfunktion des MBs-Matching wird durch ein Fuzzy-System realisiert, da sie ein effektives Werkzeug bildet, um Expertenwissen algorithmisch umzusetzen. Dadurch ist der Zusammenhang zwischen zwei aufeinander folgenden Zuständen  $\mathbf{z}_{k-1}$  und  $\mathbf{z}_k$  mit einem nichtlinearen und auch nicht stetigen Bewegungsmodell beschrieben.
- Die A-posteriori-Wahrscheinlichkeit  $p(\mathbf{z}_k|\mathbf{m}_k)$  kann im zweiten Schritt mit der Prädiktion  $p(\mathbf{z}_k|\mathbf{m}_{k-1})$  und der Likelihood-Funktion  $p(\mathbf{m}_k|\mathbf{z}_k)$  durch den Satz von Bayes ermittelt werden. Daraus folgt der *Aktualisierungsschritt* für die Messung  $\mathbf{m}_k$ . Die Berechnung der Likelihood-Funktion  $p(\mathbf{m}_k|\mathbf{z}_k)$  resultiert aus dem *Beobachtungsmodell*, welches durch ein Ähnlichkeitskriterium auf der Grundlage der erfassten merkmalsbasierten Daten definiert wird (Abb. 5-3-1). Die Likelihood wird durch die Gewichte  $\pi_k^n$  der  $n$  Partikel approximiert und erzeugt in Verbindung mit dem Sampling-Schritt aus der alten Partikelmenge  $S_k$  eine neue Partikelmenge  $S_{k+1}$ , deren Wahrscheinlichkeitsdichte der A-posteriori-Wahrscheinlichkeit  $p(\mathbf{z}_k|\mathbf{m}_k)$  entspricht.
- Ein wesentliches Problem bei den Partikelfiltern ist das *Degeneracy Problem*, das dazu führt, dass ein großer Teil des Rechenaufwandes dazu verschwendet werden muss, um Partikel, deren Auftrittswahrscheinlichkeit gegen Null tendiert, zu aktualisieren. Eine Möglichkeit dieses Problems zu umgehen, ist der Condensation-Algorithmus (vgl. Kapitel 4), der in diesem Kapitel angewendet wird.

---

<sup>94</sup> Der Zustand der Trajektorien kann hier als eine Menge betrachtet werden.

Eine genaue Betrachtung der einzelnen Verarbeitungsschritte und das Zusammenwirken der einzelnen Komponenten der schematischen Darstellung des Partikelfiltermodells für die Datenfusion in Abbildung 5-3-1 findet in den folgenden Abschnitten statt.

### 5.3.1 Zustandsmodell

Zur Beschreibung des Zustandsmodells ist die Übertragung der Verfolgung der Objekte auf die Abstraktionsschicht der MBs erforderlich. Dafür gibt es, vor allem in Konfliktsituationen, zwei übliche Strategien, die als Straight-Through-Ansatz (ST-Ansatz<sup>95</sup>) und Merge-Split-Ansatz (MS-Ansatz) zu beschreiben sind.

Unter der Verwendung des MS-Ansatzes ist die Anzahl der Freiheitsgrade in der Verkehrsanalyse erheblich geringer und ist abhängig von den sich in Konflikt befindenden Objekten. Beispielsweise entsprechen bei dem Segment-Matching durch die Berücksichtigung eines ST-Ansatzes die Freiheitsgrade den verschiedenen Möglichkeiten, wie Segmente miteinander verbunden werden können. Diese können schon bei wenigen Objekten enorm groß werden. Im Gegenteil dazu genügt durch den MS-Ansatz bereits eine Partikelmenge mit wenigen Partikeln<sup>96</sup>, um die Wahrscheinlichkeitsverteilung des Zustandsvektors zu approximieren.

Hier wird im Unterschied zu Kapitel 4 erst der Zustand  $z_k^n$  eines Partikels  $n$  durch die Objekttrajektorien  $(T_{j,k})$  mit  $N_T$  der Anzahl der Trajektorien beschrieben, die die zeitliche Verknüpfung einzelner MBs beinhalten. Diese werden als Knoten nach dem Prinzip des MS-Ansatzes betrachtet. Als Konvention für die Beschreibung der Eingangsdaten werden die Trajektorien aus dem momentanen Zustand  $z_k$  als Vorgeschichte und die MB als Messung ( $\mathbf{m}_k$ ) bezeichnet (s. Abb. 5-4-2).

$$z_k^n = \{T_{0,k}, \dots, T_{N_T-1,k}\} \quad (5-3-1)$$

Zusätzlich zur Menge der Knoten  $\Phi_{j,k}$  gibt es weitere Attribute für die Beschreibung einer Objekttrajektorie  $(T_{j,k})$ : der *Aktivitätsgrad* einer Trajektorie ( $a_{j,k}$ ) und die Menge der Indizes der Eltertrajektorie ( $T_{j,k}^P$ ). Dabei gibt der Aktivitätsgrad an, inwieweit eine Trajektorie an den Konfliktoperationen teilnehmen kann. Hier wird beispielsweise der Aktivitätsgrad einer Trajektorie auf Null gesetzt und eine neue Trajektorie erzeugt, wenn sie mit anderen Trajektorien verschmolzen wird. Aus diesen Attributen lässt sich nun eine Trajektorie (Gl. 5-3-2) beschreiben, die aus Knoten besteht. Diese enthalten zusätzlich die Position eines Objektes in der Bildsequenz, das Bewegungsvektorfeld  $\mathbf{V}$ , das Farbhistogramm  $\mathbf{h}$ , die umgebende Ellipse  $f$ , die prädizierte Ellipse  $f'$  und den Flächeninhalt  $A$  (s. Tupel 5-3-3). Der Unterschied zwischen einem Trajektorieknoten- $\Phi_{j,k-1}$  und Messknoten  $\mathbf{b}_{j,k}$  in Gleichung (5-3-3) besteht nur darin, dass der Messknoten noch keiner Trajektorie zugeordnet wurde.

<sup>95</sup> Der ST-Ansatz beinhaltet die Zuordnung eines MB zu genau einer Trajektorie, wobei eine Trajektorie genau ein Objekt beinhaltet. Der Aufwand der Operationen (z.B. Erzeugen) zwischen den Trajektorien ist dadurch auf ein Minimum reduziert. Bei Konfliktsituationen müssen aber die Objektpositionen innerhalb des MB erkannt und verfolgt werden. Das setzt eine gute Objektverfolgung innerhalb der MB voraus, die nur mit sehr hohem Aufwand aufrechterhalten werden kann, so dass die Multi-Objektverfolgung sehr oft in diesen Abschnitten instabil wird.

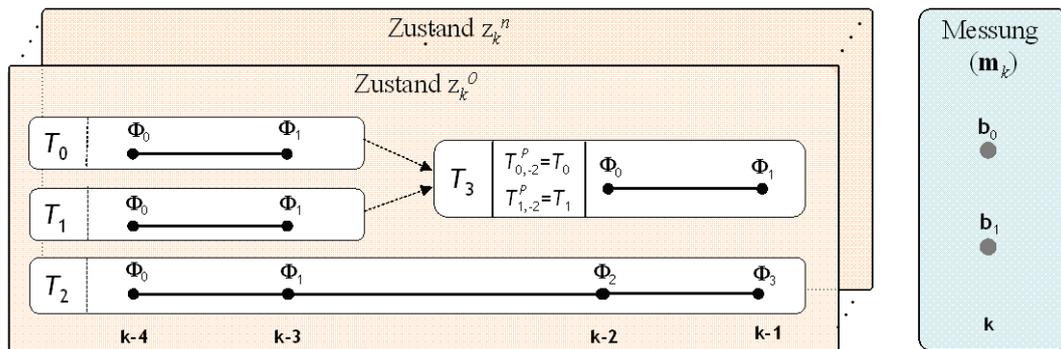
<sup>96</sup> Aufgrund des geringeren Freiheitsgrads benötigt man wenige Partikel (in unserem Fall 100), um die stochastischen Eigenschaften der Variationen zu beschreiben. Die Eigenschaften werden durch die Wahrscheinlichkeit der Variationen ausgedrückt.

## 5. DATENFUSION AUF DER BASIS VON PARTIKELFILTERN

$$T_{j,k} = \{a_{j,k}, T_{j,k}^P, \Phi_{j,k}\} \quad \text{mit } j=0, \dots, N_T-1, \forall T_{j,k} \in z_k^n \quad (5-3-2)$$

$$\mathbf{b}_{i,k} | \Phi_{j,k-1} = \{\mathbf{V}, \mathbf{h}, f, f', \mathbf{A}\} \quad (5-3-3)$$

Dabei ist  $\Phi_{j,k}$  in der Gleichung (5-3-2) die Menge der Knoten mit  $\Phi_{j,k} = \{\Phi_0, \dots, \Phi_{N_k-1}\}^{97}$  (mit  $N_k$  der Anzahl der Trajektorieknoten) und  $T_{j,k}^P$  die Menge der Indizes der Elterntrajektorie mit  $T_{j,k}^P = \{T_{0,k}^P, \dots, T_{N_{parent}-1,k}^P\}$  (mit  $N_{parent}$  der Anzahl der Elterntrajektorien zum Zeitpunkt  $k$ ). Durch diese Attribute lässt sich jedes Objekt in einer Trajektorie, durch die Elterntrajektorien einfach zurückverfolgen.



**Abbildung 5-3-2:** Darstellung der Struktur der Trajektorien. Die Trajektorien  $T_{j,k}$  setzen sich aus den Trajektorieknoten  $\Phi_{l,k}$ , den Elterntrajektorien  $T_{0,k}^P$  und den Messungen  $(\mathbf{b}_{i,k})$  zusammen.

### 5.3.2 Fuzzy-Bewegungsmodell

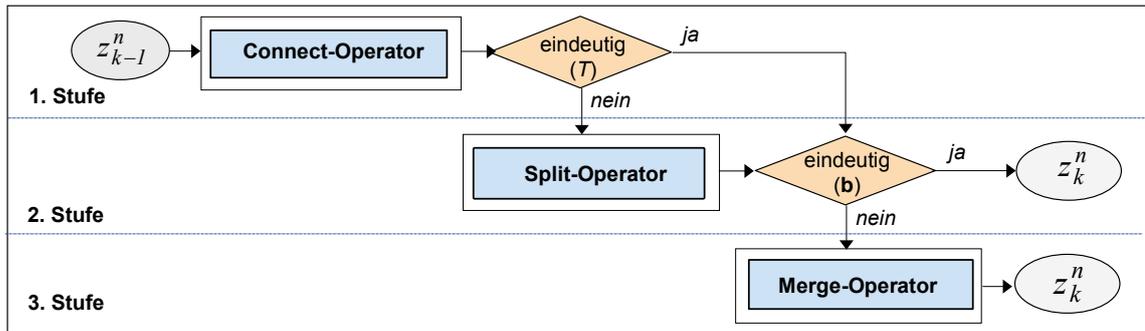
Das Bewegungsmodell definiert die Weiterentwicklung und die Variation des Zustands  $z_k^n$  unter der Einbeziehung der Bewegungs- und der Farbdaten aus dem vorhergehenden Zustand  $z_{k-1}^n$  für alle  $n$  Partikel. Dieser Vorgang entspricht der Prädiktion des Partikelfilters (Abb. 5-3-1). Dabei erfolgt die Entwicklung des Zustandes  $\mathbf{z}_k$  durch die Transitionsgleichung (5-3-3) mit  $f$  der Beschreibung des Bewegungsmodells, wobei die Variation des Zustandes durch einen normalverteilten Rauschprozess  $N(\mu, \sigma^2)$  gewährleistet wird.

$$\text{F-BM:} \quad z_k^n = f(z_{k-1}^n, N(\mu, \sigma^2)) \quad (5-3-3)$$

Die Weiterentwicklung des Zustands wird durch die Zuordnung der MBs zu den Trajektorien in drei Verarbeitungsstufen durch die Operatoren *Connect*, *Split* und *Merge* erfolgen (Abb. 5-3-3a). Ausgehend von einer Gatingmatrix (vgl. Kapitel 4), welche die Zuordnung der Trajektorien ( $T$ ) zu den Messungen ( $\mathbf{b}$ ) beinhaltet, werden Entscheidungen bezüglich der Verwendung der Operatoren getroffen. Die möglichen Verbindungen werden zunächst durch den Connect-Operator bestimmt. Werden einer Trajektorie mehrere MBs in der Gatingmatrix zugeordnet, dann wird darauf folgend der Split-Operator aktiviert. Durch ihn werden die MBs den enthaltenen Objekten der Trajektorie zugeordnet.

<sup>97</sup> Die Darstellung der Zugehörigkeit eines Knotens, oder einer Elterntrajektorie zu einer Trajektorie wird in verkürzter Form statt mit  $a_j \in T_{j,k}$ ,  $\Phi_l \in T_{j,k}$  und  $T_o^P \in T_{j,k}$  mit  $a(T_{j,k})$ ,  $\Phi_l(T_{j,k})$  und  $T_o^P(T_{j,k})$  angegeben.

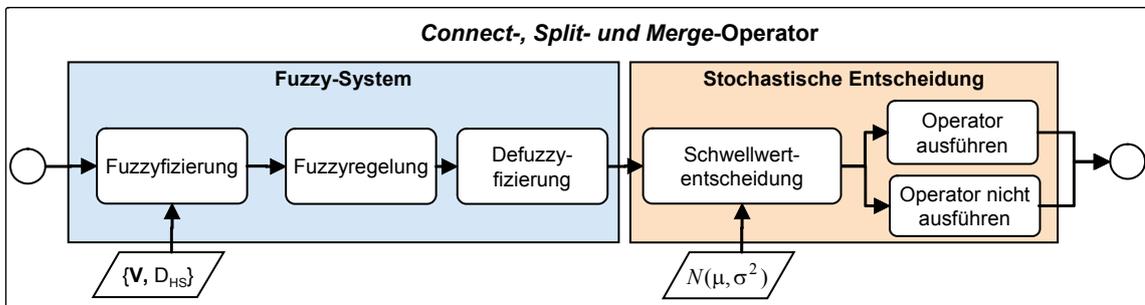
## 5. DATENFUSION AUF DER BASIS VON PARTIKELFILTERN



**Abbildung 5-3-3a:** Ablaufdiagramm des Bewegungsmodells. Der Zuordnung der MB zu den Trajektorien findet in drei Stufen über die Operatoren *connect*, *split* und *merge* statt.

Werden einem MB mehrere Trajektorien zugeordnet, so wird der Merge-Operator aufgerufen, der die Verschmelzung der Trajektorien im MB verwaltet. Die Eingangsdaten der Operatoren werden in jeweils zwei Stufen verarbeitet (Abb. 5-3-3b).

- Die erste Stufe stellt eine Fuzzyregelung dar, in der die Bewegungs- und Farbdaten durch ein Fuzzy-System nach dem Konzept des Mamdani-Assini-Reglers<sup>98</sup> [Michels 03] so verarbeitet werden, dass über ein Fuzzyregelwerk anhand der Eingangsdaten ein Zugehörigkeitsgrad als Regelausgang erzeugt wird, der die weitere Durchführung der Operationen Connect, Merge oder Split bestimmt. Die Grundidee dieser Regelung besteht darin, eine unscharfe Funktion stückweise über eine wissenbasierte Regelbasis und „so genannte linguistische Variablen“ zu definieren. Der Fuzzyregelausgang für einen Operator gibt einen Zugehörigkeitsgrad an, inwieweit die jeweilige Operation für ein MB-Trajektorie Paar ausgeführt werden soll.



**Abbildung 5-3-3b:** Der Ablauf für die Ausführung eines Operators für ein MB-Trajektorien-Paar besteht aus einem Fuzzy-System nach dem Prinzip von Mamdani-Assini-Regler und einer stochastischen Entscheidung. Durch diesen Schritt erhält das Bewegungsmodell seine Variation.

- In der zweiten Stufe „*stochastische Entscheidung*“ wird die Entscheidung getroffen (s. Abb. 5-3-3b), ob ein Operator ausgeführt wird. Diese wird über einen Schwellwert für den defuzzifizierten Regelausgangswert realisiert. Dabei besitzt in dieser Stufe der Schwellwert die Eigenschaft einer Zufallsvariable mit der Verteilungsdichtefunktion  $N(\mu, \sigma^2)$  und wird als stochastischer Schwellwert bezeichnet. Dadurch wird die Variation des Bewegungsmodells erreicht. Abhängig von der Wahrscheinlichkeitsverteilung der stochastischen Schwelle können somit auch Operationen mit geringen

<sup>98</sup> Ein sogenannter Mamdani-Regler basiert auf linguistischen Regeln des Typs (R1: **Wenn**  $x_1 = A_1$  **und**  $x_2 = A_2$ , **dann**  $y = B$ ). D.h., Prämisse und Konklusion der Regel sind mit Hilfe von linguistischen Termen definiert. Alternativ sind auch Regler (z.B. Sugeno-Regler) denkbar, bei denen sich entweder die Prämisse oder die Konklusion auf scharfe (z.B. reelle) Werte beziehen.

Zugehörigkeitsgrad ausgeführt werden. Somit wird die Übergangswahrscheinlichkeit  $p(\mathbf{z}_k | \mathbf{z}_{k-1})$  des Markovprozesses durch die Überlagerung der stochastischen Schwellwerte aller Operationen, die zu dem Zustand  $\mathbf{z}_k$  führen, bestimmt.

Das grundlegende Verhalten der Operatoren ist also bestimmt durch die Reglerübertragungsfunktion des Fuzzy-Systems (Abb. 5-3-3b) und den Rauschanteil  $N(\mu, \sigma^2)$ , der die Entscheidung über das Ausführen eines Operators bestimmt. Im Folgenden werden die anschließenden Schritte des angewendeten Fuzzy-Systems Fuzzifizierung, Fuzzyregelung und Defuzzifizierung detailliert beschrieben.

### 5.3.2.1 Fuzzifizierung

Grundsätzlich sind die Fuzzy-Sets und Fuzzy-Regeln von der Applikation abhängig und müssen an jedes System und seine Gegebenheiten dementsprechend angepasst werden<sup>99</sup>. In dieser Arbeit erfolgt die Fuzzifizierung der Eingangsdaten für die Fuzzy-Regelung über vier linguistische Variablen (LV)<sup>100</sup>, die den Fuzzyaktivitätsgrad, das Fuzzybewegungsvektormatching und das Fuzzyfarbmatching beschreiben.

I) Die erste linguistische Variable „Fuzzyaktivitätsgrad“ bestimmt, ob eine Trajektorie gültig oder ungültig ist und wird mit  $a(T_{j,k-1})$  bezeichnet. Ungültige Trajektorien können z.B. Störungen, die durch die Bewegungssegmentierung (z.B. Luftflimmern) hervorgerufen werden, beinhalten. Dieser Aktivitätsgrad ergibt sich für eine Trajektorie aus der Gleichung (5-3-4a).

$$a(T_{j,k-1}) = 1 - \frac{\tau_{Re}}{\beta_1 \cdot N_k} \quad (5-3-4a)$$

Dabei entspricht  $\tau_{Re}$  der Anzahl der Zeitschritte seitdem die Trajektorie  $T_{j,k-1}$  nicht aktualisiert wurde, und  $N_k$  ist die Anzahl der Knoten. Kurze Trajektorien weisen einen niedrigen Aktivitätsgrad auf, wenn sie längere Zeit nicht mehr aktualisiert wurden (s. Abb. 5-3-4b). Auf diese Weise wird die Anteilnahme einer Trajektorie an den Regelprozessen eingestellt.

Die Länge der Zeit wird mit der Konstanten  $\beta_1$  eingestellt und ist in unserem Fall empirisch mit  $\beta_1=2$  festgelegt. Ausgehend davon wird die linguistische Variable des Aktivitätsgrades durch zwei unscharfe Mengen repräsentiert. Die Menge  $A_{LV}(ja)$ , die dem Ausdruck, dass die Trajektorie aktiv ist, entspricht und der Menge  $A_{LV}(nein)$ , die für den Ausdruck steht, dass die Trajektorie nicht aktiv ist. Die LV wird aus dem Fuzzy-Mengensystem gemäß Gleichung (5-3-4b) gebildet, die in Abbildung 5-3-4 dargestellt sind.

$$A_{LV} = \{\mu_R^{(1)}(a) = A_{LV}(ja), \mu_R^{(2)}(a) = A_{LV}(ne)\} \quad (5-3-4b)$$

<sup>99</sup> Die hier verwendeten Sets und Regeln wurden empirisch in Versuchen mit dem Objektverfolgungssystem im IESK ermittelt und angepasst. Dafür wurden mehr als 20 reale Sequenzen aus einer stationären Kamera untersucht. Beispielsweise wurde eine Datenbasis (43 PKW in 5 unterschiedlichen Positionen) betrachtet, um das Fuzzset für die Farbkorrelation zu bestimmen. Daher geht es in dieser Arbeit darum, neue Wege bei der Datenzuordnung zu erforschen und nicht um ein komplettes echtzeitfähiges System zur Korrespondenzermittlung in Bildsequenzen zu entwickeln.

<sup>100</sup> Linguistische Variablen (Terme) dienen insbesondere dazu, linguistisch formuliertes Wissen (z.B. von Experten, wenn man ein Expertensystem modellieren möchte) mit all seinen Unschärfen angemessen in formale Sprachen zu übersetzen, wobei möglichst wenig des Reichtums menschlicher Sprache verloren gehen und gleichzeitig das Wissen mit rechnerbasierten Informationssystemen verarbeitet werden soll. Es handelt sich also um eine sprachliche Bezeichnung für technische Größen.

## 5. DATENFUSION AUF DER BASIS VON PARTIKELFILTERN

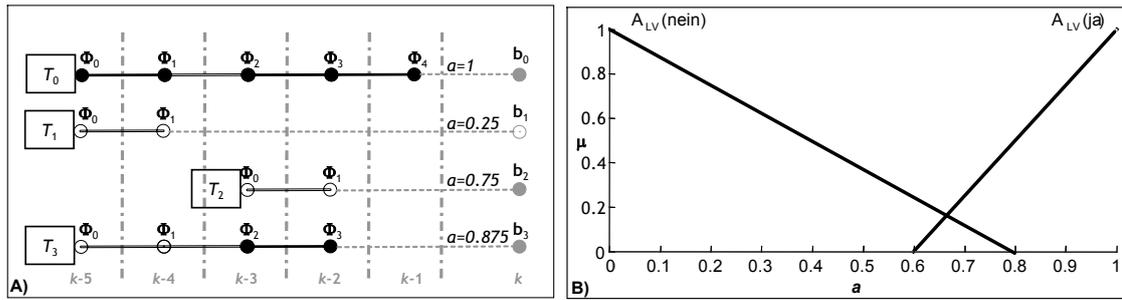


Abbildung 5-3-4: A) Beispiele für mögliche Aktivierungsgrade und B) Fuzzysset für den Aktivitätsgrad einer Trajektorie.

II) Ein weiterer Bestandteil der linguistischen Variable ist das *Fuzzybewegungsvektormatching*, das die Fuzzifizierung der m:n Zuordnung zwischen den Messungen und den Trajektorienknoten beschreibt. Dies erfolgt durch die fuzzifizierten optischen Flussdaten<sup>101</sup>, indem man die m:n Zuordnung betrachtet, die sich in eine m:1 und 1:n Zuordnung zerlegen lässt. Dies wird wie folgt kurz erläutert:

- Die Idee für die erste 1:n Zuordnungsmöglichkeit besteht in der Annahme, dass jeder Bewegungsvektor aufgrund eines definierten Gitters die Bewegung eines Flächenelements beschreibt. Um eine Verbindung zwischen einem Trajektorienknoten und n MB herstellen zu können, müssen sich daher die Flächenelemente des letzten Knotens der Trajektorie  $\Phi_{j,k-1}$  in dem MB wieder finden lassen. Nehmen die bewegten Flächenelemente des letzten Knotens  $\Phi_{j,k-1}$  der Trajektorie  $T_{j,k-1}$  einen großen Raum in der Fläche des MB ( $\mathbf{b}_{i,k}$ ) ein, so ist es sicher, dass das Objekt der Trajektorie  $T_{j,k-1}$  in dem MB ( $\mathbf{b}_{i,k}$ ) zu finden ist. Somit setzt sich das 1:n Zuordnungskriterium aus der Gesamtfläche  $A(\mathbf{b}_{i,k})$  des MBs und den definierten Flächenelementen  $d_{grid}$  der Trajektorie  $T_{j,k-1}$ , die den MB ( $\mathbf{b}_{i,k}$ ) getroffen haben, zusammen zu:

$$MT1(\Phi_{j,k-1}, \mathbf{b}_{i,k}) = d_{grid} \cdot \frac{\eta_T(\Phi_{j,k-1})}{A(\mathbf{b}_{i,k})} \quad (5-3-5a)$$

$\eta_T(\Phi_{j,k-1})$  ist die Anzahl der Bewegungsvektoren des Endknotens  $\Phi_j$  zur Zeit  $k-1$ .

- Das m:1 Zuordnungskriterium beschreibt die andere Richtung des Matchings. Es kann direkt aus der Anzahl der auf ein MB treffenden Bewegungsvektoren ermittelt werden. Das heißt bildlich gesehen, wenn sämtliche Bewegungsvektoren des Endknotens  $\Phi_{j,k-1}$  einer Trajektorie in einem MB ( $\mathbf{b}_{i,k}$ ) zusammentreffen, so stammt das Objekt des MBs ( $\mathbf{b}_{i,k}$ ) aus der Trajektorie  $T_{j,k-1}$ . Das zweite Zuordnungskriterium ist somit definiert durch:

$$MT2(\Phi_{j,k-1}, \mathbf{b}_{i,k}) = \frac{\eta_T(\mathbf{b}_{i,k})}{\eta_G(\Phi_{j,k-1})} \quad (5-3-5b)$$

Dabei ist  $\eta_T(\mathbf{b}_{i,k})$  die Anzahl der Bewegungsvektoren der Bewegungsmaske von  $\mathbf{b}_i$  zur Zeit  $k$  und  $\eta_G(\Phi_{j,k-1})$  stellt die Gesamtanzahl der Bewegungsvektoren des Endknotens der Trajektorie  $T_{j,k-1}$  dar. Die Fuzzifizierung der Zuordnungskriterien MT1 und MT2 wird über die Mengensysteme (Gl. 5-3-5c) und Abbildung 5-3-5 dargestellt. Hier

<sup>101</sup> Für die Berechnung der optischen Flussdaten wird der im Kapitel 2 vorgestellte pyramidale LK-Algorithmus mit der Adaption gemäß Kapitel 2 angewendet.

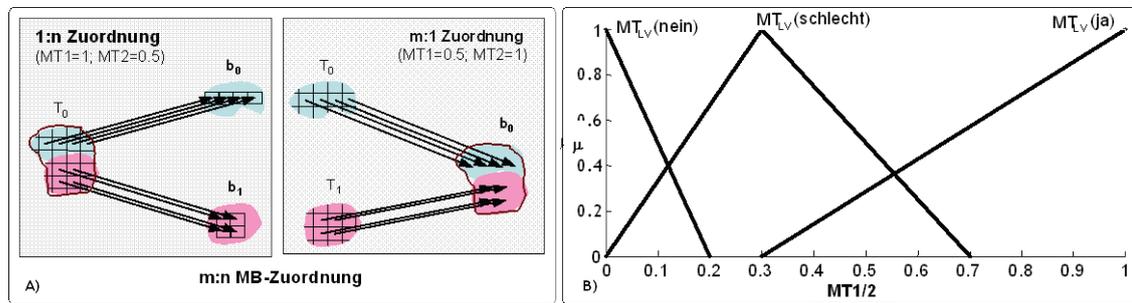
## 5. DATENFUSION AUF DER BASIS VON PARTIKELFILTERN

wurden die Abkürzungen ne für *nein* und sc für *schlecht* eingefügt.

$$MT1_{LV} = \left\{ \mu_R^{(3)}(MT1) = MT_{LV}(ja), \mu_R^{(4)}(MT1) = MT_{LV}(sc), \mu_R^{(5)}(MT1) = MT_{LV}(ne) \right\} \quad (5-3-5c)$$

$$MT2_{LV} = \left\{ \mu_R^{(3)}(MT2) = MT_{LV}(ja), \mu_R^{(4)}(MT2) = MT_{LV}(sc), \mu_R^{(5)}(MT2) = MT_{LV}(ne) \right\} \quad (5-3-5d)$$

Das Fuzzysset besteht in diesem Fall aus drei Fuzzymengen, die erhöht wurden, um die Sensitivität der Regelung zu verbessern. Eine erhöhte Sensitivität ist deshalb notwendig, weil durch das Auftreten von Störungen in der optischen Flussberechnung, bei der Regelung von kleinen Objekten, Probleme auftreten können. Denn je kleiner ein zu verfolgendes Objekt ist, desto schwerer ist die Trennung der Prädiktion der partiellen Objektflächen von den falsch prädiktierten Flächenelementen, die durch Ausreißer entstanden sind.



**Abbildung 5-3-5:** In A) ist das Prinzip der Bewertungskriterien MT1 und MT2 dargestellt, wobei die Summe der durch die Bewegungsvektoren verschobenen Flächenelemente der Trajektorie  $T_0$ , die im Folgebild den MB  $b_0$  getroffen haben, beinhaltet. Je größer die Restfläche ist, desto niedriger ist der Zugehörigkeitsgrad MT1, des MB  $b_0$  zu der Trajektorie  $T_0$ . Das gleiche gilt für den MB  $b_1$  und den daraus resultierenden Zugehörigkeitsgrad MT1. Es werden in diesem Beispiel der Trajektorie  $T_0$  beide MBs zugeordnet. Je mehr Vektorendpunkte, also innerhalb der Segmentierungsmaske des MB  $b_0$  liegen, desto größer ist der Zugehörigkeitsgrad MT2 für die Trajektorien  $T_0$  und MT2 für die Trajektorie  $T_1$  und dem MB. In B) ist das Fuzzysset für die Zuordnungskriterien dargestellt.

**III)** Das *Fuzzyfarbmatching* definiert die letzte linguistische Variable, das das Fuzzysset für den Vergleich der Objektfarbmerkmale mittels Ähnlichkeitsmaß definiert. Die Grundlage dafür ist die Interpretation der Farbhistogrammdata der erfassten Knoten und Motion-Blobs, als Vektoren. Ein heuristisches Ähnlichkeitsmaß, das hier verwendet wird, ist der *Histogrammschnitt*  $D_{HS}$ , indem die relativen Häufigkeiten als Flächen betrachtet werden. In dieser Betrachtungsweise wird die Ähnlichkeit über die gemeinsamen Flächen der Histogramme berechnet. Diese wird zwischen dem Histogramm  $h_b$  des MBs und dem Histogramm  $h_T$  des letzten Knotens einer Trajektorie durch die Gleichung (5-3-6a) mit  $h_T = \mathbf{h}(\Phi_{k-1})$ ,  $h_b = \mathbf{h}(b_{i,k})$  und  $U$  der Anzahl der Histogrammurnen berechnet.

$$D_{HS}(h_T, h_b) = 1 - \left( \sum_{u=0}^{U-1} \min(h_b(u), h_T(u)) \right) \cdot \left( \sum_{u=0}^{U-1} h_T(u) \right)^{-1} \quad (5-3-6a)$$

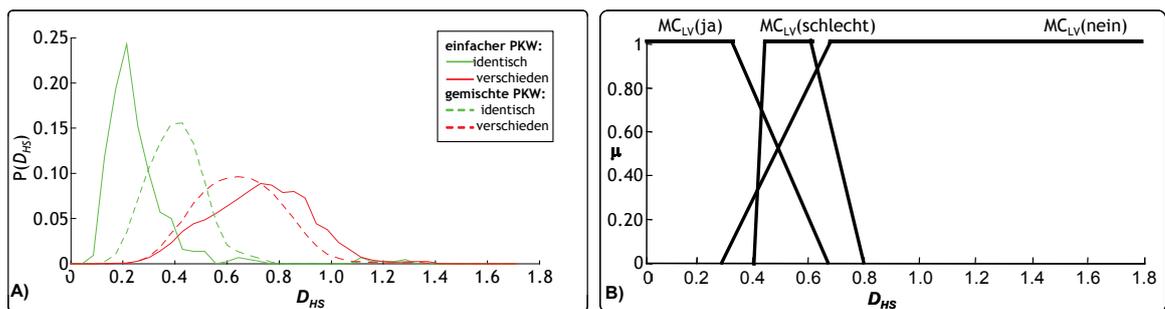
Stimmen die beiden Histogramme überein, dann liefert diese Gleichung ein Minimum. Das Fuzzysset für die Farbmerkmale besteht aus drei Fuzzymengen gemäß Gleichung (5-3-6b), die auf der Grundlage der Wahrscheinlichkeitsverteilungen der durchgeführten

## 5. DATENFUSION AUF DER BASIS VON PARTIKELFILTERN

Versuche<sup>102</sup> (s. Anhang 8.3) zur Bestimmung einer optimalen Distanzfunktion für Fahrzeuge festgelegt werden. Daraus hat sich ergeben, dass der Histogrammschnitt eine gute Separabilität der identischen und nicht identischen Auftrittswahrscheinlichkeiten der Fahrzeuge besitzt, im Besonderen bei Situationen an denen mehrere Fahrzeuge beteiligt sind<sup>103</sup>. Der dafür gewählte HS-Farbraum ist ein Unterraum des wahrnehmungsorientierten HSV-Raumes (*Hue-Saturation-Value*), wodurch der Histogrammvergleich unabhängig von Helligkeitsschwankungen wird und mit 64 Urnen wurde ein gutes Mittelmaß zwischen der Separabilität und der Rechengeschwindigkeit getroffen.

$$MC_{LV} = \{\mu_R^{(6)}(D_{HS}) = MC_{LV}(ja), \mu_R^{(7)}(D_{HS}) = MC_{LV}(schlecht), \mu_R^{(8)}(D_{HS}) = MC_{LV}(nein)\} \quad (5-3-6b)$$

In der Abbildung 5-3-6 ist das Fuzzyset für den Farbvergleich in Anlehnung an der Wahrscheinlichkeitsverteilung aus den durchgeführten Versuchen dargestellt. Dabei waren diejenigen Versuche von besonderem Interesse, in denen ein Objekt in einer Menge von Objekten/Fahrzeugen wieder erkannt wurde (gemischte PKW). Denn in dem Bewegungsmodell ist es notwendig bei komplexeren Szenarien ein Fahrzeug einem MB zuzuordnen, in dem sich mehrere PKW befinden können. Mehr Details sind aus den dargestellten Untersuchungen im Anhang 8.3 zu entnehmen.



**Abbildung 5-3-6:** Fuzzyset für den Farbvergleich in Anlehnung an die Wahrscheinlichkeitsverteilung aus Abbildung 8.3.1 (s. Abschnitt 8.3 im Anhang) .

### 5.3.2.2 Fuzzyregelung

Das Bewegungsmodell wird auf jedes der Partikel der Partikelmenge  $S_{k-1}$  angewandt. Es besteht im Einzelnen aus drei Mamdani-Assini-Fuzzyreglern den Connect-, Split,- und Merge-Operatoren, die miteinander verknüpft sind. Die Fuzzyübertragungsfunktion ist stückweise über die wissenbasierte linguistischen Variablen definiert, die aus einer endlichen Menge von Wenn-Dann-Regeln besteht. Für diese Fuzzyregler werden nun diese Regelbasen aufgestellt.

- **Connect-Operator:** Die Aufgabe des Connect-Operators ist es, die Verbindungen zwischen den Endknoten ( $\Phi_{j=0,\dots,N_T-1}$ ) der  $N_T$ -Trajektorien des Zustands  $z_{k-1}^n$  und den segmentierten MBs ( $\mathbf{b}_{i=0,\dots,m-1}$ ) der Messung ( $\mathbf{m}_k$ ) herzustellen. Die Eingangsdaten bilden die Aktivitätsgrade für alle Trajektorien  $a(T_j)$ , die aus den Bewegungsvektoren

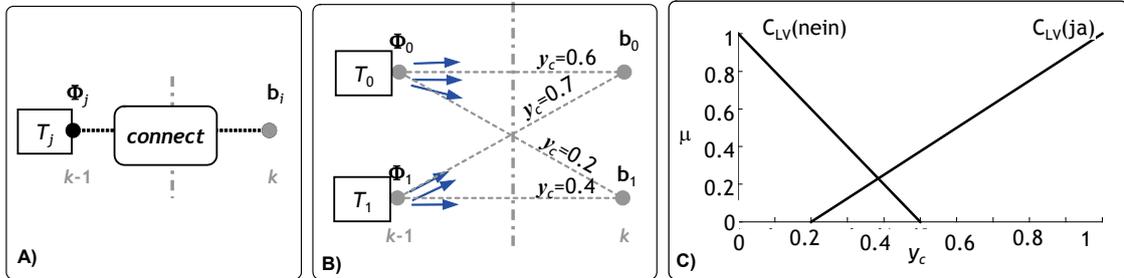
<sup>102</sup> Die Grundlage für das Finden der geeigneten Konfiguration für den Vergleich ist die durchgeführte Untersuchung mit einer Testdatenbank, die 43 verschiedene aus Verkehrsszenen per Hand markierte Fahrzeuge in jeweils fünf verschiedenen Positionen enthält

<sup>103</sup> Es wurden nicht nur Histogramme von einzelnen Fahrzeugen miteinander verglichen. Es wurde darauf Wert gelegt, dass ein Fahrzeug aus einem Histogramm, bestehend aus der Überlagerung des eigenen Histogramms und das eines zusätzlichen Fahrzeuges erkannt wird (mixed PKW).

## 5. DATENFUSION AUF DER BASIS VON PARTIKELFILTERN

definierten Bewertungskriterien  $MT1(\Phi_j, \mathbf{b}_i)$  und  $MT2(\Phi_j, \mathbf{b}_i)$  der Trajektorien und MBs, mit den zugehörigen Eingangs-Fuzzysets  $A_{LV}$ ,  $MT1_{LV}$  und  $MT2_{LV}$ . Der Regelausgang wird durch das Fuzzyset  $C_{LV}$  in Gleichung 5-3-7a und Abbildung 5-3-7 beschrieben.

$$C_{LV} = \{\mu_{R0}(y_c) = C_{LV}(ja), \mu_{R1}(y_c) = C_{LV}(ne)\} \quad (5-3-7a)$$



**Abbildung 5-3-7:** Funktionsweise des Connect-Operators. Über die Fuzzyregelung des Connect-Operators wird mit dem Connect-Ausgangsfuzzyset (B) der Zugehörigkeitsgrad für jede Verbindung einer Trajektorie mit einem MB ermittelt.

Die *Regelbasis* besteht aus folgenden Regeln:

$\forall \Phi_j \in Z_{k-1}^n, \forall \mathbf{b}_i \in \mathbf{m}_k, \forall T_j \in Z_{k-1}^n$  gilt:

- $R_1$ : wenn  $a(T_j) = A_{LV}(ja)$  und  $MT1(\Phi_j, \mathbf{b}_i) = MT_{LV}(ja)$  und  $MT2(\Phi_j, \mathbf{b}_i) = MT_{LV}(ja)$ ,  
dann gilt  $y_c(T_j, \mathbf{b}_i) = C_{LV}(ja)$
- $R_2$ : wenn  $a(T_j) = A_{LV}(ja)$  und  $MT1(\Phi_j, \mathbf{b}_i) = MT_{LV}(sc)$  und  $MT2(\Phi_j, \mathbf{b}_i) = MT_{LV}(ja)$ ,  
dann gilt  $y_c(T_j, \mathbf{b}_i) = C_{LV}(ja)$
- $R_3$ : wenn  $a(T_j) = A_{LV}(ja)$  und  $MT1(\Phi_j, \mathbf{b}_i) = MT_{LV}(ja)$  und  $MT2(\Phi_j, \mathbf{b}_i) = MT_{LV}(sc)$ ,  
dann gilt  $y_c(T_j, \mathbf{b}_i) = C_{LV}(ja)$
- $R_4$ : wenn  $a(T_j) = A_{LV}(ja)$  und  $MT1(\Phi_j, \mathbf{b}_i) = MT_{LV}(sc)$  und  $MT2(\Phi_j, \mathbf{b}_i) = MT_{LV}(ne)$ ,  
dann gilt  $y_c(T_j, \mathbf{b}_i) = C_{LV}(ne)$
- $R_5$ : wenn  $a(T_j) = A_{LV}(ja)$  und  $MT1(\Phi_j, \mathbf{b}_i) = MT_{LV}(ne)$  und  $MT2(\Phi_j, \mathbf{b}_i) = MT_{LV}(sc)$ ,  
dann gilt  $y_c(T_j, \mathbf{b}_i) = C_{LV}(ne)$
- $R_6$ : wenn  $a(T_j) = A_{LV}(ne)$  oder  $MT1(\Phi_j, \mathbf{b}_i) = MT_{LV}(ne)$  oder  $MT2(\Phi_j, \mathbf{b}_i) = MT_{LV}(ne)$ ,  
dann gilt  $y_c(T_j, \mathbf{b}_i) = C_{LV}(ne)$

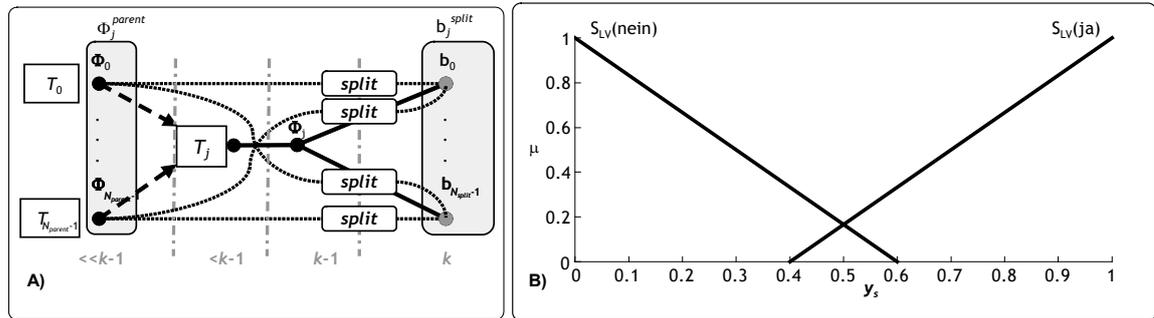
mit  $j=0, \dots, N_T-1$  und  $i=0, \dots, m-1$

Die Variable  $y_c(T_j, \mathbf{b}_i)$  ist ein Wert, der die Verbindungszugehörigkeit des MBs ( $\mathbf{b}_i$ ) zu der Trajektorie ( $T_j$ ) ausdrückt und wird durch die Defuzzifizierung der Ausgangsmengen  $C_{LV}$  gewonnen. Nachdem alle Connect-Operationen für ein Partikel ausgeführt worden sind, tritt der Split-Operator in Kraft, wenn eine Trajektorie mit mehreren MBs verbunden wird.

## 5. DATENFUSION AUF DER BASIS VON PARTIKELFILTERN

▪ **Split-Operator:** Die Aufgabe dieses Operators ist die Verbindung der Endknoten  $\Phi_j^{parent} = \{\Phi_{p=0, \dots, N_{parent}-1}\}^{104}$  der Elterntrajektorien<sup>105</sup>, die in einer Trajektorie  $T_j$  enthalten sind, mit den erfassten Motion-Blobs  $\mathbf{b}_j^{split} = \{\mathbf{b}_{s=0, \dots, N_{split}-1}\}$  herzustellen (s. Abb. 5-3-8A). Die Eingangsdaten bestehen aus den Bewegungsvektoren  $MT1(\Phi_j, \mathbf{b}_s)$  der Trajektorie  $T_j$  mit den MBs ( $\mathbf{b}_j^{split}$ ) und den Farbhistogrammdistanzen  $D_{HS}(\mathbf{h}(\Phi_p, \mathbf{b}_s))$  der Elterntrajektorien und der MBs, mit dem dazugehörigen Eingangs-Fuzzysets  $MT1_{LV}$  und  $MC_{LV}$ . Der Regelausgang wird durch das Fuzzyset  $S_{LV}$  beschrieben.

$$S_{LV} = \{\mu_{R2}(y_s) = S_{LV}(ja), \mu_{R3}(y_s) = S_{LV}(ne)\} \quad (5-3-7b)$$



**Abbildung 5-3-8:** A) Funktionsweise des Split-Operators. Wird einer Trajektorie mit mehreren MBs verbunden, so werden alle Eltern dieser Trajektorie ermittelt. Durch die Fuzzyregeln des Split-Operators und dem Split-Ausgangsfuzzyset in B) wird dann der Zugehörigkeitsgrad eines jeden verbundenen MB  $\mathbf{b}_s$  zu jeder Elterntrajektorie  $T_p$  ermittelt.

Die *Regelbasis* besteht aus folgenden Regeln:

$\forall \Phi_p (\Phi_p \in \Phi_j^{parent} \wedge \Phi_p \in Z_{k-1}^n), \forall \mathbf{b}_s (\mathbf{b}_s \in \mathbf{b}_j^{split} \wedge \mathbf{b}_s \in \mathbf{m}_k), \forall \Phi_j \in Z_{k-1}^n, \forall T_p \in Z_{k-1}^n$  gilt:

- $R_1$ : wenn  $MT1(\Phi_j, \mathbf{b}_s) = MT_{LV}(ja)$  und  $D_{HS}(\mathbf{h}(\Phi_p), \mathbf{h}(\mathbf{b}_s)) = MC_{LV}(ja)$ ,  
dann gilt  $y_s(T_p, \mathbf{b}_s) = S_{LV}(ja)$
- $R_2$ : wenn  $MT1(\Phi_j, \mathbf{b}_s) = MT_{LV}(ja)$  und  $D_{HS}(\mathbf{h}(\Phi_p), \mathbf{h}(\mathbf{b}_s)) = MC_{LV}(sc)$ ,  
dann gilt  $y_s(T_p, \mathbf{b}_s) = S_{LV}(ja)$
- $R_3$ : wenn  $MT1(\Phi_j, \mathbf{b}_s) = MT_{LV}(sc)$  und  $D_{HS}(\mathbf{h}(\Phi_p), \mathbf{h}(\mathbf{b}_s)) = MC_{LV}(ja)$ ,  
dann gilt  $y_s(T_p, \mathbf{b}_s) = S_{LV}(ja)$
- $R_4$ : wenn  $MT1(\Phi_j, \mathbf{b}_s) = MT_{LV}(sc)$  und  $D_{HS}(\mathbf{h}(\Phi_p), \mathbf{h}(\mathbf{b}_s)) = MC_{LV}(sc)$ ,  
dann gilt  $y_s(T_p, \mathbf{b}_s) = S_{LV}(ne)$
- $R_5$ : wenn  $MT1(\Phi_j, \mathbf{b}_s) = MT_{LV}(ne)$  oder  $D_{HS}(\mathbf{h}(\Phi_p), \mathbf{h}(\mathbf{b}_s)) = MC_{LV}(ne)$ ,  
dann gilt  $y_s(T_p, \mathbf{b}_s) = S_{LV}(ne)$
- mit  $s = 0, \dots, N_{split}-1, p = 0, \dots, N_{parent}-1$  und  $j = 0, \dots, N_T-1$

<sup>104</sup>  $N_{parent}$  Anzahl der Elterntrajektorien und  $N_{split}$  Anzahl der MBs

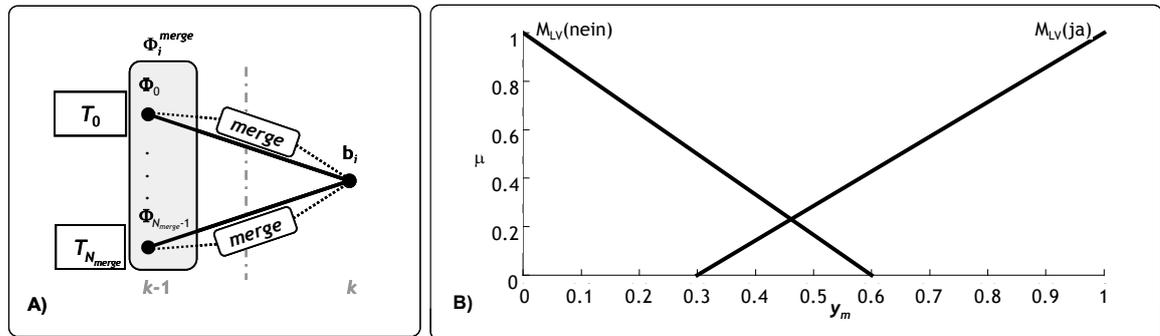
<sup>105</sup> Durch einen rekursiven Algorithmus wird so lange in Trajektorien nach Elterntrajektorien über die gespeicherten Elternindizes gesucht, bis es keine Elterntrajektorien gibt, somit ist gewährleistet, dass die Elterntrajektorie initial nur ein Objekt besitzt.

## 5. DATENFUSION AUF DER BASIS VON PARTIKELFILTERN

Die Variable  $y_s(T_p, \mathbf{b}_s)$  gibt also einen Wert an, der die Verbindungszugehörigkeit des MBs ( $\mathbf{b}_s$ ) der Menge ( $b_j^{split}$ ) zu den Objekten der Trajektorie  $T_j$  ausdrückt und durch die Defuzzifizierung der Ausgangsmengen  $S_{LV}$  gewonnen wird. Ob eine Elterntrajektorie  $T_p$  mit einem MB ( $\mathbf{b}_s$ ) verbunden wird, hängt also von zwei Faktoren ab. Zum einen vom Verbindungswert MT1 zwischen der Kindtrajektorie  $T_j$  und dem MB und zum anderen von der Farbdistanz  $D_{HS}$  zwischen der Elterntrajektorie  $T_p$  und dem MB. Besitzt die Trajektorie  $T_j$  keine Eltern, so wird eine Kopie der Trajektorien erzeugt und diese mit einem der MBs verbunden. Dieser Fall tritt dann auf, wenn eine Trajektorie mit einem MB der zwei Fahrzeuge enthält, initialisiert wurde. Da kein Vorwissen existiert, wird in diesem Fall jede aufgeteilte Trajektorie gleichberechtigt behandelt. Besitzt die Trajektorie jedoch Eltern, dann werden die Elterntrajektorien den MB ( $\mathbf{b}_s$ ) zugeordnet. Der Teil der Trajektorie der seit der Verschmelzung weiterentwickelt wurde, wird auf die Elterntrajektorie übertragen und der Aktivitätswert ( $a$ ) der Elterntrajektorie reinitialisiert, während der Aktivitätswert der leeren Trajektorie  $T_j$  auf Null gesetzt wird. Werden mehrere Trajektorien mit einem MB verbunden, dann tritt der Merge-Operator in Kraft.

- **Merge-Operator:** Die Aufgabe dieses Operators ist die Regelung der Verschmelzung von mehreren Trajektorien mit den Endknoten  $\Phi_j^{merge} = \{\Phi_{k=0, \dots, N_{merge}-1}\}^{106}$ , die mit einem MB verbunden wurden. Die Eingangsdaten bilden die Bewegungsvektoren  $MT1(\Phi_k, \mathbf{b}_i)$  der Knotentrajektorien, die mit einem MB ( $\mathbf{b}_i$ ) verbunden wurden sind, mit dem Eingangsfuzzysets  $MT1_{LV}$ . Der Regelausgang wird durch das Fuzzyset  $M_{LV}$  in Gleichung (5-3-7c) und Abbildung 5-3-9 beschrieben.

$$M_{LV} = \{\mu_{R4}(y_m) = M_{LV}(ja), \mu_{R5}(y_m) = M_{LV}(ne)\} \quad (5-3-7c)$$



**Abbildung 5-3-9:** Funktionsweise des Merge-Operators, der in Kraft tritt, wenn mehrere Trajektorien mit einem MB verbunden sind. Dann wird durch die Fuzzyregeln des Merge-Operators und des Merge-Ausgangsfuzzysets in B) erneut der Zugehörigkeitsgrad einer jeden Trajektorie zu dem MB berechnet.

Die *Regelbasis* besteht aus folgenden Regeln:

$\forall \Phi_k (\Phi_k \in \Phi_j^{merge} \wedge \Phi_k \in Z_{k-1}^n), \mathbf{b}_i \in \mathbf{m}_k$  gilt:

$R_1$ : **wenn**  $MT1(\Phi_k, \mathbf{b}_i) = MT_{LV}(ja)$ , **dann** gilt  $y_m(T_k, \mathbf{b}_i) = M_{LV}(ja)$

<sup>106</sup>  $N_{merge}$  Anzahl der mit dem MB  $\mathbf{b}_i$  verbundenen Trajektorien

## 5. DATENFUSION AUF DER BASIS VON PARTIKELFILTERN

- 
- $R_2$ : **wenn**  $MT1(\Phi_{0+s}, \mathbf{b}_i) = MT_{LV}(sc)$  **und** ... **und**  $MT1(\Phi_{N_{merge}-1-n}, \mathbf{b}_i) = MT_{LV}(sc)$   
**und**  $MT1(\Phi_k, \mathbf{b}_i) = MT_{LV}(sc)$ , **dann** gilt  $y_m(T_k, \mathbf{b}_i) = M_{LV}(ja)$
- $R_3$ : **wenn**  $MT1(\Phi_{0+s}, \mathbf{b}_i) = MT_{LV}(ne)$  **und** ... **und**  $MT1(\Phi_{N_{merge}-1-n}, \mathbf{b}_i) = MT_{LV}(ne)$   
**und**  $MT1(\Phi_k, \mathbf{b}_i) = MT_{LV}(sc)$ , **dann** gilt  $y_m(T_k, \mathbf{b}_i) = M_{LV}(ja)$
- $R_4$ : **wenn**  $MT1(\Phi_{0+s}, \mathbf{b}_i) = MT_{LV}(ja)$  **und** ... **und**  $MT1(\Phi_{N_{merge}-1-n}, \mathbf{b}_i) = MT_{LV}(ja)$   
**und**  $MT1(\Phi_k, \mathbf{b}_i) = MT_{LV}(sc)$ , **dann** gilt  $y_m(T_k, \mathbf{b}_i) = M_{LV}(ne)$
- $R_5$ : **wenn**  $MT1(\Phi_k, \mathbf{b}_i) = MT_{LV}(ne)$ , **dann** gilt  $y_m(T_k, \mathbf{b}_i) = M_{LV}(ne)$

$$\text{mit } n = \begin{cases} 1 & , \text{ wenn } k = N_{merge} - 1 \\ 0 & , \text{ sonst} \end{cases}, \quad s = \begin{cases} 1 & , \text{ wenn } k = 0 \\ 0 & , \text{ sonst} \end{cases}, \quad i=0 \dots M-1 \text{ und } k=0, \dots, N_{merge}-1$$


---

Die Regelung der Verschmelzung der Trajektorien in dem MB wird notwendig, da es möglich ist, dass unter den zu verschmelzenden Trajektorienkandidaten auch Ausreißer enthalten sein können. Ein Merkmal für die Erkennung dieser Ausreißer ist, dass sie gegenüber den normalen Kandidaten kleinere Flächensegmente besitzen. Somit können sie durch eine angepasste Regelung beseitigt werden. Die Variable  $y_m(T_k, \mathbf{b}_i)$  drückt also einen Wert aus, der die Aufrechterhaltung der Verbindung des MB ( $\mathbf{b}_i$ ) der Trajektorie  $T_k$  ausdrückt und durch die Defuzzifizierung der Ausgangsmengen  $M_{LV}$  gewonnen wird. Wenn genügend Trajektorien mit dem MB verbunden werden, wird die Verschmelzung der Trajektorien dadurch geregelt, dass der Aktivitätswert (a) jeder Trajektorie  $T_k$  auf Null gesetzt und eine neue Trajektorie  $T_{N_T}$  erzeugt wird. Die Menge der Elterntrajektorien wird dann in die Trajektorie  $T_{N_T}$  eingetragen.

### 5.3.2.3 Defuzzifizierung und stochastische Entscheidung

In der letzten Stufe des Fuzzy-Systems werden die unscharfen Fuzzyausgangsmengen  $C_{LV}$ ,  $M_{LV}$  und  $S_{LV}$  auf die scharfen Ausgangswerte  $y_c$ ,  $y_m$  und  $y_s$  abgebildet. Dafür existieren verschiedene Methoden wie z.B. die *Max-Methode*, die den Wert  $y$ , mit dem maximalen Zugehörigkeitsgrad liefert. Allerdings kann diese Methode zu einem sprunghaften Verhalten (z.B. bei lokalen Maxima und Plateaus) des Ausgangswertes führen. Alternative dazu ist die *Mean-of-Maximum-Methode*, die den Mittelwert aller Maxima bildet. Allerdings bereitet auch diese Methode Schwierigkeiten, wenn die Zugehörigkeitsfunktion der Ausgangsmenge ein Plateau besitzt, auf dem überall das Maximum angenommen wird.

Auf Grund der erwähnten Nachteile wurde hier die *Center-of-Gravity-Methode* (CoG) angewendet, die den Ausgangswert  $y$  aus dem Schwerpunkt der Fläche der Fuzzy-Ausgangsmenge berechnet. Sie setzt für die Definition der LV stetig charakteristische Funktionen voraus, die jedoch aufgrund der Übersichtlichkeit der Fuzzy-Regelung meist gegeben sind. Da die numerische Integration für die Berechnung des Schwerpunktes sehr aufwendig ist, kann sie, wie in [Bothe 93], durch eine Gewichtung der Einzelschwerpunkte  $y_i$  der Fuzzy-Mengen der Ausgangs-Fuzzy-Sets angenähert werden. Dabei erfolgt die Gewichtung über die Aktivierungsgrade, der als Maximum über die Einzelaktivierungsgrade  $\alpha$  der Regel  $R$ , die Vereinigung der resultierenden Ausgangsmengen approximiert. Mit der Approximation der CoG-Methode [Bothe 93] über die Einzelaktivierungsgrade  $\alpha_l$  mit  $l=0, \dots, 5$  für die Ausgangsmengen  $\mu_{RI}$  ergeben sich die scharfen Ausgangswerte zu:

## 5. DATENFUSION AUF DER BASIS VON PARTIKELFILTERN

$$y_c(T_j, \mathbf{b}_i) = \frac{\sum_{l=0}^1 \alpha_l \cdot \mu_{Rl}}{\sum_{l=0}^1 \alpha_l}, \quad y_s(T_p, \mathbf{b}_s) = \frac{\sum_{l=2}^3 \alpha_l \cdot \mu_{Rl}}{\sum_{l=2}^3 \alpha_l} \quad \text{und} \quad y_m(T_k, \mathbf{b}_i) = \frac{\sum_{l=4}^5 \alpha_l \cdot \mu_{Rl}}{\sum_{l=4}^5 \alpha_l} \quad (5-3-8a)$$

Die defuzzifizierten Werte geben nun einen Zugehörigkeitsgrad zur Schwellenwertentscheidung an, um die Ausführung eines Operators zu steuern.

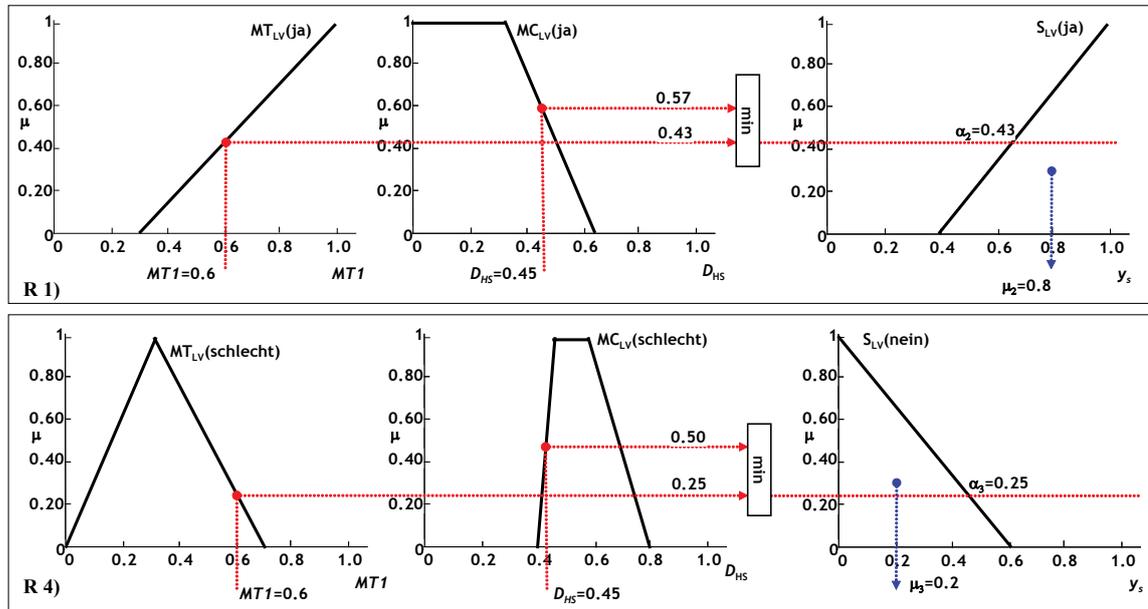


Abbildung 5-3-10a: Fuzzyregelung am Beispiel des Split-Operator für Regel 1 und 4 ( $y_s=0.58$ ).

Die Entscheidung, ob ein Operator ausgeführt<sup>107</sup> wird, tritt dann ein, wenn der Ausgangswert  $y$  eine Schwelle  $H$  überschreitet. Um die Variation des Bewegungsmodells zu berücksichtigen, werden die Schwellwerte  $H$  als stochastische Variablen modelliert, deren Verteilungsdichtefunktionen in Form der Normalverteilung  $N(\mu, \sigma^2)$  gegeben ist, so dass die Aktivierung der jeweiligen Operatoren eintritt, wenn gilt:  $y_c(T_j, \mathbf{b}_i) > H_c$ ,  $y_m(T_k, \mathbf{b}_i) > H_m$  und  $y_s(T_p, \mathbf{b}_s) > H_s$  mit:

$$f_{H_c}(y) = \frac{1}{\sqrt{2\pi\sigma_c}} e^{-\frac{(y-\mu_c)^2}{2\cdot\sigma_c^2}}, \quad (5-3-9a)$$

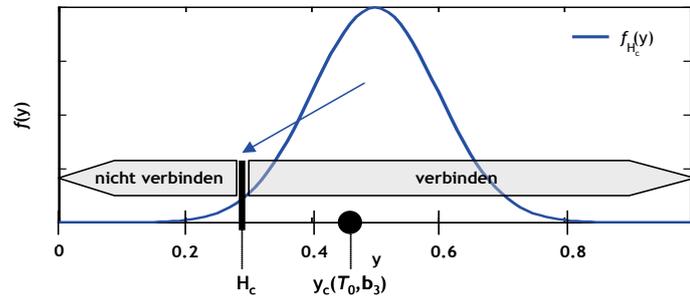
Die Schwellwerte  $H_c$ ,  $H_m$  und  $H_s$  werden demnach als Zufallswerte nach der Monte-Carlo-Methode<sup>108</sup> implementiert. Durch die Variation des Bewegungsmodells mit einem normalverteilten Rauschen können auch Trajektorienkonstellationen, die mit dem aus dem Bewegungsmodell eingebrachten Wissen unwahrscheinlich sind, erzeugt werden. Die Variation gibt dem Verfahren eine Robustheit gegenüber neuen und damit nicht in Wissensbasis (Fuzzyregelung) vertretenen Situationen.

In Abbildung 5-2-10b ist die Anwendung dieses Verfahrens beispielhaft für den Connect-Operator dargestellt. Das prinzipielle Diagramm zeigt die Normalverteilung

<sup>107</sup> Ausführen eines Operators bedeutet das Verbinden bzw. Verschmelzen oder Trennen von Knoten.

<sup>108</sup> Dazu wird ein Samplewert durch einen normalverteilten Zufallsprozess, der den Verteilungsdichtefunktionen der jeweiligen Schwellwerte entspricht, berechnet und als Schwellwert verwendet.

des Connect-Schwellwertes aus der ein Sample  $H_c$  berechnet wird. Die Fuzzyregelung hat für das Trajektorien-MB-Paar den Ausgangswert  $y_c(T_0, \mathbf{b}_3)$  geliefert. Da der Ausgangswert größer als das Sample des Schwellwertes ist, wird die Trajektorie  $T_0$  um die Messung  $\mathbf{b}_3$  ergänzt.



**Abbildung 5-3-10b:** Beispiel für die Anwendung des stochastischen Schwellwertes für den Connect-Operator. Aus der Verteilungsfunktion wird ein Schwellwertsample  $H_c$  erzeugt. Für den Ausgangswert  $y_c(T_0, \mathbf{b}_3)$  der Fuzzyregelung wird dann entschieden, ob der MB  $\mathbf{b}_3$  zu der Trajektorie  $T_0$  zugeordnet wird.

### 5.3.3 Das Beobachtungsmodell

Das Beobachtungsmodell agiert nachdem das Fuzzy-Bewegungsmodell die Partikelmenge  $S$  für den aktuellen Zustand prädiiziert hat. In Kombination mit dem Sampling-Schritt resultiert die Aktualisierung auf der Grundlage der in dem Beobachtungsmodell berechneten Likelihood-Funktion  $p(\mathbf{m}_k | \mathbf{z}_k)$  gemäß Gleichung (5-3-10a)<sup>109</sup> für jedes Partikel  $s^n$  der Partikelmenge  $S_k$ . Die Likelihood-Funktion setzt sich aus dem Glattheitskriterium  $F_{Glattheit}$  und dem Konsistenzkriterium  $F_{Konsistenz}$  zusammen und wird für jedes Partikel  $s^n$  durch die Partikelgewichte  $\pi_k^n$  angegeben. Sie stellen einfach die jeweiligen Auftrittswahrscheinlichkeiten des entsprechenden Partikelzustandes dar.

$$p(\mathbf{m}_k | \mathbf{z}_k) \propto -\left(F_{Glattheit}(\mathbf{z}_k) + F_{Konsistenz}(\mathbf{z}_k, \mathbf{m}_k)\right) \quad (5-3-10a)$$

**Glattheitskriterium:** Der erste Term ( $F_{Glattheit}$ ) in Gleichung (5-3-10a) gibt die Glattheit einer Trajektorie an, der die zeitliche Änderung der Knotenmerkmale bestimmt und sich aus dem Farbfehler ( $\hat{D}_{\chi^2}$ ) und dem Positionsfehler ( $\Delta_P$ ) zusammensetzt. Angestrebt für eine optimale Trajektoriebeschreibung, d.h. eine Trajektorie, die das Verhalten eines realen Objektes wiedergibt, sind eine minimale Änderung der Knotenfarben und eine genaue Prädiktion der Knotenbewegung durch den Positionsfehler. Diese Tatsache lässt sich durch den Gesamtglattheitsfehler beschreiben, der sich aus der Summe über alle  $N_k$ -Knoten<sup>110</sup> für  $N_T$ -Trajektorien zusammensetzt.

$$F_{Glattheit} = \sum_{j=0}^{N_T-1} \sum_{k=1}^{N_k-1} \left[ \beta \cdot \hat{D}_{\chi^2}(\mathbf{h}(\Phi)_{j,k-1}, \mathbf{h}(\Phi)_{j,k}) + \Delta_P(f'(\Phi)_{j,k-1}, f(\Phi)_{j,k}) \right] \quad (5-3-10b)$$

<sup>109</sup> Diese Gleichung drückt die indirekte lineare Proportionalität zwischen der Likelihood-Funktion und den Fehlerkriterien aus.

<sup>110</sup> Da die Trajektorien unterschiedliche Länge haben, bezieht sich der Index  $k$  auf Anfang ( $k=i$ ) und Ende der Trajektorie ( $k=N_k-1$ ) und nicht auf der Zeitlänge der Sequenz in dem Beobachtungsmodell.

Der erste Term  $\hat{D}_{\chi^2}$  in der Gleichung (5-3-10b) ist ein Maß für die Farbänderung zwischen zwei aufeinanderfolgenden Knoten  $\Phi_{j,k-1}$  und  $\Phi_{j,k}$  einer Trajektorie  $T_{j,k}$  und wird als Farbfehler bezeichnet. Der zweite Term  $\Delta_p$  gibt die Genauigkeit der Prädiktion der Knoten  $\Phi_{j,k-1}$  für den Knoten  $\Phi_{j,k}$  an. Für die Gewichtung der beiden Fehler sorgt die Konstante  $\beta$ , die durch empirische Versuche mit zwei aufgrund der Bedeutung der Farbinformation angegeben wird. Die beiden Terme werden wie folgt erklärt:

I) Um die Farbänderung beschreiben zu können, gibt es zahlreiche Farbdistanzfunktionen (s. Abschnitt 5.3.2). Im Unterschied zu dem Bewegungsmodell wird hier eine andere Farbdistanzfunktion angewendet. Denn bei dem Glattheitskriterium wird davon ausgegangen, dass die Anzahl der Objekte in aufeinander folgenden Trajektorieknoten gleich sein sollen. Die verwendete Farbdistanzfunktion muss daher sensitiv gegenüber Änderungen der Anzahl der Objekte sein. Das heißt, wenn das Objekt aus dem Knoten  $\Phi_{j,k-1}$  im darauf folgenden Knoten  $\Phi_{j,k}$  enthalten ist, der Knoten  $\Phi_{j,k}$  aber noch ein zusätzliches Objekt enthält, so soll die Farbdistanzfunktion dies durch eine *Zunahme* anzeigen. Ausgehend davon muss nun eine sensitivere Distanzfunktion gewählt werden, da der in dem Bewegungsmodell angewendete Histogrammschnitt  $D_{HS}$  diesen Situationen gegenüber tolerant ist. Auf der Grundlage der durchgeführten Untersuchungen (s. Anhang 8.3) in Verbindung mit einer geringeren False-Acceptance-Rate<sup>111</sup> (FAR) und False-Rejection-Rate (FRR), erfüllt der  $\chi^2$ -Abstand<sup>112</sup> aus Gleichung 5-3-10c diese gestellte Forderung und wird deshalb für die Farbfehlerbeschreibung eingesetzt.

$$D_{\chi^2}(\mathbf{h}(\Phi)_{j,k-1}, \mathbf{h}(\Phi)_{j,k}) = \sum_{u=0}^{U-1} \left( \frac{(\mathbf{h}(\Phi)_{j,k-1,u} - \mathbf{h}(\Phi)_{j,k,u})^2}{2 \cdot (\mathbf{h}(\Phi)_{j,k-1,u} + \mathbf{h}(\Phi)_{j,k,u})} \right) \quad (5-3-10c)$$

Für eine einfache Anwendung wäre diese Distanz zur Beschreibung der Farbänderung ausreichend. Allerdings ist es möglich (in Anlehnung an Abschnitt 5.3.2) die statistische Eigenschaft des Farbfehlers auf das Beobachtungsmodell zu übertragen. Somit kann der Zusammenhang zwischen dem Farbfehler und dem  $\chi^2$ -Abstand auf der Grundlage der Häufigkeitsverteilung der Histogrammdistanzen für das HS-64-Urnen-Histogramm genauer beschrieben werden (s. Anhang 8.3 für mehr Details). Von Interesse sind die  $\chi^2$ -Distanzen für die identischen Fahrzeuge  $P^I(\mathbf{D}_{\chi^2})$  (grün) und für die verschiedenen Fahrzeuge  $P^J(\mathbf{D}_{\chi^2})$  (rot) für einfache Stichproben (durchgezogene Linie). Daraus lässt sich die Wahrscheinlichkeit  $p(\mathbf{D}_{\chi^2})$  errechnen, die zwei identische Objekte mit einem  $\chi^2$ -Abstand ausdrückt.

<sup>111</sup> Anhand dieser Farbmerkmale soll die Zuordnung in eine der beiden Klassen erfolgen. Dabei werden für die beiden zu trennenden Klassen zwei Häufigkeitsverteilungen bestimmt. Sind diese beiden Klassen durch das Merkmal voneinander zu trennen, so werden sich zwei deutlich unterscheidbaren Kurven ergeben. Über die Größe des Grenzwertes, der das Grenzelement des Klassifikators darstellt, kann die Größe der FRR- oder der FAR-Fläche angepasst werden. Wird der Grenzwert liberaler eingestellt, steigt die Anzahl der falsch positiv erkannten Signale und die Anzahl der falsch negativ erkannten Signale sinkt.

<sup>112</sup> Die getroffene Auswahl orientiert sich an den Arbeiten von [Paulus 01, Knauer 05, Senst 2008].

$$p(\mathbf{D}_{\chi^2}) = \frac{P^r(\mathbf{D}_{\chi^2})}{P^r(\mathbf{D}_{\chi^2}) + P^f(\mathbf{D}_{\chi^2})} \quad (5-3-10d)$$

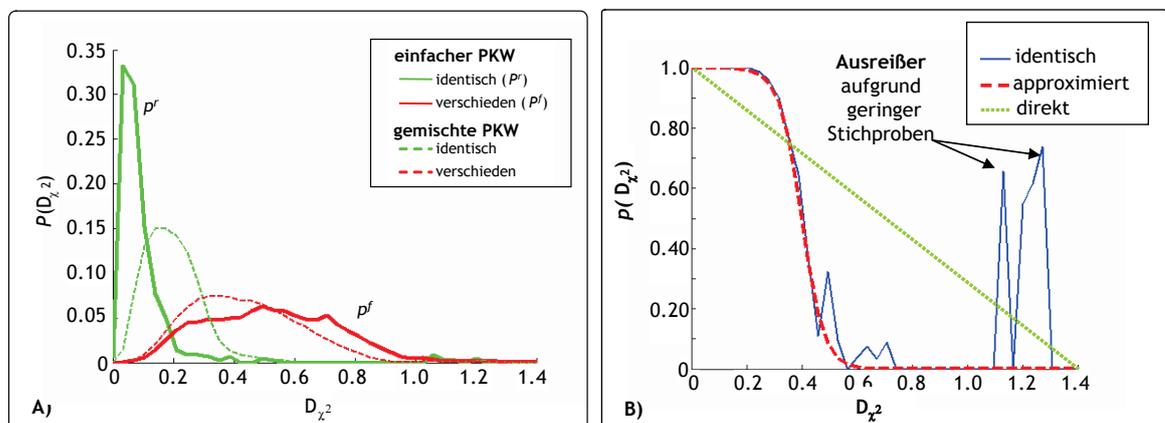
Abbildung 5-2-12B (blau) veranschaulicht die Wahrscheinlichkeit  $p(\mathbf{D}_{\chi^2})$ . Um den Wert  $p(\mathbf{D}_{\chi^2}) = 1,2$  entstehen allerdings Ausreißer, die dadurch bedingt sind, dass die Stichprobenanzahl für diese Distanzwerte so gering ist, dass sie nicht aussagekräftig sind. Aus diesem Grund können sie außer Acht gelassen werden. Es zeigt, dass sich die Wahrscheinlichkeitsfunktion  $p(\mathbf{D}_{\chi^2})$  durch die Exponentialfunktion  $\tilde{p}(\mathbf{D}_{\chi^2})$  approximiert werden kann (rot-gestrichelt):

$$\tilde{p}(\mathbf{D}_{\chi^2}) = \frac{1}{1 + e^{-\frac{(\mathbf{D}_{\chi^2} - 0.4)}{0.04}}} \approx p(\mathbf{D}_{\chi^2}) \quad (5-3-10e)$$

Daraus lässt sich der Farbfehler mit diesem Vorwissen über die Exponentialfunktion von der Farbdistanz zwischen den Histogrammen zweier Knotenpunkte  $\Phi_{j,k-1}$  und  $\Phi_{j,k}$  bestimmen, mit dem die totale Übereinstimmung der beiden Histogramme durch eine Null angegeben wird.

$$\hat{D}_{\chi^2}(\mathbf{h}(\Phi)_{j,k-1}, \mathbf{h}(\Phi)_{j,k}) = 1 - \tilde{p}\left(\mathbf{D}_{\chi^2}(\mathbf{h}(\Phi)_{j,k-1}, \mathbf{h}(\Phi)_{j,k})\right) \quad \text{mit } \hat{D}_{\chi^2} \in [0,1] \quad (5-3-10f)$$

Die Farbdistanz  $\hat{D}_{\chi^2}$  in Gleichung (5-3-10f) stellt somit eine an die Versuchsdatenbasis angepasste Funktion für die Bewertung der Histogrammdistanzen dar.



**Abbildung 5-2-12:** Darstellung der Häufigkeitsverteilung für das  $\chi$ -Distanzmaß aus [Senst 08] (s. Anhang 8.3) (A). B) Approximation der Wahrscheinlichkeit  $p$ , dass zwei Fahrzeuge (einfache PKW) identisch sind, durch die Gleichung 5.3.10e (B).

II) Der formbasierte Fehlerterm  $\Delta_p$  in der Gleichung (5-3-10b) drückt die Genauigkeit der Prädiktion der Knoten  $\Phi_{j,k-1}$  für den Knoten  $\Phi_{j,k}$  durch den Vergleich der prädiktieren Ellipse  $f'(\Phi)_{j,k-1}$  mit der darauf folgenden Knoten-Ellipse  $f(\Phi)_{j,k}$  aus. Dieser ist definiert durch den normierten euklidischen Abstand der Mittelpunkte der prädiktieren Ellipse des Vorgängerknoten  $\Phi_{j,k-1}$  und der Ellipse des jetzigen Knotens  $\Phi_{j,k}$ . Im

## 5. DATENFUSION AUF DER BASIS VON PARTIKELFILTERN

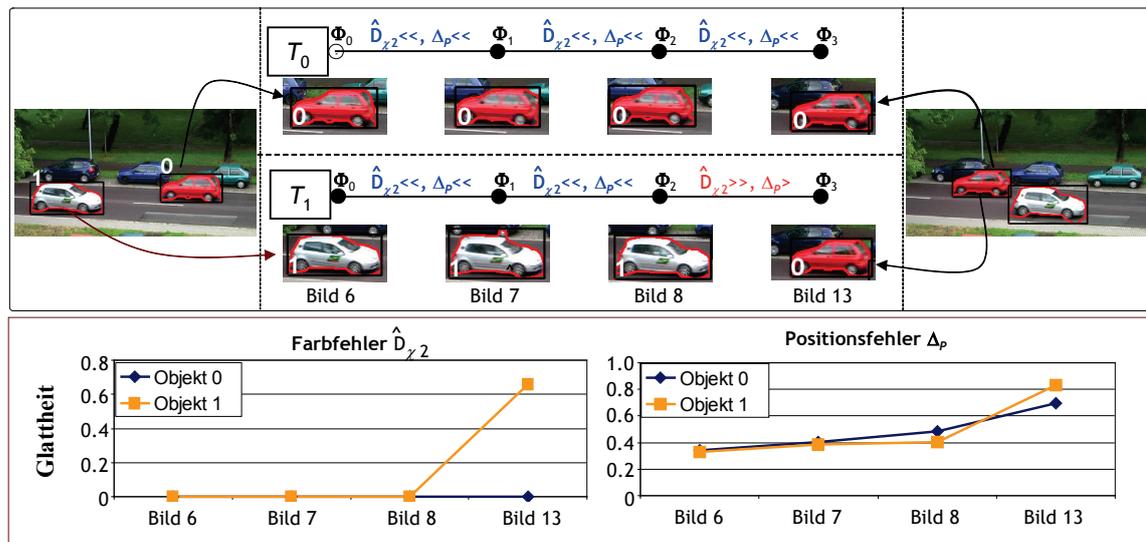
optimalen Fall stimmt die Position der prädiktierten Ellipse des Knotens  $\Phi_{j,k-1}$  mit der Position der Ellipse des darauf folgenden Knotens überein.

$$\Delta_P(f(\Phi)_{j,k}, f'(\Phi)_{j,k-1}) = \frac{1}{\sqrt{a' \cdot b'}} \cdot \sqrt{(x_m - x'_m)^2 + (y_m - y'_m)^2} \quad (5-3-10g)$$

mit  $x_m, y_m \in f(\Phi)_{j,k}$  und  $x'_m, y'_m, a', b' \in f'(\Phi)_{j,k-1}$ .

Als Normierungsfaktor wurden die Halbachsen der prädiktierten Ellipse genutzt, so dass der Prädiktionsfehler in Relation zu der Ellipsen-Größe steht. Ohne diesen Normierungsfaktor würden große Ellipsen/MB das Gütekriterium stärker beeinflussen und das Beobachtungsmodell würde sich an der Genauigkeit der Trajektorienschätzung großer Objekte ausrichten.

Das Zusammenwirken der beiden Kriterien in Gleichung (5-3-10b) wird in Abbildung 5-3-12 dargestellt, in der zwei Trajektorien ( $T_0$  und  $T_1$ ) für den Verlauf einer Konfliktsituation betrachtet werden, in den zwei Objekte verschmelzen und zu trennen sind. Dabei zeigt sich im Knoten  $\Phi_3$  die Auswirkung der falschen Zuweisung des Objektes (0) zu der Trajektorie ( $T_1$ ). Der Grund dafür ist auf die Variabilität des Bewegungsmodells zurückzuführen. Die daraus resultierende Zuordnung lässt sich durch eine signifikante Änderung des Glattheitskriteriums identifizieren (Abb. 5-3-12). In Folge dessen sinkt die Likelihood-Funktion dieses Zustandes und führt zu einer geringeren A-posteriori-Wahrscheinlichkeit.



**Abbildung 5-3-12:** Prinzip des Glattheitskriteriums. In der Trajektorie  $T_0$  wird im Zeitschritt 2 dem roten Auto ein falscher Nachfolger zugeordnet. Die Folge daraus ist, dass der Glattheitsfehler der Trajektorie  $T_1$  stärker zunimmt als der Glattheitsfehler der störungsfreien Trajektorie  $T_0$ .

**Konsistenzkriterium:** Das Konsistenzkriterium ( $F_{Konsistenz}$ ) soll die Länge der Trajektorie bewerten, indem die Anfangs- und Endknoten  $\Phi_{j,k}^{An} \in \Phi_{j,k}$ ,  $\Phi_{j,k}^{En} \in \Phi_{j,k}$  durch die weiteren Fehlermaße  $\Delta f_{ANF}$  und  $\Delta f_{END}$  untersucht werden. Dadurch wird bewertet, ob das Erzeugen bzw. Beenden einer Trajektorie gerechtfertigt ist. Ohne das Konsistenzkriterium würden kurze Trajektorien bevorzugt, da kurze Trajektorien weniger zu vergleichende Knotenpunkte besitzen. Dies entspricht aber nicht den in den beobachteten Szenen vorhandenen Sachverhalten.

## 5. DATENFUSION AUF DER BASIS VON PARTIKELFILTERN

$$F_{\text{Konsistenz}} = \sum_{j=0}^{N_T-1} \left[ \Delta f_{\text{ANF}}(\Phi_{j,k}^{\text{An}}) + \Delta f_{\text{END}}(\Phi_{j,k}^{\text{En}}) \right] \quad (5-3-11a)$$

Das Fehlermaß  $\Delta f_{\text{ANF}}$  für den Anfangsknoten und den Endknoten  $\Delta f_{\text{END}}$  lässt sich durch folgende Gleichungen definieren, in denen Farbfehler ( $\hat{D}_{\chi^2}$ ) und den Ellipsenabstandsfehler ( $\Delta_r$ ) berücksichtigt werden. Der Farbfehler wird aus dem Histogramm des Knotens  $\mathbf{h}(\Phi_{j,k}^{\text{An}})$  bzw.  $\mathbf{h}(\Phi_{j,k}^{\text{En}})$  und des Hintergrundes  $\mathbf{h}(\text{HG})$  gerechnet. Dieses enthält die Farbdaten der zu untersuchenden Bewegungsmaske, die im Hintergrundbild gespeichert sind. Die Idee, die zum Hintergrundhistogrammvergleich führt, ist folgende:

- Wenn eine Trajektorie erzeugt wird, so muss der Inhalt des ersten Knotens noch im Hintergrundbild nachzuweisen sein, da dieser vor einem Zeitschritt noch zum Hintergrund gehörte.
- Je geringer also der Farbunterschied des Knotens mit dem Hintergrund ist, desto größer ist die Wahrscheinlichkeit, dass die Erzeugung einer neuen Trajektorie für den MB gerechtfertigt ist und der MB keine Fortsetzung einer anderen Trajektorie darstellt.

$$\Delta f_{\text{ANF}}(\Phi_{j,k}^{\text{An}}) = \beta \cdot \hat{D}_{\chi^2}(\mathbf{h}(\Phi_{j,k}^{\text{An}}), \mathbf{h}(\text{HG})) + \Delta_r(\mathbf{f}(\Phi_{j,k}^{\text{An}}), \mathbf{f}(\Phi_{i,k})). \quad (5-3-11b)$$

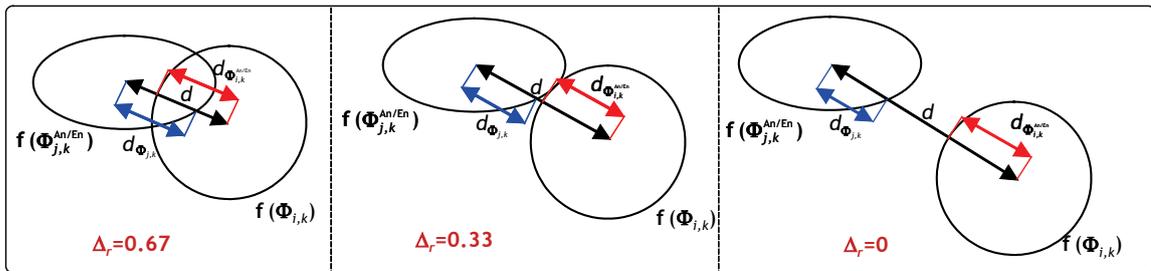
$$\Delta f_{\text{END}}(\Phi_{j,k}^{\text{En}}) = \beta \cdot \hat{D}_{\chi^2}(\mathbf{h}(\Phi_{j,k}^{\text{En}}), \mathbf{h}(\text{HG})) + \Delta_r(\mathbf{f}(\Phi_{j,k}^{\text{En}}), \mathbf{f}(\Phi_{i,k})). \quad (5-3-11c)$$

Der Ellipsenabstandsfehler ( $\Delta_r$ ) gibt ein Maß für den Abstand zwischen der Ellipse eines anderen Trajektorieknotens  $\Phi_{i,k}$ , die dem Anfangsknoten  $\Phi_{j,k}^{\text{An}}$  bzw. Endknoten  $\Phi_{j,k}^{\text{En}}$  der Trajektorie  $T_{j,k}$  am nächsten ist. Je größer dieser Abstand ist, desto geringer ist die Wahrscheinlichkeit, dass die neu erzeugte bzw. nicht weiter entwickelte Trajektorie in Wirklichkeit die Weiterführung einer anderen Trajektorie darstellt. Der Ellipsenabstandsfehler zwischen zwei Ellipsen ist gegeben durch die Gleichung (5-3-11d), wobei  $c_r=1,5$  ist.

$$\Delta_r(\Phi_{j,k}^{\text{An/En}}, \Phi_{i,k}) = \begin{cases} 0 & , \text{wenn } c_r \cdot (d_{\Phi_{j,k}^{\text{An/En}}} + d_{\Phi_{i,k}}) - d < 0 \\ 1 - \frac{d}{c_r \cdot (d_{\Phi_{j,k}^{\text{An/En}}} + d_{\Phi_{i,k}})} & , \text{sonst} \end{cases} \quad (5-3-11d)$$

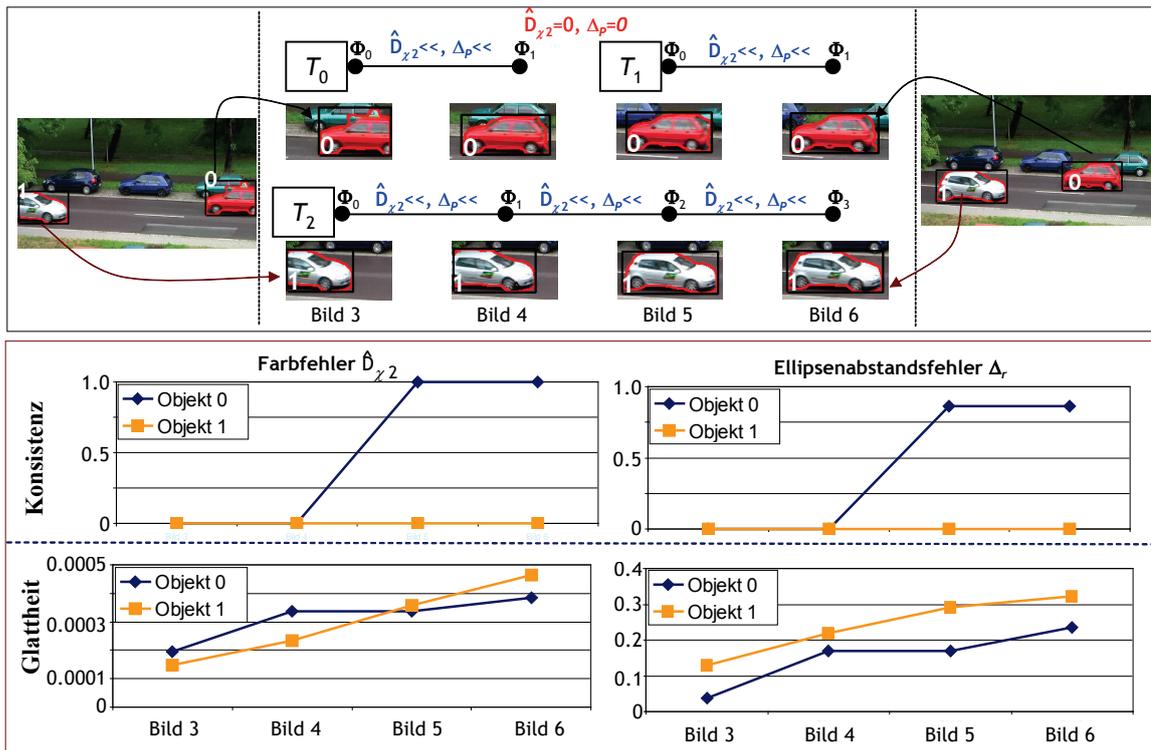
Die Koeffizienten  $d_{\Phi}$  dienen der Normierung des Ellipsenabstands  $d$ . Sie sind in der Abbildung 5-2-14 dargestellt. Diese Koeffizienten entsprechen den Strecken zwischen den Ellipsenmittelpunkten und dem Schnittpunkt der Verbindungsgeraden  $d$  mit den Ellipsenrändern, also der Breite der Ellipsen in Richtung der Verbindungsgeraden. Ein Abstandsfehler tritt dann auf, wenn der Abstand  $d$  zweier Ellipsen geringer als 1,5-mal der Summe ihrer gerichteten Breiten  $d_{\Phi}$  ist. Die Berechnung dieser Koeffizienten ist im Anhang 8.2 beschrieben.

## 5. DATENFUSION AUF DER BASIS VON PARTIKELFILTERN



**Abbildung 5-3-13:** Schematische Darstellung der Berechnung der Normierungskoeffizienten  $d_a$  und  $d_b$  für den Ellipsenabstandsfehler

Das Zusammenwirken zwischen dem Konsistenz- und dem Glattheitskriterium wird in Abbildung 5-3-14 gezeigt, in der ein konfliktloser Teilabschnitt einer Bildsequenz dargestellt wird. Durch das Bewegungsmodell ist es jedoch möglich, dass die Trajektorie ( $T_0$ ) im Knoten  $\Phi_1$  nicht weiter geführt wird und eine neue Trajektorie ( $T_1$ ) erzeugt wird. Bei der Betrachtung des Glattheitsfehlers wird deutlich, dass diese Situation bevorzugt wird, so dass zwischen dem Endknoten der Trajektorie  $T_0$  und dem Anfangsknoten der Trajektorie  $T_1$  kein Glattheitsfehler berechnet werden muss. Aus diesem Grund wurde das Konsistenzkriterium eingeführt, dass bei einer unberechtigten Unterbrechung der Trajektorie einen Fehler erzeugt.



**Abbildung 5-3-14:** Prinzip des Konsistenzkriteriums. Beide Trajektorien sind nach Beginn des ersten Bildes entstanden und enden vor dem Ende der Szene. Für die Trajektorie  $T_1$  existiert das Fahrzeug im ersten und letzten Bild, was sich in dem Fehlerwert widerspiegelt. Für die Trajektorie  $T_0$  existiert das Fahrzeug nicht in Bild 0 und 1. Der Konsistenzfehler ist geringer als bei  $T_1$ . Die Knoten  $T_{n0}$ ,  $T_{n1}$ ,  $T_{m0}$  und  $T_{m1}$  sind die nächstgelegenen Messungen zu den Anfangs- bzw. Endknoten der Trajektorien.

### 5.3.4 Sampling

Nach dem Aktualisierungskriterium muss nun die Partikelmenge so modifiziert werden, dass sie dem Satz von Bayes entspricht. Die A-priori-Wahrscheinlichkeit  $p(\mathbf{z}_k | \mathbf{m}_{k-1})$  ist durch das Bewegungsmodell bestimmt worden. Dadurch besitzt der Zustand  $\mathbf{z}_k$  (in Form der Partikelmenge  $S_k$ ) die Partikel mit den durch die Variation des Bewegungsmodells gegebenen unterschiedlichen Trajektorien, als Lösung für die auftretenden Korrespondenzprobleme. Die Bewertung dieser prädiktierten Lösungen in Form der unterschiedlichen Trajektorien ist Bestandteil des Beobachtungsmodells. Dieses liefert für die Partikel  $s_k^{n*}$  und ihre enthaltenen Trajektorien den Glattheits- und Konsistenzfehler, der in Form der Partikelgewichte  $\pi_k^{n*}$  in Bezug auf die gesamte Partikelmenge  $S_k$  die Likelihood-Funktion des Zustandes  $\mathbf{z}_k$  bestimmt.

Die Modifikation der A-priori-Wahrscheinlichkeit  $p(\mathbf{z}_k | \mathbf{m}_{k-1})$  nach dem Satz von Bayes findet über diese Approximation der Likelihood-Funktion  $p(\mathbf{m}_k | z_k^n)$  für jedes Partikel  $s_k^{n*}$  der Partikelmenge  $S$  durch den Sampling-Schritt statt. Dabei beinhaltet die Partikelmenge  $S$   $N$ -Partikel ( $s_k^{n*}$ ) ( $n^*=100$ ) mit  $n$  verschiedenen Zuständen  $z_k^n$ . Jedes Partikel  $s_k^{n*}$  wird ein Gewicht  $\pi_k^n$  zugeordnet.

$$\pi_k^{n*} = \frac{p(\mathbf{m}_k | z_k^n)}{\sum_{n=1}^N p(\mathbf{m}_k | z_k^n)} \quad (5-3-12a)$$

Die Gewichte sind entsprechend der Gleichung (5-3-12a) normiert, dass sie sich zu Eins summieren, so dass aus ihnen direkt die Kulminationskoeffizienten  $c_k$  Gleichung (5-3-12 b) berechnet werden können.

$$c_k^{n*} = \begin{cases} 0, & \text{wenn } n = 0 \\ c_k^{n*-1} + \pi_k^{n*}, & \text{sonst} \end{cases} \quad (5-3-12b)$$

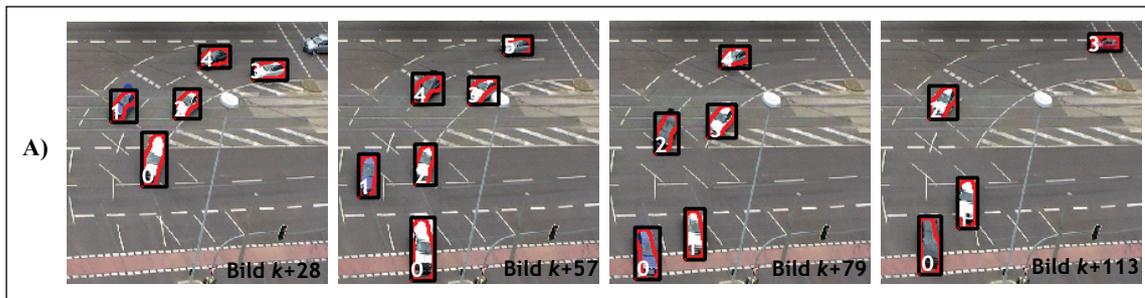
Das *resampling* wird dementsprechend durch den CONDENSATION-Algorithmus ausgeführt und dadurch eine neue Partikelmenge  $S_{k+1}$  mit  $N$  Partikeln  $s_{k+1}^{n*}$  erzeugt, deren Wahrscheinlichkeitsverteilung der A-posteriori-Wahrscheinlichkeit  $p(\mathbf{z}_k | \mathbf{m}_k)$  entspricht. Ein Vorteil dieser Methode zur Aktualisierung der Partikelmenge ist es, dass sich die Anzahl der möglichen Zustände verringert, denn Zustände mit einer geringen Auftrittswahrscheinlichkeit und kleinen Gewichten werden nicht in die neue Partikelmenge übernommen. Das wirkt dem Degeneracy Problem entgegen. Denn durch die begrenzte Anzahl der Partikel ist die Abdeckung der „interessanten“ Zustände besser, wenn sehr unwahrscheinliche Zustände aus der Partikelmenge entfernt und nicht weiter durch das Bewegungsmodell weiter verfolgt werden. Somit bleibt immer eine gewisse Varianz der A-posteriori-Wahrscheinlichkeit erhalten.

## 5.4 Untersuchungen und Ergebnisse

Im Folgenden werden Untersuchungsergebnisse vorgestellt, die durch Anwendung der entwickelten Datenfusionsstruktur gewonnen wurden. Dazu wurden Resultate aus realen wie auch aus künstlich erzeugten Bildsequenzen betrachtet, die komplizierte Situationen von Verkehrsaufnahmen darstellen. Die Korrespondenzprobleme, die aus den bildspezifischen Konfliktsituationen resultieren, werden beschrieben und analysiert, um die Leistungsfähigkeit des entwickelten Systems zur Datenfusion bei der Analyse von realen Bildsequenzen nämlich genau bei dem Einfluss von nicht-kooperativen Messsituationen, zu demonstrieren.

### 5.4.1 Szene ohne Konfliktsituation

Die erste Bildsequenz entstand durch die Aufnahme einer Verkehrsszene, die zunächst der Erklärung der Verarbeitungskette des entwickelten Systems dient. Vorteilhafte Eigenschaften für die Analyse sind die hohe Bildhelligkeit sowie geringe perspektivisch bedingte Größenverzerrungen der einzelnen Objekte. Nachteilig wäre hier die höhere Anzahl von Objekten, die zu den Problemen bei den Datenassoziationen führen können. Die Bewegungssegmentierung mittels des MD-Ansatzes im Abschnitt 5.2.1 verläuft vollständig automatisch und liefert in jedem Zeitschritt die optimal separierten MBs, die als Merkmalsträger für die Zustandsbeschreibung (s. Abschnitt 5.2) dienen (Abb. 5-4-1a). Da die betrachtete Szene keine Konfliktsituationen enthält, werden sich die MBs konstant verhalten (Abb. 5-4-1a). Mit dem Zustandsmodell stellt jeder MB einen Knoten dar, der sich durch Bewegungsvektorfelder, Farbhistogramme und Flächen charakterisieren lässt.



**Abbildung 5-4-1a:** Multi-Objektverfolgung mit Partikelfilter ohne Konfliktsituationen. Dieser Bildsequenzausschnitt beinhaltet die aus der Bewegungssegmentierung erhaltenen MBs.

Anhand dieser resultierenden Informationen werden mit dem Fuzzy-Bewegungsmodell zunächst die Zuordnungen der Trajektorien zu den Objekten geschätzt, indem die Partikel nach der Vorschrift des Bewegungsmodells transformiert und variiert werden. Aufgrund der resultierenden Gatingsmatrix ist die Zuordnung von Trajektorien und Motion-Blobs eindeutig, welche dazu führt, dass der Connect-Operator aktiviert wird. Dazu werden anhand der Bewegungsvektordinformationen MT1 und MT2 die MBs mit den Trajektorien durch den Connect-Operator zugeordnet. Der Ausgang  $y_c$  des Connect-Operators ist in der Tabelle 5-4-1 angegeben. Der Wertebereich des Connect-Operator liegt im Intervall  $[0.167 \ 0.733]$ , was es ermöglicht, dass bei genügend hoher Varianz

## 5. DATENFUSION AUF DER BASIS VON PARTIKELFILTERN

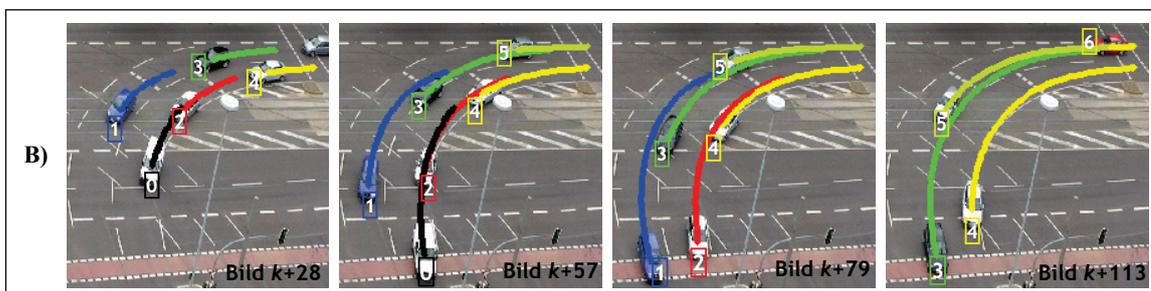
des Schwellwertes  $H_S$  Verbindungen zustande kommen, die in dem Bewegungsmodell nicht vorgesehen waren.

**Tabelle 5-4-1:** Zuordnungstabelle als Ausgang des Connect-Operators mit  $y_c$  als Ausgang. Alle Verbindungen werden grün unterlegt. MT1 und MT2 vgl. Gleichung (5.2.2c) und (5.2.2d)

		Motion-Blobs ( $b_i$ )														
		0			1			2			3			4		
$T_i$		$y_c$	MT1	MT2	$y_c$	MT1	MT2	$y_c$	MT1	MT2	$y_c$	MT1	MT2	$y_c$	MT1	MT2
0		0.73	0.87	0.94	0.17	0.00	0.00	0.17	0.00	0.00	0.17	0.00	0.00	0.17	0.00	0.00
1		0.17	0.00	0.00	0.73	0.93	0.83	0.17	0.00	0.00	0.17	0.00	0.00	0.17	0.00	0.00
2		0.17	0.00	0.00	0.17	0.00	0.00	0.73	0.83	0.87	0.17	0.00	0.00	0.17	0.00	0.00
3		0.17	0.00	0.00	0.17	0.00	0.00	0.17	0.00	0.00	0.17	0.00	0.00	0.73	0.93	1.0
4		0.17	0.00	0.00	0.17	0.00	0.00	0.17	0.00	0.00	0.73	0.93	0.83	0.17	0.00	0.00

Für den Schwellwert  $H_C = N(\mu=0.5, \sigma^2=0.12)$  entstehen bei einer Partikelmenge mit 100 Partikeln 1 Partikel bei dem die Trajektorien nicht mit dem MB verbunden werden. Für dieses Partikel ergeben sich ein Gesamtglattheitsfehler von 13,41 und ein Konsistenzfehler von 23,72. Verglichen mit den 99 Partikeln (vgl. Tabelle 5-4-1), die jeweils einen Glattheitsfehler von 14,00 und einen Konsistenzfehler von 6,00 aufweisen, zeigt sich, dass der Konsistenzfehler für die Bewertung unterbrochener Trajektorien die entscheidene Rolle spielt. Dementsprechend besitzt das eine Partikel einen kleineren Gewichtswert als die 99 korrekt geschätzten. Durch die sehr geringe Auftrittswahrscheinlichkeit von 0,01 und dem kleinen Gewichtswert wird das Partikel mit der falschen Schätzung im Sampling-Schritt nicht in die neue Partikelmenge aufgenommen. So dass die A-posteriori geschätzte Partikelmenge Partikel besitzt, in denen eine eindeutige Korrespondenz zwischen den Trajektorien und den MBs hergestellt wurde.

Da durch den Connect-Operator die Korrespondenzzuordnung zwischen den Trajektorien und den MBs eindeutig hergestellt wurde, werden der Split- und Merge-Operator nicht ausgeführt. Weil keine Konfliktsituationen in der Zuordnungstabelle auftreten, ist es nicht überraschend, dass keine Störungen in den ermittelten Objekttrajektorien vorkommen (Abb. 5-4-1b). Aus diesen Ergebnissen und dem Sachverhalt ist das entwickelte System, das auf Fuzzy-Bewegungs- und Beobachtungsmodell beruhende Datenfusion von Bilddaten prinzipiell gültig.



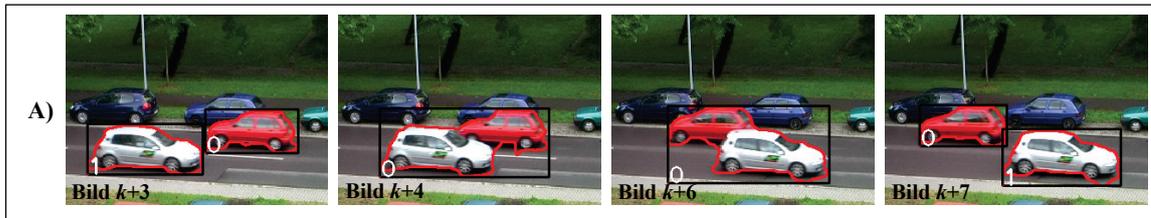
**Abbildung 5-4-1b:** Multi-Objektverfolgung mit Partikelfilter ohne Konfliktsituationen. Die durch das Partikelfiltersystem ermittelten Objekttrajektorien werden farblich dargestellt.

### 5.4.2 Szenen mit einfachen Konfliktsituationen

Reale Bildszenen enthalten im dichten Straßenverkehr mehrere Objekte mit Konfliktsituationen, die in der Regel ein Korrespondenzproblem bei der zeitlichen Datenzuwei-

## 5. DATENFUSION AUF DER BASIS VON PARTIKELFILTERN

sung verursachen. Solche problematischen Situationen treten nicht nur durch Teilverdeckung, sondern auch bei den Beleuchtungsänderungen, Schattenwurf und Interaktionen von Objekten in der Straßenverkehrsszene auf. Dies kann in der Abbildung 5-4-2a demonstriert werden, in der sich zwei Fahrzeuge auf einer Straße begegnen. Dieser Sachverhalt führt zu einem Korrespondenzproblem aufgrund der Teilverdeckung (s. Abb. 5-4-2a, Bild  $k+3$  bis Bild  $k+6$ ). Um das Problem zu lösen, sollen nun die Objekte zunächst mittels des MD-Ansatzes erfasst und je einem MB zugeordnet werden (Bild  $k+2$ ).



**Abbildung 5-4-2a:** Dargestellt ist hier die Verschmelzung und Einteilung von Objekten aufgrund der Verdeckung in realen Bildsequenzen.

Aufgrund der Teilverdeckung werden die MBs verschmolzen, sodass der resultierende MB mehrere Objekte beinhalten kann. Daraus wird die Zuordnung von einzelnen Objekten in der MB-Korrelationsebene nicht ermöglicht. Um diese Zuordnungsaufgabe lösen zu können, enthält das vorgeschlagene System zur Datenfusion das Fuzzy-Bewegungsmodell, das Partikel mit Zuständen erzeugt, die die prädiktierten Lösungen des Zuordnungsproblems gemäß des Fuzzy-Systems im Abschnitt (5.3.2) aufweisen. Dadurch wird durch das Fuzzy-Bewegungsmodell eine Menge von möglichen Zuständen (in dem Kurzzeitraum) erzeugt, die in wahrscheinlichkeitstheoretischer Sicht das prädiktierte A-priori-Wissen  $p(\mathbf{z}_k | \mathbf{m}_{k-1})$  repräsentieren.

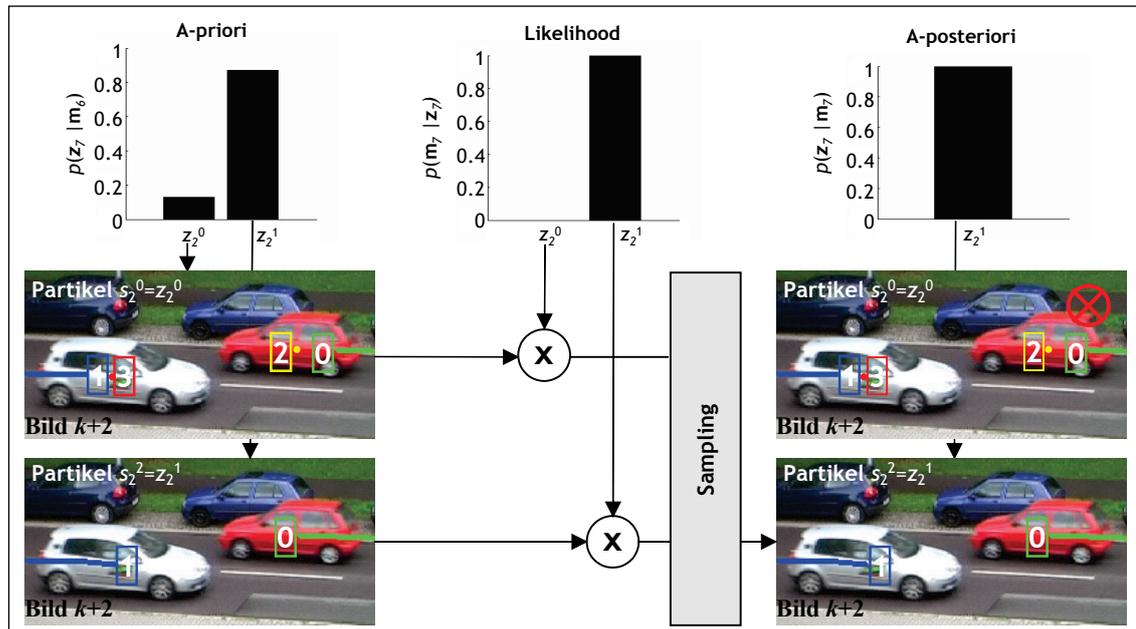
Die Lösungen bilden eine Variation von Zuständen, die nicht nur aus der möglichen Zuordnung des MBs 0 aus dem Bild  $k+6$  zu dem MBs 0 und 1 aus dem Bild  $k+7$  bestehen, sondern zusätzlich in der Aufteilung der Elterntrajektorien im Bild  $k+2$  auf die MBs, da theoretisch jedem Blob keine, eine oder beide Elterntrajektorien zugeordnet werden können. Die Lösung der Objektverfolgung ist durch das Partikel mit dem größten Likelihood-Wert gegeben.

Die Abbildung 5-4-2b und 5-4-2c zeigen die Wahrscheinlichkeitsverteilungen für die Entwicklung der Partikelmenge des Bildes  $k+2$  (Abb. 5-4-2b) und  $k+7$  (Abb. 5-4-2c). Die Prädiktion der Zustände durch das Bewegungsmodell aus dem Bild  $k+1$  bildet die A-priori-Wahrscheinlichkeit  $p(\mathbf{z}_2 | \mathbf{m}_1)$  und ist im linken Teil des Bildes dargestellt. Partikel  $s_2^0$  und  $s_2^2$  sind die Partikel, die alle auftretenden Zustände des Systems zu diesem Zeitpunkt enthalten. Der Zustand  $z_2^1$  des Partikels  $s_2^2$  ist die ungestörte Fortführung der Trajektorien 0 und 1 und besitzt auch die höchste A-priori-Wahrscheinlichkeit.

In Partikel  $s_2^0$ , das den Zustand  $z_2^0$  besitzt, werden für den roten PKW und den weißen PKW neue Trajektorien (2, 3) erzeugt. Diese mit einer geringen A-priori-Wahrscheinlichkeit versehene Möglichkeit, wurde durch den im Bewegungsmodell enthaltenen Rauschanteil generiert. Dieser bewahrt eine gewisse Variation, so dass auch Zustände, die in der Nähe der besten Lösung des Bewegungsmodells liegen, gebildet werden. Durch diese Variation können Fehler in der Definition des Bewegungsmodells ausgeglichen werden. Dies ist eine wesentliche Eigenschaft in dieser entworfenen

## 5. DATENFUSION AUF DER BASIS VON PARTIKELFILTERN

Datenfusionsstruktur. Mit dem Zeitverlauf wird die Möglichkeit der Zustandsvariationen immer größer, weil mit jedem Bild neue zusätzliche Zuordnungsmöglichkeiten im Betracht gezogen werden.



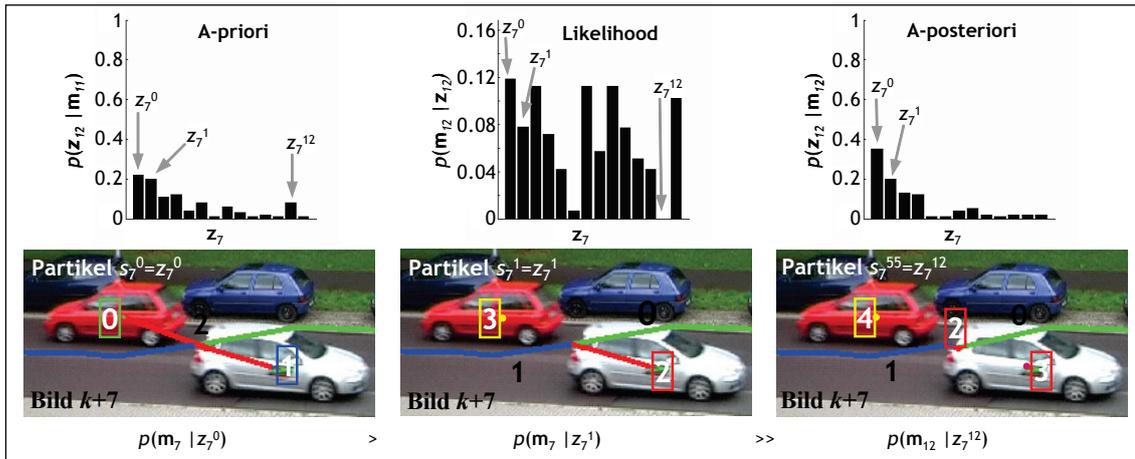
**Abbildung 5-4-2b:** Wahrscheinlichkeitsbasierte Darstellung der konfliktfreien Objektverfolgung mit dem Partikelfilter. Die Zustände der A-priori-Wahrscheinlichkeit werden durch das Bewegungsmodell erzeugt. Die Streuung der Zustände ist nicht sehr hoch, so dass keine Konflikte auftreten. Durch die Likelihood-Funktion, die durch die Partikelgewichte bestimmt wird und dem Sampling-Schritt ist die A-posteriori-Wahrscheinlichkeit nach dem Satz von Bayes gegeben. Es zeigt sich, dass durch den Sampling-Schritt Partikel mit sehr kleinen A-posteriori-Wahrscheinlichkeiten aus der Partikelmenge entfernt werden, um so dem Degeneracy Problem entgegen zu wirken.

Allein durch den Connect-Operator ist eine Variation<sup>113</sup> von  $2^{(N_T \times N_M)}$  maximal gegeben. Im Abbildung (5-4-2c) erkennt man die 13 verschiedenen Zuordnungsmöglichkeiten, wobei das Partikel  $s_7^0$  die korrekte Zuordnung mit dem Zustand  $z_7^0$  enthält. Die Bewertung dieser Zustände (Variationen) wird durch das Beobachtungsmodell im Sinne der Likelihood durch die Berechnung der Glattheits- und Konsistenzfehler der erzeugten Trajektorien durchgeführt und erzeugt damit die Partikelgewichte  $\pi_k^n$ , die als Ausgangsmenge für den Prädiktionsschritt im nächsten Bild Verwendung findet. Die Anpassung des A-priori-Wissens durch die Likelihood-Funktion der Partikel wird über den Satz von Bayes berechnet und ergibt die A-posteriori-Wahrscheinlichkeit  $p(\mathbf{z}_k | \mathbf{m}_k)$ . Dadurch können im zeitlichen Sinne punktuelle fehlerhafte Schätzungen der A-priori-Wahrscheinlichkeit oder durch das Bewegungsmodell falsch prädizierte Zustände kompensiert werden. Das ist in den einfachen Szenen noch nicht zu beobachten, da das Bewegungsmodell in der Regel eine gute Prädiktion liefert. Die Verarbeitung der prädizierten Partikelmenge durch die erzeugte Gewichtung der Partikel findet im Sampling-Schritt statt. Dieser entfernt durch die Erzeugung einer neuen Partikelmenge

<sup>113</sup>  $N_T$  ist die Anzahl der Trajektorien und  $N_m$  die Anzahl der Motion-Blobs (Messungen)

## 5. DATENFUSION AUF DER BASIS VON PARTIKELFILTERN

Zustände, die sehr geringe Wahrscheinlichkeiten besitzen und wirkt so dem Degeneracy Problem entgegen.



**Abbildung 5-4-2c:** Wahrscheinlichkeitsbasierte Darstellung der Objektverfolgung in der Konfliktsituation (Teilung) mit dem Partikelfilter. Es zeigt sich, dass die Streuung der Zustände bei Konfliktsituationen wesentlich höher ist, da die Anzahl der möglichen Trajektorienverknüpfungen steigt. Einige der möglichen Zustände sind dargestellt.

Abbildung 5-4-2d beschreibt die Objektverfolgung für die gesamte Verkehrsszene. Dabei werden die Trajektorien durch die abrupte Verschiebung der MB-Schwerpunkte bei der Verschmelzung und Teilung sehr ungenau und geben ungenügend die Bewegung des Objektes in der Szene wieder (5-4-2d/ Reihe B). Um die Bewegungstrajektorien glätten zu können, wurden die Trajektorien durch B-Splines (Abbildung 5-4-2d/ Reihe C) nachbearbeitet.

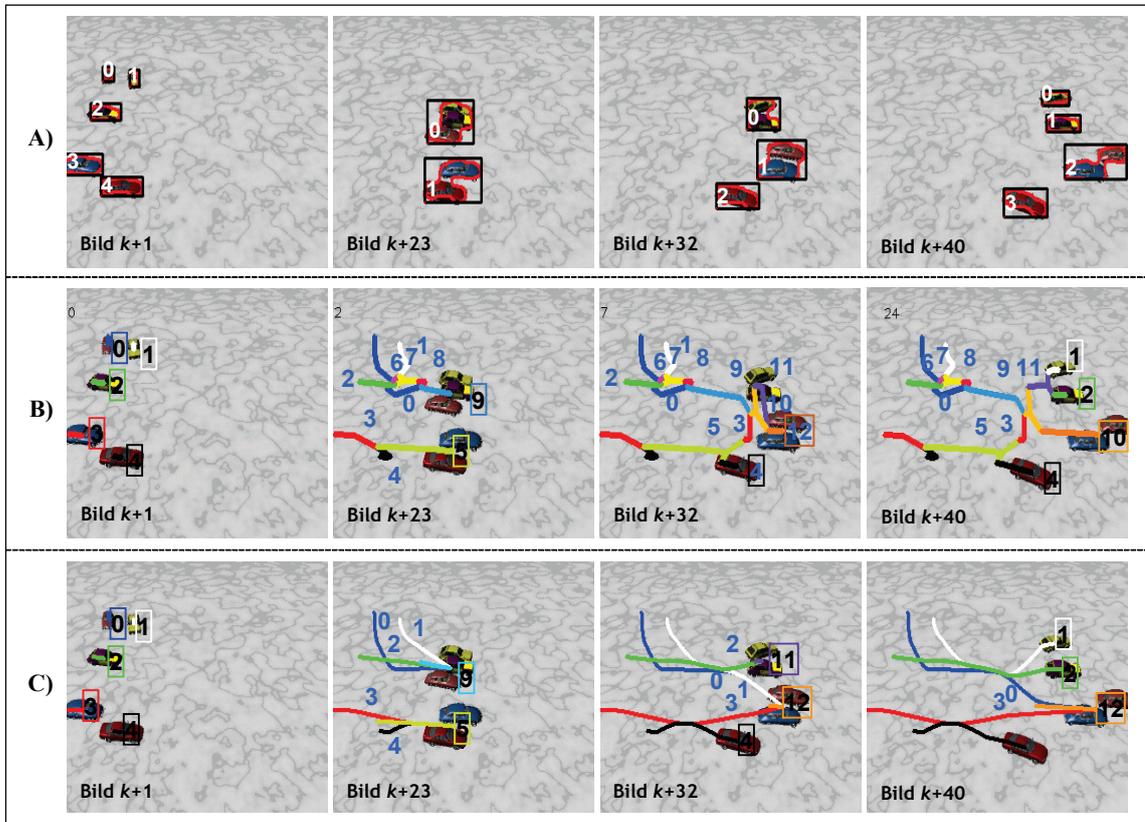


**Abbildung 5-4-2d:** Objektverfolgung mit dem Partikelfiltersystem für eine einfache Verkehrsszene. A) Bewegungsegmentierung mit dem MD-Ansatz. B) Trajektorien des Partikels mit dem größten Likelihood-Wert. C) B-Spline geglättete Trajektorien des Partikels mit dem größten Likelihood-Wert.

Die Grundidee basiert darauf, dass mit den Trajektorienknoten als Kontrollpunkte für einen B-Spline<sup>114</sup>, der Bewegungsverlauf gut angepasst werden kann, dabei werden Kontrollpunkte von Verschmelzungen geringer gewichtet als Kontrollpunkte (de Boor-Punkte) von einzelnen Objekten.

### 5.4.3 Szenen mit gehäuften Konfliktsituationen

Die Leistungsfähigkeit der vorgeschlagenen Datenfusionsstruktur in Situationen, in denen eine größere Anzahl sich dicht nebeneinander bewegender Objekte vorliegt, war Untersuchungsgegenstand weiterer Bildfolgen. Hierbei wurden enthaltene Störungen, wie Objektinteraktionen wie Teilverdeckungen, die Verschmelzungen oder Zerfallen einzelner Motion-Blobs, untersucht<sup>115</sup>. Ein Beispiel für komplexere Interaktion der Objekte auf der MB-Ebene ist in der Abbildung 5-4-3 dargestellt<sup>116</sup>. Das Bewegungsmodell muss hier Zustände mit MBs verarbeiten, in denen bis zu drei Objekte enthalten sind, die untereinander ausgetauscht werden können, siehe Bild  $k+23$  und Bild  $k+32$ .



**Abbildung 5-4-3:** Objektverfolgung mit dem Partikelfiltersystem in einer synthetischen Verkehrsszene mit sehr dichtem Verkehr. A) Bewegungssegmentierung. B) Trajektorien des Partikels mit dem größten Likelihood-Wert. C) B-Spline-geglättete Trajektorien des Partikels mit dem größten Likelihood-Wert.

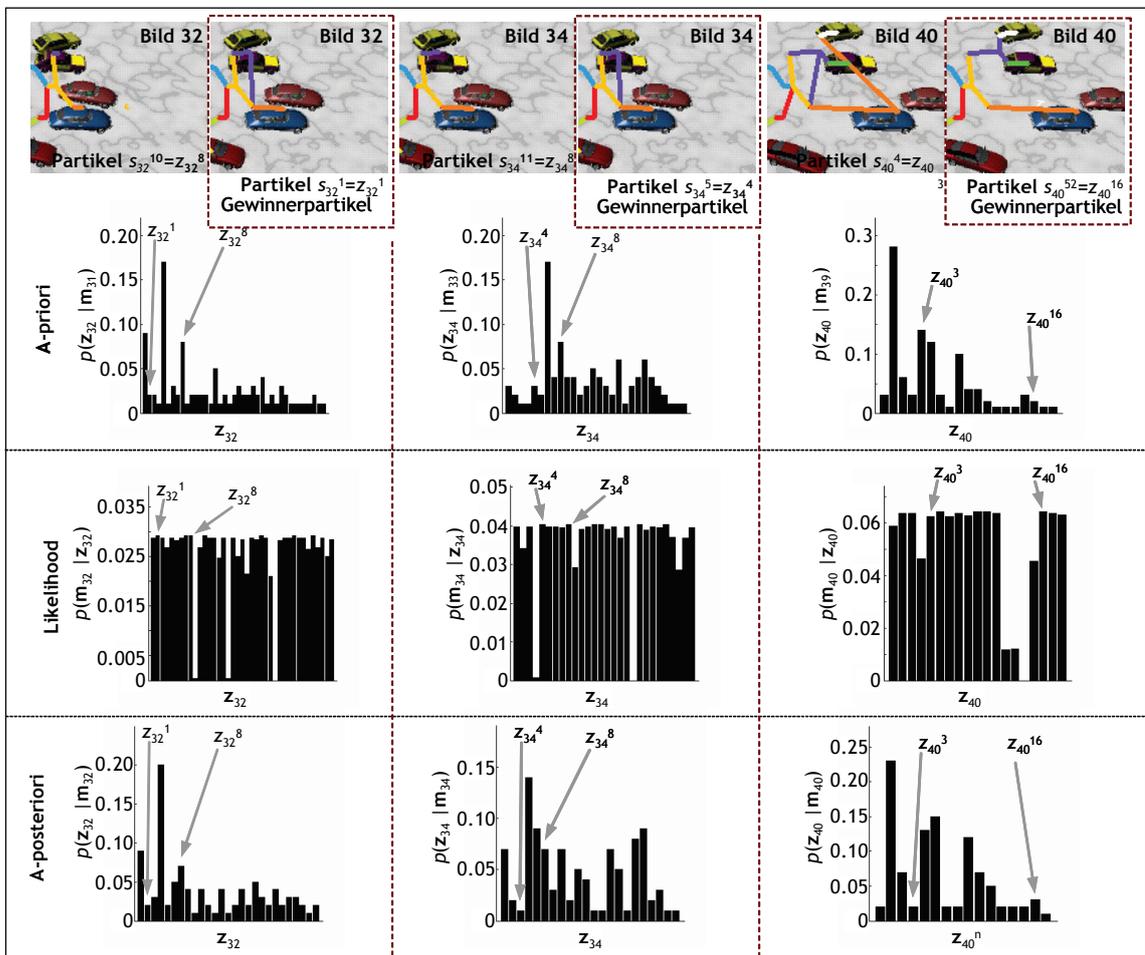
<sup>114</sup> Im Kontext numerischer Verfahren, wo Splines häufig eingesetzt werden, ist die Wahl der Basis entscheidend für eventuelle Rundungsfehler und damit für die praktische Einsetzbarkeit. Diese so genannten B-Spline-Basisfunktionen haben einen kompakten Träger, sie sind nur auf einem kleinen Intervall von Null verschieden. Änderungen der Koeffizienten wirken sich also nur lokal aus.

<sup>115</sup> Weitere Analysen und Darstellungen sind im Anhang 5 und 6 zu finden.

<sup>116</sup> Da es hier erstens um die Leistungsfähigkeit und zweitens um die Validierung des Systems unter dem Einfluss von gestörten Situationen geht, wurde mit einer synthetischen Bildsequenz gearbeitet, die die Interaktionen von Objekten verdeutlicht.

## 5. DATENFUSION AUF DER BASIS VON PARTIKELFILTERN

Die Entwicklung der Partikelmenge ab Bild  $k+32$  für diese Szene ist in Abbildung 5-4-4 dargestellt. Es zeigt sich, dass in Bild  $k+32$  das Partikel  $s_{32}^1$  mit dem Zustand  $z_{32}^1$  als Gewinnerpartikel ausgewählt wurde. Eine derart komplexe Situation kann dazu führen, dass das Partikel mit dem größten Likelihood-Wert, wie in Bild  $k+32$  (5-4-3B) nicht dem Trajektoriezustand mit dem tatsächlichen Verlauf der Fahrzeuge entspricht. Denn es ist zu erkennen, dass der dargestellte Zustand nicht der Realität entspricht. Das bedeutet, dass das Beobachtungsmodell ungenau ist, da es diesem Partikel einen höheren Likelihood-Wert, als dem Partikel  $s_{32}^{10}$  mit dem Zustand  $z_{32}^8$ , der die Realität am Besten widerspiegelt und der eine geringfügig kleinere Likelihood besitzt. Es sind aber beide Partikel in der Partikelmenge des Systems enthalten. Die Weiterentwicklung des Partikels  $s_{32}^{10}$  ist der Partikel  $s_{34}^5$  im Bild  $k+34$ . Auch dieser Partikel wird vom Beobachtungsmodell mit einem höheren Likelihood-Wert bewertet, als der Partikel  $s_{34}^{11}$ , welcher die Weiterentwicklung des Partikels  $s_{32}^{10}$  aus dem Bild  $k+32$  bewertet. Es zeigt sich aber in Bild  $k+40$ , dass die Weiterentwicklung des Partikels  $s_{32}^{10}$  zu  $s_{34}^{11}$  und schließlich zu  $s_{40}^{52}$  einen Zustand erzeugen, der einen höheren Likelihood-Wert besitzt als die Nachfolger der Partikel  $s_{32}^1$  und  $s_{34}^5$ .

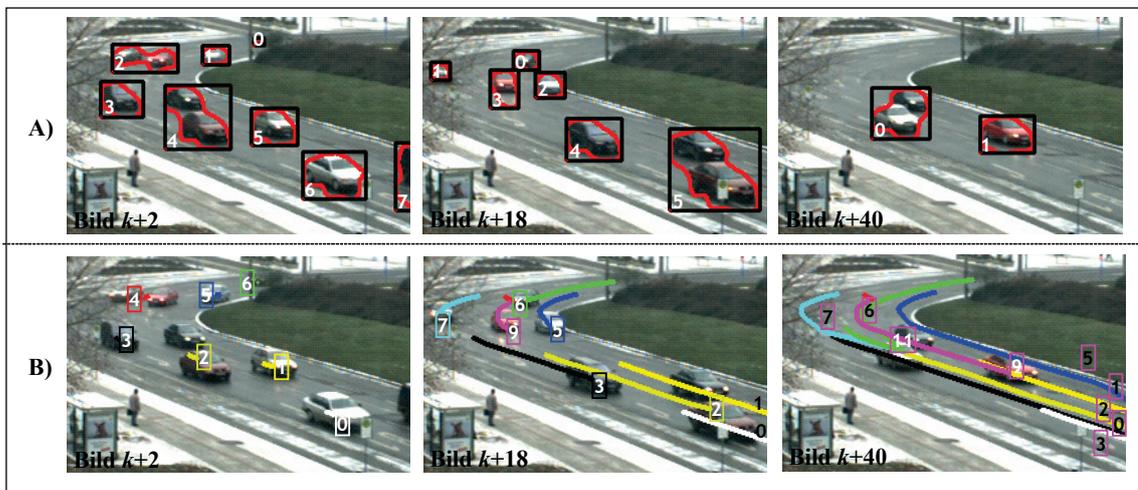


**Abbildung 5-4-4:** Darstellung der Entwicklung der Partikelmenge des Bild  $k+32$ -  $k+40$ . Ab Bild  $k+32$  (links) entspricht der Gewinnerpartikel, d.h. der Partikel mit dem höchsten Likelihood-Wert, nicht mehr dem Verlauf der Szene. Der Zustand, der dem Verlauf der Szene entspricht (obere Bilderreihe), ist aber noch in der Partikelmenge enthalten und erlangt erst im Bild  $k+40$  die maximale Likelihood (rechts).

## 5. DATENFUSION AUF DER BASIS VON PARTIKELFILTERN

Bildlich gesehen stand das System im Bild  $k+32$  vor der Entscheidung zwischen fast gleichwahrscheinlichen Möglichkeiten, um die Objekte zu verfolgen. Würde ein nicht stochastisches System zur Modellierung des Zustandes benutzt, so wäre die Verfolgung an dieser Stelle fehl gelaufen. Der Vorteil des Partikelfilters ist es also, dass der Zustand als stochastisches Modell betrachtet wird, welches robust auf Fehlentscheidungen durch die Prädiktion reagiert. Die Ausgangsdaten der Multi-Objektverfolgung mit dem Partikelfilter bestehen somit nicht nur aus dem Gewinnerpartikel, die Information aus der Objektverfolgung ist vielmehr in der gesamten Partikelmenge enthalten.

Die Abbildung 5-4-5 zeigt eine reale komplexe Verkehrssituation, wie sie in Abbildung 5-4-3b künstlich erzeugt wurde. Durch die relativ hohe Verkehrsdichte werden auch hier in der Bewegungssegmentierung (Reihe A) mehrere Fahrzeuge in einem MB zusammengefasst. Dadurch entstehen komplexe Trajektorienverläufe (Reihe B) mit vielfältigen Split- und Merge-Situationen. Die Bewegungsverläufe können erfasst und dargestellt werden.

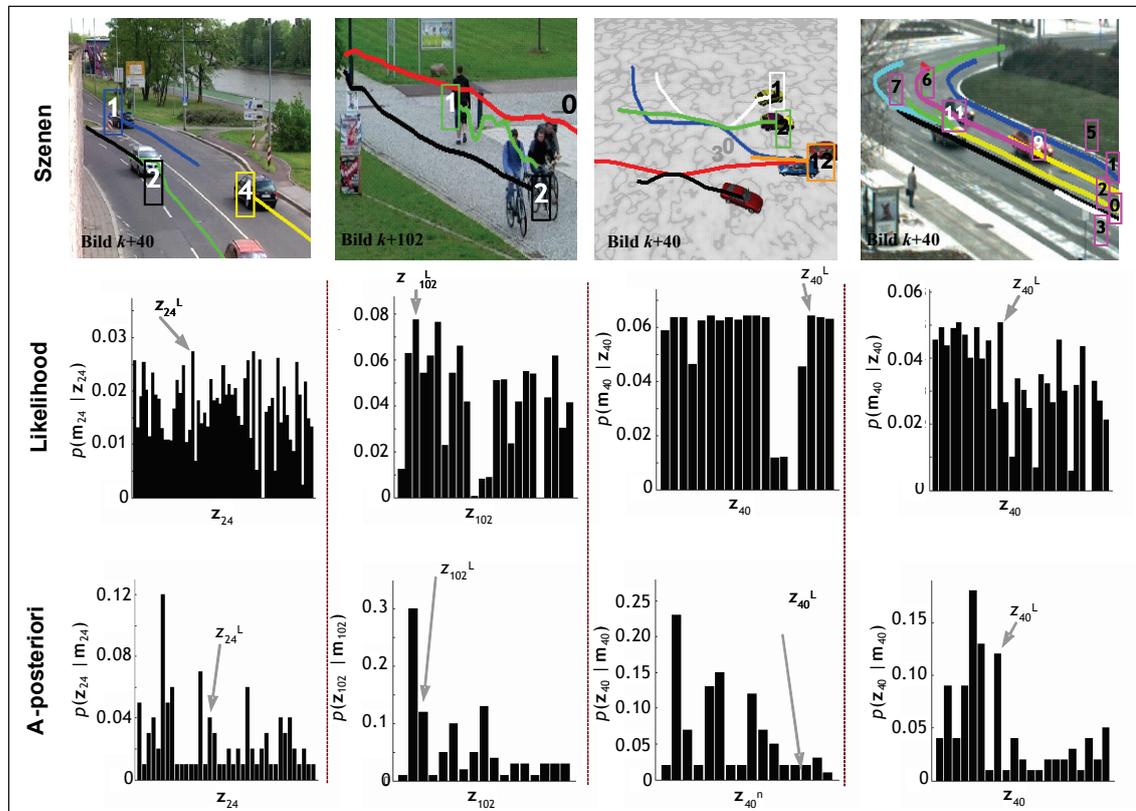


**Abbildung 5-4-5:** Objektverfolgung mit dem Partikelfiltersystem in einer Verkehrsszene mit sehr dichtem Verkehr. A) Bewegungssegmentierung. B) Trajektorien des Partikels mit dem größten Likelihood-Wert. C) B-Spline-geglättete Trajektorien des Partikels mit dem größten Likelihood-Wert.

In Abbildung 5-4-6 sind die A-posteriori- und Likelihood-Funktionen der entwickelten Partikelmengen für die Szenen mit gehäuften Konfliktsituationen dargestellt. Die Likelihood wird also für die Auswertung der Information Partikelmenge benutzt und erzeugt das Gewinnerpartikel.

In den Diagrammen fällt auf, dass die Gewinnerpartikel nicht die größte A-posteriori-Wahrscheinlichkeit besitzen. Das weist darauf hin, dass das Bewegungsmodell in diesen Stellen ungenau ist. Die Prädiktion durch den optischen Fluss stimmt nur bedingt mit der Realität überein. Deshalb bedarf es einer Anpassung des Fuzzy-Systems. Daraus zeigt sich die in Abschnitt (5.3) beschriebene gegenseitige Unterstützung der intensitätsbasierten und merkmalsbasierten Daten. Während in Abbildung 5-4-6 gezeigt wurde, dass bei dem Beobachtungsmodell, mit merkmalsbasierten Daten, Ungenauigkeiten dazu führen, dass das Gewinnerpartikel falsch ausgewählt wurde, hat sich dieser Fehler durch das Bewegungsmodell kompensiert. Die korrekten Trajektorien sind in Bild  $k+40$  dargestellt. Es zeigt sich aber auch, dass die Bewegungsprädiktion bei komplexen Szenen im Gegensatz zu einfachen Szenen ungenauer wurde, d.h. das

Maximum der A-priori-Wahrscheinlichkeit ist nicht das Partikel mit der höchsten Likelihood. Somit kompensieren in umgekehrter Weise die merkmalsbasierten Daten durch das Beobachtungsmodell die Prädiktion aus dem Bewegungsmodell. (Weitere Analysen und Darstellungen sind im Anhang 8.5 und 8.6 zu finden).



**Abbildung 5-4-6:** Wahrscheinlichkeitsverteilungen der Partikelmengen am Ende der Szenen 1-4. Der Zustand des Gewinnerpartikels  $z^L$  mit dem maximalen Likelihood-Wert besitzt nicht zwangsläufig die maximale A-posteriori-Wahrscheinlichkeit. Dies ist ein Hinweis darauf, dass die stochastischen Eigenschaften der vorkommenden Konfliktsituation durch das Bewegungsmodell ungenau wiedergegeben werden.

### 5.5 Schlussfolgerung und Zusammenfassung

Das Hauptziel dieses Kapitels ist es, mit dem vorgestellten Optischen Flussverfahren aus Kapitel 2 und Verfahren aus Kapitel 3 ein hybrides Multi-Objektverfolgungssystem zu entwickeln. Dies wird durch eine intelligente Kombination von intensitäts- und merkmalsbasierten Daten erreicht. Dazu wurde ein Partikelfiltersystem auf der Grundlage des Condensation-Algorithmus entwickelt, das durch Abstrahierung der Probleme der auftretenden Konfliktsituationen ein Zustandsmodell verwendet, in dem diese Probleme als Rauschen bzw. Messungenauigkeit modelliert werden. Somit konnte eine Bayes-Filterung auf den Prozess der Objektidentifikation, die als Datenassoziation bezeichnet wird, angewendet werden. Zusätzlich bot die Struktur des Partikelfilters eine optimale Möglichkeit, intensitäts- und merkmalsbasierte Daten in einem System zu integrieren. Durch die Verwendung eines Fuzzy-Systems in der Zustandsprädiktion wurde zusätzliches Expertenwissen in das System eingebracht, so dass die Variationen der Zustände über ein Expertensystem reguliert werden können, die sich aus dem intensitätsbasierten Daten und den Farbdaten ergeben.

### 5.5.1 Diskussion der vorgeschlagenen Datenfusionsstruktur

Die Untersuchungen der Testfarbbildsequenzen haben gezeigt, dass die angestrebten Zielstellungen der Forschungsarbeit mit der vorgeschlagenen Datenfusionsstruktur grundsätzlich erreicht wurden und sich wie folgt darstellen lassen. Um das Prinzip zu vereinfachen, wird die Datenfusionsstruktur zur Multiobjektverfolgung in drei Verarbeitungsschritte geteilt.

*1. In dem Segmentierungsschritt werden alle Objekte mittels des MD-Ansatzes erfasst und initialisiert.*

Der angewandte MD-Ansatz hat es ermöglicht, auf geeignete Weise die Bilddaten so zu reduzieren, dass ausschließlich bewegte Regionen in der Bewegungsanalyse betrachtet werden. Es konnte durch die Kombination zweier Differenzbilder eine Verbesserung gegenüber der einfachen Differenzbildtechnik erzielt werden, welche nicht in der Lage ist genaue Konturen von bewegten Objekten aufgrund des Ghosting-Effektes zu ermitteln. Weiterhin wurden Erweiterungen durch die Anwendung einer Hystersebinarisierung realisiert, da die normale Differenzbildtechnik relativ anfällig gegenüber Störungen ist, die durch Kamerarauschen oder atmosphärische Störungen hervorgerufen werden.

Allerdings enthält die umfangreiche Zusammensetzung des entwickelten Systems aus den verschiedensten Methoden Stellen, die in optimierter Weise zur Verbesserung des Systems beitragen können. Der Schwerpunkt dabei ist die Verbesserung der Segmentierungsergebnisse durch die Integration eines Schatteneliminierungsfilters oder durch die Einbeziehung von 3D-Tiefdaten aus einem Stereokamerasystem, um eine robuste Knotenformbeschreibung zu ermöglichen. Ebenso sind Methoden zur Entfernung anderer Störeffekte wie reflektierte Fahrzeugbeleuchtung, die zur Verschmelzung der Blobs mehrerer Objekte führen, erforderlich. Für die Analyse von Szenen mit direktem aber auch indirektem Schatten ist eine weitere Ausnutzung des adaptiven Farbraumes [Al-Hamadi 01] denkbar. Die Untersuchungen der Arbeit beschränkten sich gegebenenmaßen auf gleichmäßig beleuchtete Szenen. Die verwendeten Farbräume enthalten noch den Helligkeitsanteil. Mit der Abkopplung dieser Komponente können erweiterte Objektbeschreibungen aufgestellt werden, die invariant gegen Beleuchtungsschwankungen sind und somit den Einsatzbereich des Verfahrens erweitern.

*2. Im Datenaufbereitungsschritt werden Merkmale aus den erfassten Regionen aus dem Segmentierungsschritt extrahiert, die als farbbasierte, formbasierte und bewegungsbasierte Merkmale für den nächsten Schritt vorliegen.*

Durch die Datenaufbereitung wurden das Bewegungsvektorfeld, das Farbhistogramm, die umgebende Ellipse, die prädizierte Ellipse und der Flächeninhalt zur Beschreibung eines durch den MD-Ansatz erfassten Knotens angewendet. Dadurch wurden redundante und unwichtige Merkmale unterdrückt, so dass diese extrahierten Merkmale die gesamte relevante Information des Knotens besitzen, um eine zuverlässige Zuordnung in dem letzten Schritt der Bewegungsanalyse zu ermöglichen.

Die Verfahren zur Bestimmung des Bewegungsvektorfeldes benutzen die Approximation der Bildwerttreue zwischen zwei Bildern als Maß für die Bewegung. Zur Ermittlung des Flussvektorfeldes ist dies richtig, solange die Helligkeit eines Objektes zwischen zwei Bildern sich nicht ändert. Diese Annahme ist jedoch nicht immer erfüllt, da in praktischen Anwendungen zum einen unvermeidlichen Beleuchtungsänderung

auftauchen können, zum anderen größere Verschiebungen einen negativen Einfluss haben. Die hier eingesetzte pyramidale Lösungsstrategie ermöglicht das LK-Verfahren auch auf größere Verschiebungen anzuwenden ohne Abstriche an dessen Robustheit zu machen. Da kleine Verschiebungen bei ausreichenden Texturen viel genauer berechnet werden können als größere, nimmt durch den Einsatz einer Coarse-to-Fine-Warping-Strategie die Schätzqualität im Allgemeinen deutlich zu. Dazu kann der koronale Effekt des pyramidalen LK-Verfahrens durch den MD-Ansatz reduziert werden und somit kann das LK-Verfahren in seiner pyramidenartigen Implementierung ein robustes und akkurates Verfahren für die Berechnung des optischen Flusses gerade bei großen Verschiebungen darstellen.

Nachteile bei dem angewendeten Verfahren zur Berechnung des Vektorfeldes liegen in der Vektorausweitung und dem Schattenwurf eines Objektes, da dieser auch indirekt zum Objekt gehört. Die Vektorausweitung kann durch weitere Untersuchung durch Kombination<sup>117</sup> des HS-Verfahrens mit dem LK-Verfahren oder die Benutzung eines gewichteten Fensters, z.B. eines Gaußfensters, anstelle des Rechteckfensters für die Berechnung eines Vektors beim LK-Verfahren in Grenzen gehalten werden. Beide Möglichkeiten konnten jedoch auf Grund des begrenzten Zeitrahmens nicht implementiert werden. Hierfür wurde der MD-Ansatz zur Begrenzung der Vektorfeldausweitung angewendet, der für den Zweck hier gute Ergebnisse geliefert hat (vgl. Kapitel 2). Des Weiteren könnte für Schattenwurfeliminierung ein Filter auf der Grundlage des CSC-Algorithmus verwendet werden.

Zur Beschreibung der Farbinformation bewegter Bilddaten wurden mehrere Ähnlichkeitsmaße angewendet, um die Konstellation der Aufgaben des Bewegungsmodells und Beobachtungsmodells zu erfüllen. Hierfür wurden mehrere *Histogrammdistanzfunktionen* (z.B. Minkowski-Abstand, Histogrammschnitt, Jeffrey-Divergenz und  $\chi^2$ -Abstand) in mehreren Farbräumen (z.B. RGB, HSV und YCbCr) untersucht. Um die geeignete Variante zu bestimmen, wurde eine Datenbasis bestehend aus 43 PKW in 5 verschiedenen Positionen mit dem Ziel für jede Variation die Verteilung der Histogrammdistanzen für identische und verschiedene Objekte zu ermitteln, aufgestellt. Dadurch war es auch möglich die identischen und verschiedenen Häufigkeitsverteilungen zu definieren und für weitere Zwecke im Bewegungsmodell und Beobachtungsmodell zu fixieren.

Allerdings beschränken sich die Untersuchungen der Arbeit gegebenenmaßen auf gleichmäßig beleuchtete Szenen. Interessant wäre hier der Einsatz des adaptiven Farbraumes zur optimalen Berücksichtigung der Farbinformation. Zur Verbesserung der Zuverlässigkeit der Farbfehlerbeschreibung kann man hier noch Wichtungsfaktoren einfügen, die die Kombination der Farbfehler der Histogramme in Abhängigkeit ihrer Zuverlässigkeit wichtet.

Die Grundlagen für die Ellipsenbeschreibung sind das ermittelte Verschiebungsvektorfeld und der MD-Ansatz. Allerdings ist die Merkmalsqualität davon stark abhängig. Zur Verbesserung der Merkmalsqualität bei der formbasierten Objektbeschreibung könnte man weiterhin ein verbessertes Objektmodell auf Grundlage der aktiven Konturen (z.B. Aktive Shape Modell, Drahtgittermodell) in Abhängigkeit von der Umgebung in Betracht ziehen.

---

<sup>117</sup> Ähnlich zu der Arbeit von Bruhn, A.; Weickert, J.; Schnörr, C.: Lucas/Kanade Meets Horn/Schunck: Combining Local and Global Optic. First online version published in October, 2004

3. *Im letzten Schritt erfolgt die Korrespondenzanalyse, mit der die gewonnenen Merkmalsdaten verarbeitet und mittels Partikelfiltersystem fusioniert werden.*

Dieser Schritt verarbeitet die Konfliktsituationen, in denen die Korrespondenzanalyse durch die Fusion der bewegungsbasierten und merkmalsbasierten Daten erfolgt. Daher hat es die Erfassung der bewegten Bilddaten in Konfliktsituationen erfordert, zwei Verfahren auf der Basis von Partikelfilter zu fusionieren, um eine Datenzuweisung in Konfliktsituationen (z.B. Verdeckung, Verschmelzung und Zerlegung) durchzuführen. Dies nutzt die Stärken der intensitäts- und merkmalsbasierten Verfahren so aus, dass die Schwächen in beiden Verfahren kompensiert werden können.

Der hier realisierte Einsatz von Partikelfiltern schöpft die Möglichkeiten und Vorteile eines Partikelfilters noch nicht vollständig aus, da beispielsweise der Condensation Algorithmus durch Important-Sampling erweitert werden kann. In dieser Arbeit wurden die Partikelfilter in einer einfachen Modellierung verwendet. Für speziellere Anwendungen sind komplexere Eingaben als die Blobs möglich. Zur Differenzierung von Zielobjekten sind dabei beispielsweise verschiedene Objektmodelle vollstellbar, die dann auf der Assoziationsebene organisiert werden. Dies stellt auch eine Möglichkeit dar, die Trennung verschmolzener Objekte in komplizierten Situationen zu erreichen, die in diesem Verfahren durch die Differenzbildtechnik als ein Objekt angenommen werden. Welche dieser Ansätze auch weiterverfolgt werden, hängt von der gewünschten Anwendung ab. Diese Arbeit stellt dafür eine Grundlage dar.

In den Experimenten wurde gezeigt, dass dieses System zuverlässige Ergebnisse in konfliktfreien Szenen liefert. In verschiedenen Konfliktsituationen wurden gezeigt, dass die Hybridlösung auch bei gehäufte Teil- und Kompletterdeckung von Objekten korrekte Ergebnisse liefert und so eine Verbesserung zu den Resultaten des PLK- und Multi-Matching-Verfahren (vgl. Kapitel 3) erreicht werden konnte. Allerdings ist dieses System nicht in der Lage während der Verschmelzung von verschiedenen Objekten, die genaue Position innerhalb des MB zu bestimmen. Das erhöhte jedoch gleichzeitig die Robustheit der Bestimmung der Trajektorien nach dem Auflösen der Konfliktsituationen.

Der Partikelfilter eignet sich im Besonderen durch seine diskrete Filterverteilung, um die Wahrscheinlichkeitsverteilung des Zustandes, die durch seine Variation diskret ist, wiederzugeben. Des Weiteren ermöglichen die einbezogenen intensitätsbasierten Daten, in ungestörten Situationen eine hohe Genauigkeit in der Zustandsprädiktion im dynamischen Teil des Systems.

Während der Konfliktsituationen erlangt die Datenfusionsstruktur durch den Merge-Split-Ansatz unter der Verwendung von Farbhistogrammen eine gute Stabilität bei Verschmelzungs- und Trennungsszenarien.

Die Effizienz der Datenfusionsstruktur ist aufgrund der häufig durchzuführenden und rechenaufwendigen Histogrammvergleiche nicht optimal, da diese bei langen Sequenzen für viele Knoten und für eine große Anzahl von Partikel durchgeführt werden müssen. Durch eine intelligente Strukturierung der Datenhaltung<sup>118</sup> besteht in diesem Bereich jedoch eine Optimierungsmöglichkeit. Es besteht eine hohe Redundanz in dem Zustandssystem, da viele Partikel gleiche Zustände besitzen und ähnliche Zustände teilweise gleiche Trajektorien enthalten.

Die Komplexität der Implementierung ist hauptsächlich bei der Verarbeitung von Objektzuordnungen bei komplexen Merge- und Split- Konflikten gegeben. Sonst

---

<sup>118</sup> Die Speicherung der Knoteninformation

zeichnet sich das System durch einen geringen Aufwand in der Implementierung aus. Vor allem durch den hierarchischen Aufbau des Zustandsmodells ist die Datenhaltung des Systems prinzipiell ohne großen Aufwand zu gewährleisten. Auch die Partikelfilterstruktur und die Fuzzy-Systeme sind durch die Verwendung von Dreiecksmengen effizient zu integrieren.

### 5.5.2 Besonderheit

- Ein wesentlicher Vorteil der Hybriden-Struktur zur Objektverfolgung gegenüber herkömmlichen Verfahren besteht darin, dass die Stärken der intensitäts- und merkmalsbasierten Verfahren so ausgenutzt werden, dass die Schwächen in beiden Verfahren kompensiert werden können. Dazu sind keine kalibrierte Kamera, kein Straßenmodell und keine Kenntnisse über das Aussehen der zu detektierenden Objekte erforderlich. Darüber hinaus lassen sich Objekte mit signifikanter Relativbewegung zur Kamera zuverlässig detektieren.
- Als weiterer Vorteil der Datenfusion gilt, dass in der Bewegungsbestimmung weitgehend keine Parameter einzustellen sind. Hinzu kommt, dass das Verfahren robust gegenüber Teilverdeckung und Verschmelzung oder Zerfallen von Objekten aufgrund des verwendeten Zustandsmodells ist. Dazu beinhaltet die Struktur ein Fuzzy-System, welches ein effektives Werkzeug bildet, um Expertenwissen algorithmisch umzusetzen. Darüber hinaus erlaubt das System hier eine Entscheidung über die Verbindung von einer Trajektorie zur der möglichen Elterntrajektorie, um mehrere Objekte in einem MB zu identifizieren.
- Weitere entscheidende Vorteile liegen in der robusten Detektion und Analyse von formveränderlichen Objekten. Darüber hinaus liefern die Objektrepräsentationen eine gute Grundlage zur Beschreibung der Objekte zwecks Szenen Klassifikation. Eine Klassifikation der sich bewegenden Objekte (z.B. Personen, PKW, LKW, Motorrad) kann auf diese Grundlagen in Verbindung mit einem Klassifikator implementiert werden (z.B. durch k-Nearest-Neighbor-Klassifikator, das Multi-Layer-Perzeptron und die Support-Vector-Machine). Herausforderungen bestehen einerseits in der Güte der dem Klassifikator angebotenen Lern- oder Testdaten aufgrund von Fehlsegmentierungen sowie in der Auswahl geeigneter Merkmale, die den Merkmalsraum gemäß der gewünschten Klassenkonfiguration<sup>119</sup> in zufrieden stellendem Maße separieren. Als Merkmale können hier Kreisförmigkeit, Rechteckigkeit, Länglichkeit und Hauptachsenwinkel angewendet werden. Die Kreisförmigkeit führte in vielen Situationen zu einer guten Separierbarkeit, jedoch werden Probleme bei einer Unterscheidung von Mensch und Fahrzeug auftreten. Für die Länglichkeit gilt ähnliches. Vorteile brachte der Einsatz der Länglichkeit bei der Unterscheidung von Fahrzeugen in Frontal – und Seitenansicht. Die Rechteckigkeit erwies sich als mittelmäßiges bis gutes Merkmal. Massive Probleme können hier bei der Unterscheidung von Radfahrern und Menschen auftreten. Auch bei der Detektion von Gruppen führt dieses Merkmal nicht zu einer stabilen Erkennung.
- Weitere Vorteile wiederum liegen in der Möglichkeit der Parallelisierung der Berechnung der Merkmale in dem Bewegungs- und Beobachtungsmodell, welche eine Optimierung in Richtung eines echtzeitfähigen Systems ermöglichen können.

---

<sup>119</sup> Die Klassenkonfiguration legt fest, in welche Klassen die Objekte durch den Klassifikator eingeteilt werden sollen. Beispiele: Mensch und Fahrzeug / Mensch, Fahrzeug und Radfahrer

- Das entwickelte Verfahren wurde unter allgemein gehaltenen Annahmen entworfen, sodass für spezielle Einsätze zusätzliche, unterstützende Informationen herangezogen werden. Beispielsweise kann durch die Verwendung entsprechender Umgebungsmodelle, die Szene in Bereiche für Straßen, Fußwege, Überwege, Ampeln etc. geteilt werden, um so eine weitere Grundlage für Annahmen über die Objekte und deren Verhalten zu bieten. Ebenso ist die Verarbeitung spezieller Informationen auf der Ebene der Partikelfilter denkbar.

### 5.5.3 Grenzen

Die Ergebnisse der Zustandsschätzung können an Genauigkeit verlieren bzw. falsch sein, wenn es zu komplizierten Aufnahmesituationen kommt. Es hat sich gezeigt, dass für eine erfolgreiche Objektidentifikation eine minimale Anzahl von separat segmentierten Objekten vorhanden sein muss. In Fußgängerszenen und sehr dichten Verkehr ist es nicht möglich durch die Verwendung der Differenzbildtechnik einzelne Objekte in einem Motion-Blob zu separieren. Das kann dazu führen, dass das Fuzzy-Bewegungsmodell nicht in der Lage ist die Komplexität der notwendigen Hypothesen auf der Grundlage der Histogramm Daten zu erzeugen. Da das System zur Auflösung der Konfliktsituationen dem Split-Operator vorwiegend die Information der Farbhistogramme nutzt, ist die Genauigkeit des Bewegungsmodells abhängig von der Farbbarkeit der Objekte, das bedeutet, dass je ähnlicher die Objekte in ihrer Farbbarkeit sind desto ungenauer wird die Objektidentifikation mit dem Bewegungsmodell.

### 5.5.4 Zusammenfassung und Ausblick

Die umfangreiche Zusammensetzung des entwickelten Verfahrens aus verschiedenen Methoden enthält Stellen, die in optimierter Weise zur Verbesserung des Verfahrens beitragen können. Durch die Verbesserung der Segmentierungsergebnisse können schon zu Beginn der Verarbeitungskette Konfliktsituationen vermieden werden. So zum Beispiel mit der Integration einer Schatteneliminierung.

Für die Bewegungsschätzung wurde eine Fuzzy-Regelung verwendet, deren Fuzzy-Sets relativ grob sind. Auf eine genauere Untersuchung der Fuzzy-Sets ist nicht eingegangen worden, weil gezeigt werden sollte, dass durch die Verwendung des Partikelfilters als Zustandsschätzer Fehler in der Bewegungsprädiktion ausgeglichen werden können. So wird nach dem Prinzip von Occam's Razor<sup>120</sup> auf eine Spezialisierung des Fuzzy-Systems auf die Testszenen verzichtet, um eine generelle Anwendung zu bewahren. Trotzdem ist es möglich durch eine Verfeinerung dieser Fuzzy-Sets auf bestimmte Situationen mit angepassten Fuzzy-Sets zu reagieren, so besteht beispielsweise die Möglichkeit unterschiedliche Fuzzy-Sets für unterschiedliche Objekttypen zu verwenden. Auch könnten z.B. Neuro-Fuzzy-Methoden eine Verfeinerung der Fuzzy-Sets und der Regelbasis bewirken und somit die Bewegungsprädiktion verbessern.

Des Weiteren ist es möglich dem Beobachtungsmodell unterstützende Information einzubinden. Beispielsweise können unter Verwendung entsprechender Umgebungsmodelle, bestimmte Bereiche in den Szenen z.B. einzelne Straßen oder Ampeln markiert werden, in denen bestimmte Aktionen verschmelzen, nicht erlaubt sind.

---

<sup>120</sup> Occam's Razor bevorzugt die einfachste Hypothese, die konsistent mit allen Beobachtungen ist. Das heißt, wenn zwei unterschiedlich komplexe Beschreibungen von Modellen mit gleicher Genauigkeit auf den gegebenen Daten existieren, bevorzugt die einfachere. Diese Heuristik hat in vielen Anwendungsbeispielen zu guten Ergebnissen geführt, weil effektiv ein Overfitting des Klassifikators an die Trainingsbeispiele verhindert werden kann.

## **5. DATENFUSION AUF DER BASIS VON PARTIKELFILTERN**

---

Grundsätzlich gibt es eine Vielzahl an Optimierungsmöglichkeiten, die von der gewünschten Anwendung abhängen. Diese Arbeit stellt eine Grundlage dar, die zeigt wie in zwei Abstraktionsebenen einerseits intensitäts- und merkmalsbasierte Daten vereint werden können und andererseits die Ungenauigkeiten eines Expertensystems mit Hilfe eines stochastischen Verfahrens kompensiert werden können.

### 6 Zusammenfassung der Gesamtarbeit

Die Analyse von bewegten Bildern hat neben den aktuellen Forschungen im Bereich der intelligenten Überwachungstechnik, Umfelderkennung und Fahrerassistenzsysteme einen wesentlichen Einfluss auf Mensch-Maschine-Interaktionssysteme (MMI)<sup>121</sup>. Unabhängig von der Art und Komplexität der Objekte handelt es sich bei MMI prinzipiell um eine Überlagerung von mehreren dynamischen Verarbeitungsstufen, die ohne die Analyse von bewegten Bildern nicht zu bewerkstelligen sind. Die erste Stufe beschreibt die automatische Lokalisierung von Regionen des potentiellen Nutzers und seines Umfeldes. Die zweite Stufe beschreibt die lokale und globale Bewegungserfassung der Objekte. Die dritte Verarbeitungsstufe steht für die Erfassung und Zuordnung der Regionen während der Interaktionen und bezieht sich somit auf die Dynamik innerhalb der erfassten Regionen. Da diese Verarbeitungsstufen nicht linear unabhängig sind, sondern Korrelationen zwischen den Stufen zu erwarten sind, werden aus den Bildsequenzen Merkmale extrahiert, die diesen Stufen sowie deren Korrelation Rechnung tragen. Hierbei stellen die Schritte der automatischen Objekterfassung, Zustandsbeschreibung und der Datenassoziation einen erheblichen Anspruch an robuste Methoden zur Analyse von bewegten Bildern.

Da derzeit keine einzelne Methode eine Informationsdichte liefert, wie sie für die genannten Applikationen benötigt werden, zeichnet sich der Trend einer multisensoriellen Umfelderkennung mit einer anschließenden Fusion der Daten von mehreren Kamerasystemen und/oder zumeist heterogenen Sensordaten ab.

Die Herausforderung an die Objekterfassung und Zustandsbeschreibung liegt darin, dass ein komplexes Umgebungs- und Objektmodell erforderlich ist, jedoch mit keinem verfügbaren Kamerasystem die Möglichkeit besteht, alle Zustandsgrößen bei nicht kooperativen Messsituationen gleichzeitig zu beobachten. Die Beobachtbarkeit ist darüber hinaus abhängig vom Umfeld und der Perspektive, aus der ein Objekt betrachtet wird. Die Kombination von mehreren Kameras und/oder heterogenen Sensoren schafft die Möglichkeit, ein komplexes Umgebungsmodell sowie ein Objektmodell zu erstellen, führt allerdings zu mehr Komplexitäten, Rechen- und Kostenaufwand. Somit stellt die Verwendung von herkömmlichen Systemen zur Objektverfolgung die Effizienz und die Verwertbarkeit der Ergebnisse in Frage. Die Auswertung solcher fehlerhafter Messergebnisse führt entweder zu einer Verfälschung nachfolgender Klassifikationsergebnisse oder bedingt weiteren Rechenaufwand, wenn mit zusätzlichen Kriterien versucht wird, solche Messungen nachträglich zu eliminieren.

Um sich bewegende Objekte in einer monokularen Farbbildsequenz automatisch erfassen und verfolgen zu können, wurden vier wesentliche Forschungsaspekte im Rahmen der vorliegenden Arbeit thematisiert, die insbesondere durch folgende Innovationen einen Beitrag zur Reduzierung der Komplexität und somit zur Lösung des Korrespondenzproblems in komplexen Situationen geleistet haben:

- Zum einen wurde ein systematischer Rahmen zur Entwicklung genauer intensitätsbasierter Ansätze auf der Grundlage der optischen Flussverfahren und Korrelationsverfahren vorgestellt. Dieser erlaubt eine Verbesserung der zurzeit bekannten Schätzver-

---

<sup>121</sup> Der Einsatz dafür kann zum Warn- und Sicherheitssystem in der Überwachung oder Verhaltensanalyse im Bereich der Ereigniserkennung „*Event Recognition*“ oder aktiv zur Vermeidung vom Misserfolg bei den Interaktionen sein.

fahren, die in der Literatur zu finden sind. Gegenstand dieser Arbeiten waren die intensitätsbasierten optischen Flussverfahren und ihre Verwendung zur Selektion und Verfolgung mehrerer bewegter Objekte. Die hier vorgeschlagenen Lösungsansätze wurden im Kapitel 5 für die Zustandsbeschreibung eines Knotens im Zustandsmodell optimal eingesetzt. In einem weiteren Schritt wurde eine adaptive Systemstruktur auf der Grundlage der Korrelationen zwischen Bildausschnitten entwickelt, die die Farbinformation optimal zur Verbesserung der Störsicherheit ausnutzt. Die funktionelle Beschreibung dieser Systemstruktur zur Bewegungsanalyse legt zunächst weder ein Bewegungsmodell noch den verwendeten Farbraum fest. Dabei wird die Bewegungsbestimmung von Objekten nicht wie bisher überwiegend in der Schwarz-Weißdomäne sondern in der Farbdomäne durchgeführt.

- Zum zweiten wurde ein neues Matching-Paradigma zur merkmalsbasierten Korrespondenzermittlung in Farbbildsequenzen vorgeschlagen, das sich vom klassischen Matching durch seine hierarchische Struktur und Adaptivität unterscheidet. Dieses Konzept segmentiert, verfolgt und erkennt nicht nur fest gewählte Regionen, sondern auch beliebige *veränderliche* Objekte in Bildfolgen automatisch. Zusätzlich ermöglicht es die Berechnung in Echtzeitnähe und macht erstmals dieses äußerst robuste merkmalsbasierte Matching für die Bewegungsbestimmung und Verhaltensbeschreibung praktisch einsetzbar. Die Echtzeitfähigkeit der Berechnung durch dieses Matching-Paradigm kann sich aus der inhärenten möglichen Parallelität des angewandten Teilverfahrens und seiner Schnelligkeit ergeben.
- Zum dritten wurde ein Partikelfiltersystem in Verbindung mit einer Datenassoziation für die Multi-Objektverfolgung entwickelt, dessen Verarbeitungsstufen die ursprünglichen Schritte eines Partikelfilters beinhalten, die in dieser Arbeit aber so modifiziert werden. Dadurch wird eine zuverlässige Bestimmung von Bewegungstrajektorien in Konfliktsituationen ermöglicht. Die Modifikationen beziehen sich zuerst auf die Prädiktion, die durch ein erweitertes *adaptives Bewegungsmodell* erreicht wird. Zusätzlich wurde der ursprüngliche Ablauf des Partikelfilters um einen 3-stufigen Assoziationsschritt mit Datenausschluss, Datenzuweisung und Datenverwaltung ergänzt, um die Problematik der Fluktuationserkennung und den Umgang mit Objektinteraktionen zu lösen.
- Zuletzt wurde ein neuer Lösungsweg auf Grundlage der Datenfusion entwickelt, der eine Zustandschätzung und robuste Datenzuweisung von Objekten in Konfliktsituationen ermöglicht. Die Struktur für die Datenfusion nutzt die Stärken der intensitäts- und merkmalsbasierten Verfahren so aus, dass die Schwächen in beiden Verfahren kompensiert werden können. Die Grundlagen dafür sind die Bayes-Schätzung, die Fuzzy-Logik und die vorgestellten Bewegungsanalyseverfahren.

Ausgehend von diesen Aspekten wurde die vorliegende Arbeit in sechs Kapitel gegliedert, die die Basis für weitere Forschungsarbeiten im Bereich der Analyse von bewegten Farbbildern unter dem Einfluss von nicht kooperativen Messsituationen dienen und wie folgt zusammenzufassen sind:

Im Anschluss an die Einleitung und die Einordnung der eigenen Arbeiten im *ersten Kapitel* wurden im *zweiten Kapitel* zunächst die bekannten intensitätsbasierten Ansätze hinsichtlich ihrer Leistungsfähigkeit, des Rechenaufwandes und der Robustheit untersucht und bewertet. Experimente zur Evaluation der Qualität belegen die Eignung der Verfahren zur Bestimmung der Verschiebungsvektoren eines Objektes. Die notwendigen Optimalitätsbedingungen wurden für zwei einfache Variationsmethoden

präsentiert. Sie bilden eine mathematische Grundlage und Rechtfertigung für den zweiten Abschnitt der Arbeit. Dazu wurde in einer Reihe von Experimenten der Einfluss von Restriktionen auf den optischen Fluss untersucht. Anschließend wurden eine Zusammenfassung und ein Ausblick auf Erweiterungen dieser Arbeit gegeben.

Im dem zweiten Teil des zweiten Kapitels wurde eine adaptive Systemstruktur zur Bewegungsanalyse unter der Berücksichtigung von nichtkooperativen Messsituationen auf der Grundlage intensitätsbasierter Korrelationsverfahren entwickelt. Dabei wurden sowohl Methoden der Bildverarbeitung als auch der Einfluss von Farbinformationen betrachtet. Daraus resultierten zwei Verarbeitungsschritte: die Aufbereitung der Farbinformation zur Objekterkennung und die eigentliche Bewegungsanalyse in Bildsequenzen. Ausgehend von den grundlegenden Überlegungen wurde ein Farbraum in Abhängigkeit von den Bildinhalten definiert. Somit berechnet man hier eine Transformationsmatrix, deren Anwendung dazu führt, dass die Farbkanäle weitgehend unkorreliert sind, d.h. die höchstmögliche Informationsmenge transportiert wird. Dieser adaptive Farbraum gewährleistet in Verbindung mit einem Korrelationsverfahren die Bestimmung der Bewegungsgrößen eines Objektes auch in nichtkooperativen Messsituationen.

Für die Realisierung des bildspezifischen Erkennungssystems wurde der Einsatz von überwacht angelegten Feed-Forward-Netzen vorgeschlagen, die die Aufgabe haben, die Wichtung der kanalspezifischen Ähnlichkeitsfunktionen für das Multi-Matching anzupassen. Es wurde die Bedeutung der einzelnen Komponenten des vorgeschlagenen Systems und deren Zusammenwirken kurz beschrieben. Die Eignung der vorgeschlagenen Systemstruktur zur robusten Bewegungsbestimmung und Objektverfolgung wurde anhand realer Videosequenzen nachgewiesen, die häufig vorkommende bildspezifische Störungen enthalten. Es wurde gezeigt, dass mittels der vorgeschlagenen Systemstruktur eine höhere Störsicherheit bei der Bewegungsanalyse beim Auftreten dieser Störeffekte erreicht wird. Es werden zuverlässige Verschiebungsvektoren von Objekten unter dem Einfluss von nicht-kooperativen Messsituationen (z.B. Beleuchtungsänderungen, Schattenwurf und Teilverdeckung) mit der entwickelten adaptiven Systemstruktur ermittelt, die mit anderen derzeit verfügbaren intensitätsbasierten Matchingsverfahren nicht auswertbar sind.

In *Kapitel 3* wurde ein neues Matching-Paradigma zur Verfolgung multipler Objekte vorgeschlagen, welches aus dem hierarchischen Multi-Matching abgeleitet wurde. Für die Entwicklung wurde von einem statischen Aufnahmesystem, das Farbbildsequenzen liefert, ausgegangen. Die zu analysierenden Szenen enthalten Objekte unterschiedlicher Art und unterschiedlichen Verhaltens, die sich größtenteils unabhängig von einander bewegen.

Um eine möglichst exakte Segmentierung bewegter Objekte zu erreichen, wurde eine modifizierte Differenzbildtechnik in Verbindung mit einem Nachverarbeitungsalgorithmus zur Unterdrückung von verbleibenden Fehlern entwickelt, bei dem jeweils zwei Einzeldifferenzbilder kombiniert werden. Im Zusammenhang mit der Nachverarbeitung konnte eine genaue Segmentierung für die sich bewegenden Regionen erzielt werden. Da durch die Bewegungssegmentierung auch bewegte Schattenregionen erfasst werden, welche die Bewegungsanalyse des gesamten Bildes stören können, wurde weiterhin ein Algorithmus zur nachträglichen Segmentierung und Entfernung von Schatten auf der Basis von Farbvalenzen entworfen. Die als Motion-Blobs (MBs) vorgestellten Regionen, welche durch den Vorgang der Bewegungssegmentierung ermittelt wurden,

können recht groß sein und die Abbildungen mehrerer Objekte, wie z.B. beim Verschmelzen oder der Teilverdeckung von Objekten im dichten Verkehr, beinhalten. Um diesem Sachverhalt gerecht zu werden, wurden MBs-Regionen in kleinere Segmente zerlegt, was durch den Farbsegmentierungsalgorithmus realisiert wurde. Aus der Menge der ermittelten Farbsegmente und der übergeordneten MBs-Regionen gehen zwei Merkmalsebenen hervor, welche genutzt werden, um in der Kombination das Korrespondenzproblem zu lösen.

Anders als in bisherigen merkmalsorientierten Verfahren konnte durch die Betrachtung zweier Merkmalsebenen die hohe Anzahl an Bildmerkmalen sehr einfach, schnell und exakt reduziert werden. Die Schritte, die zur Umsetzung dieses Matching-Paradigmas führen, wurden begründet und detailliert beschrieben. Im Anschluss wurde das spezielle merkmalsorientierte Prinzip der Zuordnung von Bildregionen spezifiziert. Erweiterungen und Optimierungsansätze auf der Grundlage von Simulated-Annealing-Algorithmen zur Verbesserung der Matchinggüte durch die Gewichtsoptimierung wurden entwickelt und getestet. Die durch das vorgeschlagene Multi-Matching ermittelten Korrespondenzen führen zu Bewegungstrajektorien, welche Auskunft über die Bewegung in längeren Bildsequenzen geben können und somit das wichtigste Ergebnis des Zuordnungsprozesses darstellen.

Das Matching wurde in separaten Durchgängen realisiert und basiert jeweils auf einer Reihe von Ähnlichkeitsmaßen. Die durch den Matching-Prozess ermittelten Korrespondenzen zwischen den Segmenten führen zu Verschiebungsvektoren, welche prinzipiell genutzt werden können, um Aussagen über Objektbewegungen und Objektverhalten zu treffen, die als Ausgangspunkt für die erweiterte Meta-Objekt-Korrelation genutzt werden können. Es wurde anhand realer, gestörter Beispiele demonstriert, dass die Bewegung einzelner Objekte durch Häufungen von Verschiebungsvektoren im räumlich-zeitlichen Kontext repräsentiert wird. Bei der Berechnung der Verschiebungsvektoren konnte eine hohe Robustheit gegenüber in realen Situationen häufig kombiniert auftretenden Problemen, wie Beleuchtungsschwankungen, Teilverdeckungen, Schattenwurf, Objektdeformation (durch perspektivisch bedingte Größenveränderung) und Rotation erreicht werden.

In den Untersuchungen realer Bildfolgen hat sich gezeigt, dass das entwickelte merkmalsorientierte Matching-Konzept eine echtzeitfähige Bewegungsanalyse ermöglichen kann. Dies stellt einen Fortschritt gegenüber aufwändigen intensitätsorientierten Ansätzen dar. Weiterhin liegt die Qualität der in den Experimenten erzielten Ergebnissen über den Möglichkeiten der störanfälligen, intensitätsorientierten Verfahren, wie zum Beispiel „den Optischen Fluss“. Das Verfahren wird jedoch durch die vorausgesetzte Verwendung einer statischen Kamera in seinen Einsatzgebieten eingeschränkt. Eine weitere Aufgabe zur Verbesserung des Ergebnisses der Bewegungssegmentierung in Außenaufnahmen liegt in der automatischen Erfassung und Entfernung von Störregionen, die sich etwa durch Bewegungen von Bäumen bzw. Blättern, o.Ä. ergeben.

In den Untersuchungen hat sich gezeigt, dass solche Regionen einen sehr hohen Texturgehalt sowie ein unzuverlässiges Verschiebungsvektorfeld aufweisen. Auf dieser Grundlage können die Regionen klassifiziert und entfernt werden. In ähnlicher Weise können auch andere unvorhergesehene Bewegungsquellen bzw. Störungsquellen, die in der jeweiligen Anwendung auftreten, entfernt werden. Es ist allerdings klar, dass derzeit kein System alle möglichen Ursachen, die in der Realität zu Fehlern bei der Bewegungserfassung führen können, beherrscht. Die in dieser Arbeit geschaffene Technik

zur Bewegungserfassung und merkmalsorientierten Bestimmung der entsprechenden Bewegungsparameter, d.h. der Verschiebungsvektoren, dürfte jedoch eine gute Grundlage in vielen Situationen darstellen. Die Ergebnisse, die in experimentellen Untersuchungen erreicht wurden, geben dafür genug Grund zur Annahme.

In *Kapitel 4* wurde ein statistisch-basiertes Verfahren entwickelt, das auf der Basis von Partikelfiltern und Datenassoziationen die Verfolgung multipler Objekte sowie die stabile Ermittlung ihrer Trajektorien auch in Konfliktszenen ermöglicht. Die Bewegungsanalyse und die Merkmalsextraktion sind die Basis für die Verwendung von Partikelfiltern. Zur Formulierung und Repräsentation der Objekte wurden Bild- und Farbparameter definiert. Auf Bildebene sind dies lediglich ortsrelevante Daten wie Position oder Richtung. Zusätzlich ist eine Farbcharakteristik für die Objekte entwickelt worden. Die Bild- und Farbparameter dienen der Formulierung des Bewegungs- und des Beobachtungsmodells.

Zur Modellierung der zu schätzenden Objektzustände in unterschiedlichen Situationen wurden drei Modelle formuliert und angewendet: Standardmodell, Erfahrungsmodell, Standmodell. Dadurch entstand ein adaptives Bewegungsmodell, das sowohl dynamisch auf die Bewegungsänderungen des jeweiligen Objekts reagiert als auch die Partikelevolution den äußeren Umständen anpasst. Die Entscheidung über den Einsatz der Bewegungsmodelle erfolgt in der Datenassoziation. Für die Multi-Objektverfolgung ist ein Partikelfiltersystem aufgestellt worden, das aus unabhängigen Partikelfiltern besteht, die jeweils ein Objekt repräsentieren. Durch die Unabhängigkeit wird die individuelle Bewegungsschätzung mit dem adaptiven Bewegungsmodell möglich. Zudem wird durch die Zusammensetzung aus Partikelfiltern der Zustandsraum und somit Variationsmöglichkeiten auch die Anzahl der notwendigen Partikel, verringert. Die Verfolgung multipler Objekte und deren Trajektorien setzt räumliche und zeitliche Identifizierung voraus. Die Zuordnungen erfolgen im Assoziationsschritt in drei Stufen: Datenausschluss, Datenzuweisung, Datenverwaltung. Durch Datenausschluss und Datenzuweisung wird die Szene zerlegt, werden die Konfliktsituationen zwischen Objekten und Messungen soweit wie möglich aufgelöst und durch die JPDA Empfehlungen für die Objekt-Messung-Zuordnungen gegeben. Die Verwendung von Statusvariablen gibt Aufschluss über die Bewegungs- bzw. Konflikthistorie der Objekte und unterstützt im Verwaltungsschritt die Wahl des jeweiligen Bewegungsmodells und die endgültige Objekt-Messung-Zuordnung. Damit ist die Grundlage für den Aktualisierungsschritt des Partikelfiltersystems und somit die Schätzung des nächsten Zeitschritts geschaffen.

Die Organisation der einzelnen Partikelfilter und der Objekte mit ihren Trajektorien wird mit einer verbundwahrscheinlichkeitbasierten Methode bewerkstelligt und in einem Verwaltungsschritt assoziiert und organisiert. Die Partikelevolution wird in unterschiedlichen Situationen mit unterschiedlichen Bewegungsmodellen steuert. Die Zusammensetzung der Methode wurde mit Untersuchungen innerhalb des Kapitels untermauert. Dazu wurden in diesem Kapitel Untersuchungen mittels realer Bildsequenzen beschrieben. Zum einen wurde damit die Verwendung bestimmter Techniken der im Kapitel 3 entworfenen Schritte begründet. Zum anderen wurde die Funktionsfähigkeit der entwickelten Methode zur Multi-Objektverfolgung untersucht. Anhand der Ergebnisse werden die erreichten Ziele, aber auch die Grenzen der vorgeschlagenen Methode erläutert. Abschließend erfolgt eine Zusammenfassung der Erkenntnisse dieser Arbeit. Letztendlich werden Lösungsansätze zur Verfeinerung der Methode, zur

## 6. ZUSAMMENFASSUNG DER GESAMTARBEIT

---

Behebung der zuvor erläuterten Probleme, aber auch für weiterführende Arbeiten gegeben.

In *Kapitel 5* wurde basierend auf den Erkenntnissen aus Kapitel 2, 3 und 4 ein neuer Lösungsweg vorgeschlagen und als hybrides System dargestellt, das durch Fusion der intensitätsbasierten und merkmalsbasierten Ansätzen mittels Partikelfilter, die Vorteile beider Verfahren nutzt und dadurch ihre Nachteile kompensiert. Dazu wurde ein Partikelfiltersystem auf der Grundlage des Condensation-Algorithmus entwickelt, das durch Abstraktion der Probleme der auftretenden Konfliktsituation ein Zustandsmodell einführt, in dem diese Probleme als Rauschen bzw. Messungenauigkeit modelliert werden. Somit konnte eine Bayes-Filterung auf den Prozess der Objektidentifikation, in einigen Fällen auch als Datenassoziation angewandt werden.

Durch die Verwendung eines Fuzzy-Regelsystems in der Zustandsprädiktion wurde zusätzliches Expertenwissen in das System eingebracht, so dass die Variationen Zustände, die sich aus den intensitätsbasierten Daten und den Farbdaten ergeben, über ein Expertensystem reguliert werden können. In den Experimenten wurde gezeigt, dass dieses System zuverlässige Ergebnisse in konfliktfreien Szenen liefert. Es konnte in verschiedenen Konfliktsituationen gezeigt werden, dass die Hybridlösung auch bei gehäufte Teil- und Kompletterdeckung von Objekten weitgehend korrekte Ergebnisse liefert und so eine Verbesserung gegenüber den Resultaten des PLK-Algorithmus und des merkmalsbasierten Verfahrens erreicht werden konnte. Allerdings ist dieses System nicht in der Lage, während der Verschmelzung von verschiedenen Objekten, die genaue Position innerhalb des MBs zu bestimmen. Dafür wurde ein Konsistenz- und Glattheitskriterium definiert. Das erhöhte gleichzeitig die Robustheit der Bestimmung der Trajektorien nach dem Auflösen der Konfliktsituationen.

Anschließend wurden Untersuchungen von realen Bildsequenzen beschrieben, die Konfliktsituationen beinhalten. Damit wurde die Verwendung bestimmter Techniken in den in Kapitel 5 vorgeschlagenen Schritten begründet. Anhand der Ergebnisse werden die erreichten Ziele, aber auch die Grenzen der Methode erläutert. Allerdings enthält die umfangreiche Zusammensetzung des entwickelten Verfahrens aus verschiedenen Methoden Stellen, die bei der Optimierung zur signifikanten Verbesserung des Verfahrens beitragen können. Durch die Verbesserung der Segmentierungsergebnisse können schon zu Beginn der Verarbeitungskette Konfliktsituationen vermieden werden. Für die Bewegungsschätzung wurde eine Fuzzy-Regelung verwendet, deren Fuzzy-Sets relativ grob sind. Auf eine genauere Untersuchung der Fuzzy-Sets ist nicht eingegangen worden, weil gezeigt werden sollte, dass durch die Verwendung des Partikelfilters als Zustandsschätzer Fehler in der Bewegungsprädiktion ausgeglichen werden können. So wird nach dem Prinzip von Occam Razor auf eine Spezialisierung des Fuzzy-Systems auf die Testszene verzichtet, um eine generelle Anwendung zu bewahren. Trotzdem ist es möglich, durch eine Verfeinerung dieser Fuzzy-Sets auf bestimmte Situationen mit angepassten Fuzzy-Sets zu reagieren. So ist es z.B. möglich unterschiedliche Fuzzy-Sets für unterschiedliche Objekttypen zu verwenden. Auch könnten z.B. Neuro-Fuzzy-Methoden eine Verfeinerung der Fuzzy-Sets und der Regelbasis bewirken und somit die Bewegungsprädiktion verbessern.

Des Weiteren ist es möglich dem Beobachtungsmodell unterstützende Informationen einzubinden. Beispielsweise können unter Verwendung entsprechender Umgebungsmodelle, Bereiche in den Szenen, z.B. Straßen oder Ampeln markiert werden, in denen Aktionen wie „verschmelzen“, nicht erlaubt sind. Grundsätzlich gibt es eine Vielzahl an

## **6. ZUSAMMENFASSUNG DER GESAMTARBEIT**

---

Optimierungsmöglichkeiten, die von der gewünschten Anwendung abhängen. Diese Arbeit stellt eine Grundlage dar, die zeigt, wie in zwei Abstraktionsebenen intensitäts- und merkmalsbasierte Daten vereint werden können und die Ungenauigkeiten eines Expertensystems mit Hilfe eines stochastischen Verfahrens kompensiert werden können. Den Abschluss dieser Arbeit bildet das sechste Kapitel, welches zugleich eine Zusammenfassung der Erkenntnisse der vorliegenden Arbeit bildet und den Ausblick dargestellt.



## 7. Abkürzungsverzeichnis

<b>Kapitel 2</b>	
<b>CCD</b>	<b>Charged Coupled Device</b>
$u, v$	Die Flussgeschwindigkeit in $x$ - und $y$ -Richtung
<b>G</b>	Die Gradientenmatrix
<b>b</b>	Der Bildfehlervektor
SAE	Die Summen aller relativen Winkelfehler (Sum of all relative Angular Errors)
SRE	Normalisierte Summen aller relativen Betragsfehler (Sum of all Relativ Errors)
$v_c$ und $v_e$	Das Vektorfeld der Grundwahrheit $v_c$ und $v_e$ das geschätzte Vektorfeld
$\alpha$	Der Wichtungsfaktor
$I_x$ und $I_y$	Die Bildgradienten
$\lambda_1$ und $\lambda_2$	Die Eigenwerte der Gradientenmatrix <b>G</b> für ein Fenster der Größe $m \times m$
$\lambda_S$	Der Grenzwert, der auch als Schwellwert zu betrachten ist
PLK	Pyramidales Lucas-Kanade-Verfahren
APLK	Adaptives pyramidales Lucas-Kanade-Verfahren
MD	Die modifizierte Differenzbildtechnik
MAD	Mittlere Absolute Differenz, Ähnlichkeitsmaß
$MAD^{K_1K_2K_3}$	Mittlere absolute Differenz im $K_1K_2K_3$ -Farbraum; Gesamtkriterium
MSE	Mean Square Error, Ähnlichkeitsmaß
nKKF	Normierte Kreuzkorrelationsfunktion, Ähnlichkeitsmaß
RGB	Roter, Grüner und Blauer Kanal des RGB-Farbraumes
$K_1$	Erste (achromatische) Komponente des adaptiven Farbraumes
$K_2$	Zweite (chromatische) Komponente des adaptiven Farbraumes
$K_3$	Dritte (chromatische) Komponente des adaptiven Farbraumes
<b>q</b>	Richtungsvektor
$q_{ij}$	Faktorladungen/ Eigenvektorkomponenten
$Q_k^{K_i}$	Ähnlichkeitsmaße der Komponenten $K_i$ des adaptiven Farbraumes zum Zeitpunkt $k$
$t$	Die kontinuierliche Zeitkoordinate
$k$	Die diskrete Zeitkoordinate
<b>v</b>	Der Verschiebungsvektor (bzw. Geschwindigkeit)
<b>VV</b>	Der Verschiebungsvektor (Abkürzung)
VVF	Das Verschiebungsvektorfeld
$x, y$	Die Pixelkoordinaten im Bild/ Kantenlänge des Suchbereichs
MLP	Multi-Layer-Perzeptrons (Neuronales Erkennungssystem)
<b>Kapitel 3</b>	
RHS	Recursive-Histogram-Splitting (globales Farbsegmentierungsverfahren)
S&M	Split-and-Merge (lokales Segmentierungsverfahren)
CSC	Color-Structure-Code-Algorithmus (Regionenwachstumsverfahren)
DI	Die Differenzbilder
$V_a(x, y)$	Die Valenzmaskenwerte
MB	Resultierender Motion-Blob aus dem MD-Ansatz
$C_i^k$	Die Matrix $C_i^k$ entspricht einem MB und beinhaltet ein Farbvektorfeld
J	Die Farbsegmentmenge J, die die Eingangsbildmerkmale der zweiten Ebene repräsentieren

## 7. ABKÜRZUNGSVERZEICHNIS

---

$Ko$ und $G$	Die Kompaktheit und der Gültigkeitsstatus
SL	Die relative Position innerhalb einer MB-Menge (CSC-Multi-Matching)
SI	Die minimale Inter-Frame-Distanz (CSC-Multi-Matching)
SC	Die Ähnlichkeit des Farbwertes (CSC-Multi-Matching)
SA	Die Ähnlichkeit der Größe korrelierter CSC-Segmente (CSC-Multi-Matching)
SM	Das Gesamtähnlichkeitsmaß (CSC-Multi-Matching)
$\mathbf{F}_v^k$	Normierte RGB-Farbvektoren ( <b>rot</b> , <b>grün</b> und <b>blau</b> / Farbwertanteile)
$\mathbf{w}$	Die Gewichtungsfaktoren
MO	Das Meta-Objekt
MOK	Die Meta-Objekt-Korrelation
TAA	Der Threshold-Accepting-Algorithmus
SAA	Der Simulated Annealing Algorithmus
$E(\mathbf{w})$	Die Gütefunktion
$fSRE$	Der Restfehler
$Pa_r$	Die Anzahl der falsch zugeordneten Segmente/ Paarungen
$Pa_f$	Die Anzahl der richtig zugeordneten Segmente/ Paarungen
NeS	Die Anzahl der nicht erkannten Segmente/ Paarungen
$L$	Das Schwerpunktzentrum einer CSC-Patch-Menge
$L^*$	Der Mittelpunkt der Gesamtausdehnung (das räumliche Zentrum)

### Kapitel 4

$\mathbf{z}_k$	Der Zustandsvektor (Objektzustände z.B. Position) zum Zeitpunkt $k$
$\mathbf{m}_k$	Der Messvektor (Messungen oder die Beobachtungen)
$p(\mathbf{z}_k   \mathbf{m}_k)$	Die A-posteriori-Wahrscheinlichkeit
$p(\mathbf{m}_k   \mathbf{z}_k)$	Die resultierende Likelihood-Funktion (Ähnlichkeitsfunktion)
$p(\mathbf{z}_k   \mathbf{m}_{k-1})$	Die A-priori-Wahrscheinlichkeit
$p(\mathbf{z}_{k-1}   \mathbf{m}_{k-1})$	Die alte A-posteriori-Wahrscheinlichkeit
$p(\mathbf{z}_k   \mathbf{z}_{k-1})$	Die Auftrittswahrscheinlichkeiten des Bewegungsmodells
KF	Der Kalman-Filter
$B$	Die Steuermatrix
MHT	Der Multi-Hypothesen-Tracker
EKF	Der erweiterte Kalman-Filter
$S_k$ und $\pi_k$	Die Partikelmenge und zugehörige Gewichte
$\mathbf{z}_{b,k}$ und $\mathbf{z}_{co,k}$	Die Bildparameter und die Farbparameter
$(p_x, p_y)$	Die Position eines Motion-Blobs
$(d_x, d_y)$	Die Richtung und Geschwindigkeit mit dem Bewegungsvektor
UV	Ein Unterraum von dem HUV-Farbraum
MM <sub>N</sub> :	Das Standardbewegungsmodell
MM <sub>E</sub>	Das Erfahrungsbewegungsmodell
MM <sub>S</sub>	Das Standmodell
MM <sub>A</sub>	Das adaptive Bewegungsmodell
$SM_{Par}$	Das Ähnlichkeitsmaß für die Partikelgewichtung
$SM_{Obj}$	Das Ähnlichkeitsmaß für die Objektvergleiche
$SM_{pos}$	Das Ähnlichkeitsmaß für die Position
$SM_{spe}$	Das Ähnlichkeitsmaß für die Geschwindigkeit
$SM_{dir}$	Das Ähnlichkeitsmaß für die Bewegungsrichtung
PDA	Probabilistic Data Association

## 7. ABKÜRZUNGSVERZEICHNIS

---

JPDA	Joint Probabilistic Data Association
MHT	Die Multi-Hypothesen-Technik
$K=(k_{ij})$	Die Kandidatenmatrix der Größe $O \times M$
$\theta_{ij}$	Die Zuordnungspaare zwischen Prädiktion $z_{(i)}$ und Messung $m_{(j)}$
EZ	Die eindeutige Zuordnung zwischen einem Objekt und einer Messung
EO	Die Hindeutung auf einzelne Objekte
EM und MZ	Die einzelne Messung und die mehrdeutige Zuordnung
$\mathbf{H}_k$	Die Matrix des Systemmodells
HUV	Die Helligkeit, UV-Farbdifferenzsignale (Gegenfarben); Farbraum
$H_n$	Die N-Hypothesen jeder Kandidatenmenge
$p(H_n)$	Die Verbundwahrscheinlichkeit für alle Hypothesen $H_n$
$L(\theta_{ij})$	Einzelwahrscheinlichkeiten
$p(\theta_{ij})$	Die Assoziationswahrscheinlichkeiten der Zuordnungen $\theta_{ij}$
$A$	Die Assoziationsmatrix

### Kapitel 5

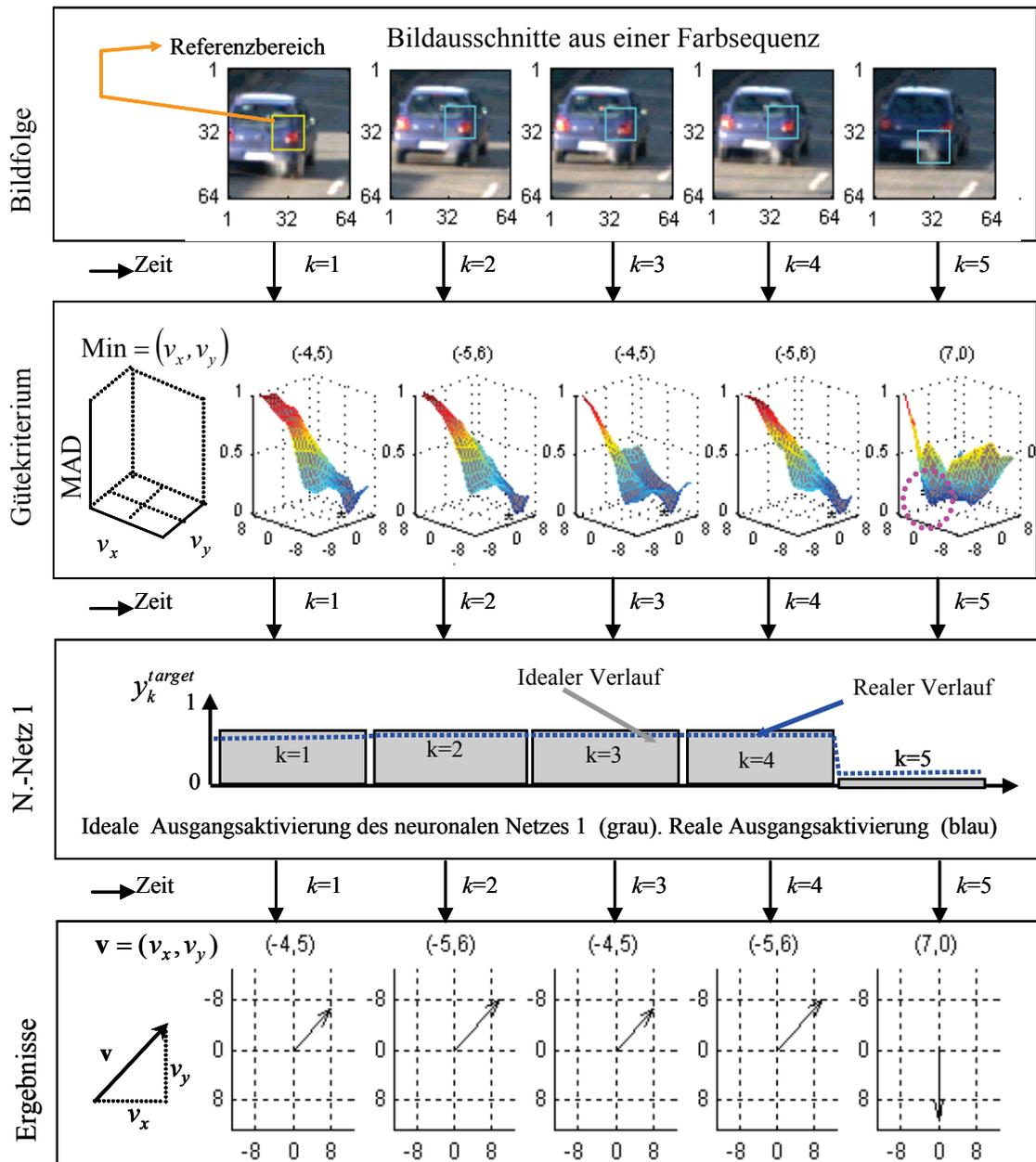
$H_1$ und $H_2$	Der Hystereseschwellwert
$\mathbf{h}$	Das Farbhistogramm
$\mathbf{V}$	Das Bewegungskvektorfeld
$\mathbf{f}$ und $\mathbf{f}'_{\text{ell}}$	Die umgebende Ellipse und die prädizierte Ellipse
$\text{MIN}_\theta$	Der <i>Minkowski</i> -Abstand für die relativen Häufigkeiten $h_{\text{rel}1}$ und $h_{\text{rel}2}$
$HS$	Der <i>Histogrammschnitt</i> (ein Abstandmaß für Histogramme)
ST	Der Straight-Through-Ansatz
MS	Der Merge-Split-Ansatz
$(T_{j,k})$	Die Objekttrajektorie
$\Phi_{j,k}$	Die Menge der Knoten einer Objekttrajektorie
$(\mathbf{a}_{j,k})$	Der <i>Aktivitätsgrad</i> einer Trajektorie
F-BM:	Das Fuzzy-Bewegungsmodell
LV	Die linguistischen Variablen
$\eta_T(\Phi_{j,k-1})$	Die Anzahl der Bewegungsvektoren des Endknotens $\Phi_j$ zur Zeit $k-1$
$N$	Die Anzahl der Partikel einer Partikelmenge
$N_k$	Die Anzahl der Knoten einer Trajektorie
$N_{\text{merge}}$	Die Anzahl der Trajektorien, die mit einem MB verbunden wurden
$N_{\text{parent}}$	Die Anzahl der Trajektorien, aus der eine Trajektorie entstanden ist (Elterntrajektorie)
$N_{\text{split}}$	Die Anzahl der MBs, die mit einer Trajektorie verbunden wurden
$N_T$	Die Anzahl der Trajektorien eines Partikels
$\Delta_p$	Der formbasierte Fehlerterm
$F_{\text{Konsistenz}}$	Das Konsistenzkriterium
$\Delta f_{\text{ANF}}$	Das Fehlermaß $\Delta f_{\text{ANF}}$ für den Anfangsknoten
$\Delta f_{\text{END}}$	Das Fehlermaß $\Delta f_{\text{END}}$ für den Endknoten
SVM	Support-Vector-Machine (Klassifikation)
NN-Verfahren	Nearest-Neighbor-Verfahren.



## 8. Anhang

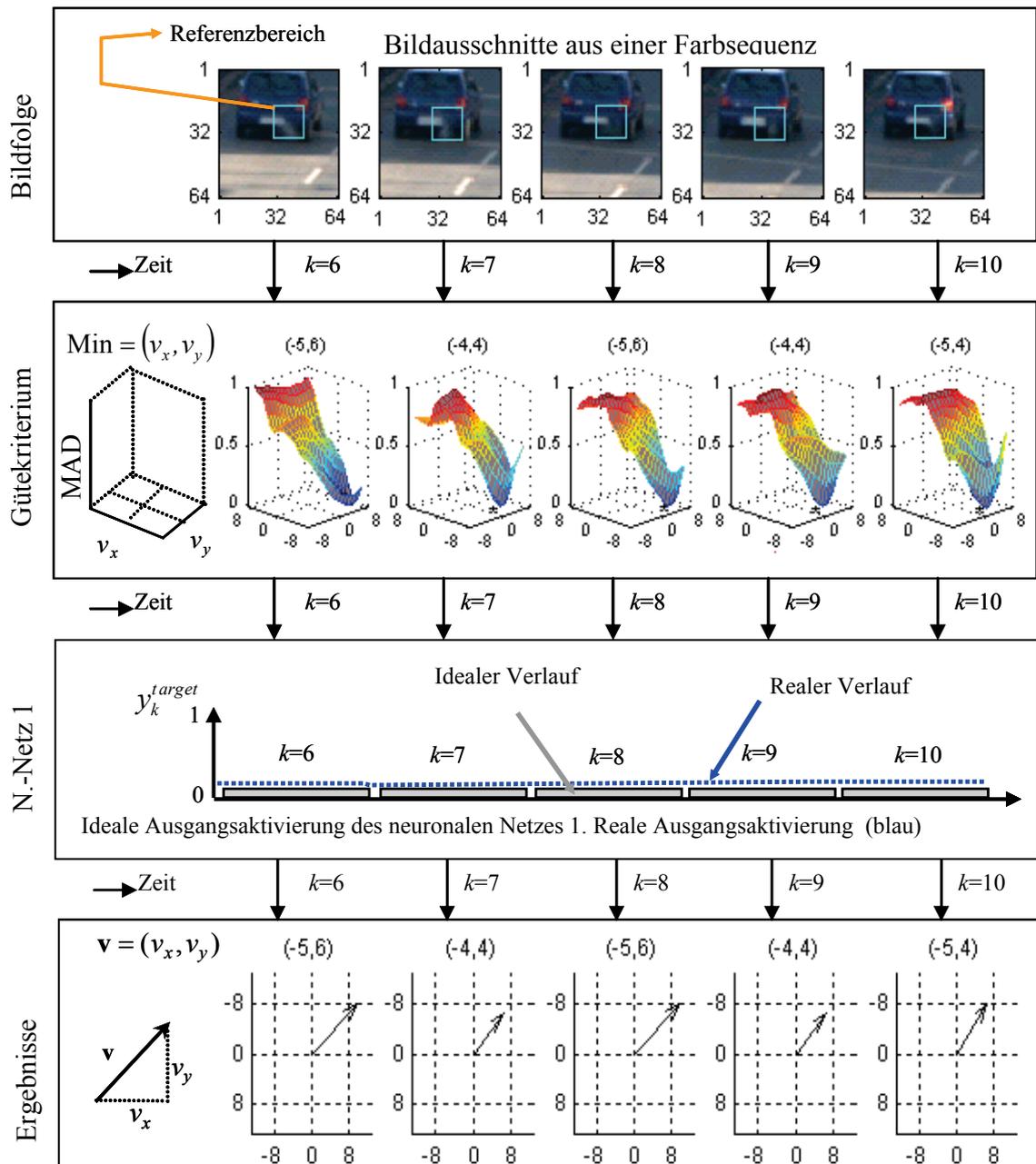
### 8.1 Der Einfluss von nicht-kooperativen Messsituationen

A1) Achromatische (helligkeitsabhängige) Komponente  $K_1$  des adaptiven Farbraumes. Das Objektbild wird von einem Schattenwurf in Bild  $k=5$  und der Beleuchtungsänderung im Bild  $k=10$  beeinflusst.



**Abbildung 8-1-1A1:** Darstellung des Signalflusses in der Verarbeitungskette (Abb. 2-2-1) am Beispiel einer Bildsequenz mit Schattenwurf. Es sind die MAD-Funktionen der achromatischen Komponente für den markierten Referenzbereich in der Sequenz dargestellt. Der Referenzbereich (gelbmarkiert) ist im Bild  $k=1$  bezüglich des Objektes positioniert. Ab Bild  $k=5$  erkennt man starke Verformungen der MAD-Funktion. Diese sind auf die Überlagerung der inhomogenen Intensitätsänderung des Objektes innerhalb der Sequenz zurückzuführen. Das Minimum der MAD-Funktion verschiebt sich nun in Richtung der hellsten Zone, da der tatsächlich gesuchte Bereich große Ähnlichkeit zum Referenzbereich (s. Bild  $k=5$ ) aufweist (**Fortsetzung**).

## A1) Fortsetzung der Analyseergebnisse.



**Fortsetzung zur Abbildung 8-1-1A1:** Das Nebenminimum dominiert und führt zum falschen Verschiebungsvektor ( $k=5$ ), der die tatsächliche Bewegung des Objektes nicht beschreibt. Die MAD-Funktion bei  $k=5$  wird somit anhand des neuronalen Netzes als nicht zuverlässig interpretiert. Daher muss ab dem Bild  $k=5$  eine Ausgangsaktivierung nahe Null (nichtzuverlässig) auftreten. Anhand des unten dargestellten Verschiebungsvektors erkennt man die signifikante Änderung bei  $k=5$ . Bei Auswertung der MAD-Funktion würde bis  $k=10$  eine Fehlanalyse erfolgen.

A2) Chromatische (helligkeitsunabhängige) Komponente  $K_2$  des adaptiven Farbraumes. Das Objekt wird von einem Schattenwurf in Bild  $k=5$  und der Beleuchtungsänderung im Bild  $k=10$  beeinflusst.

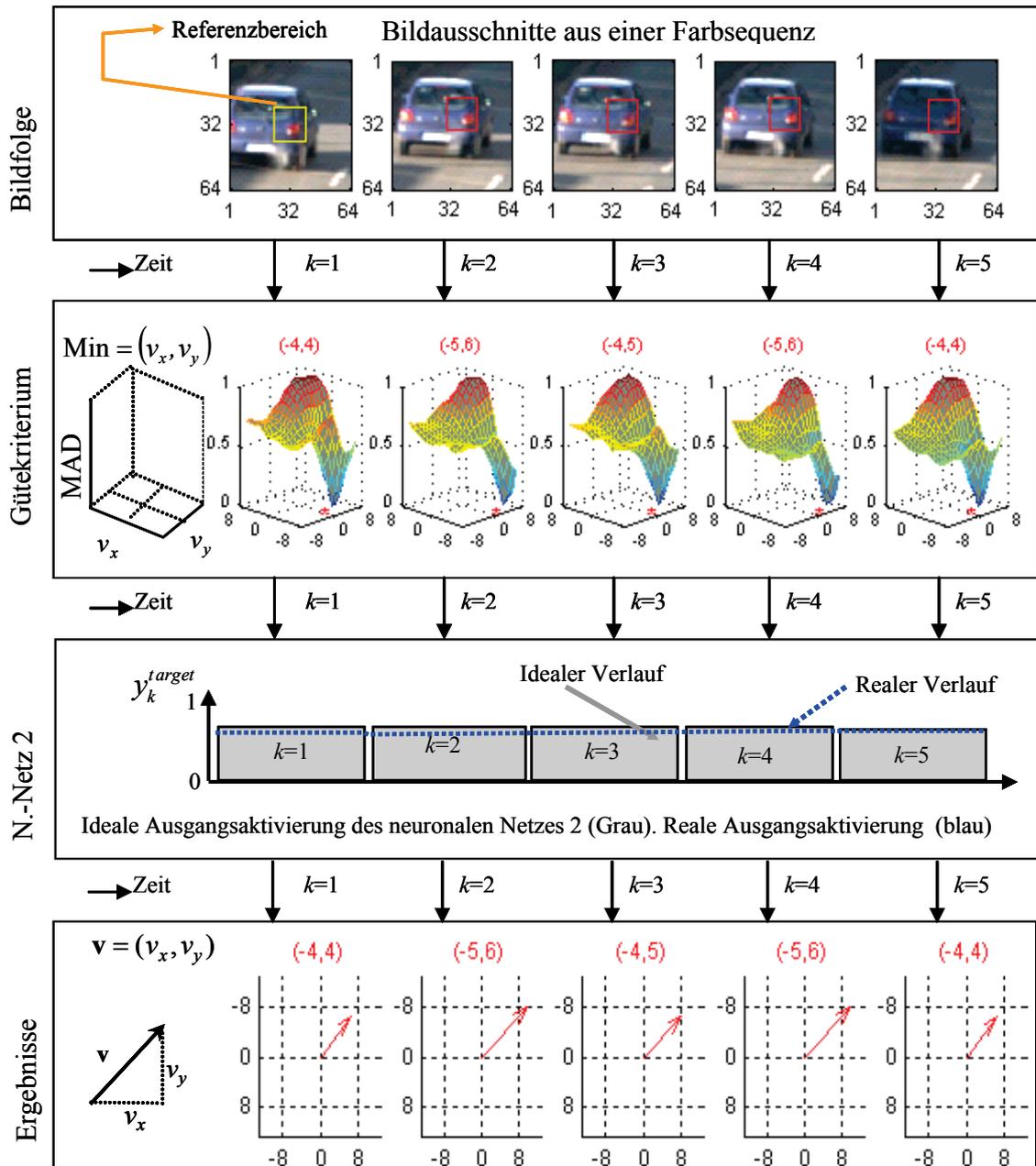
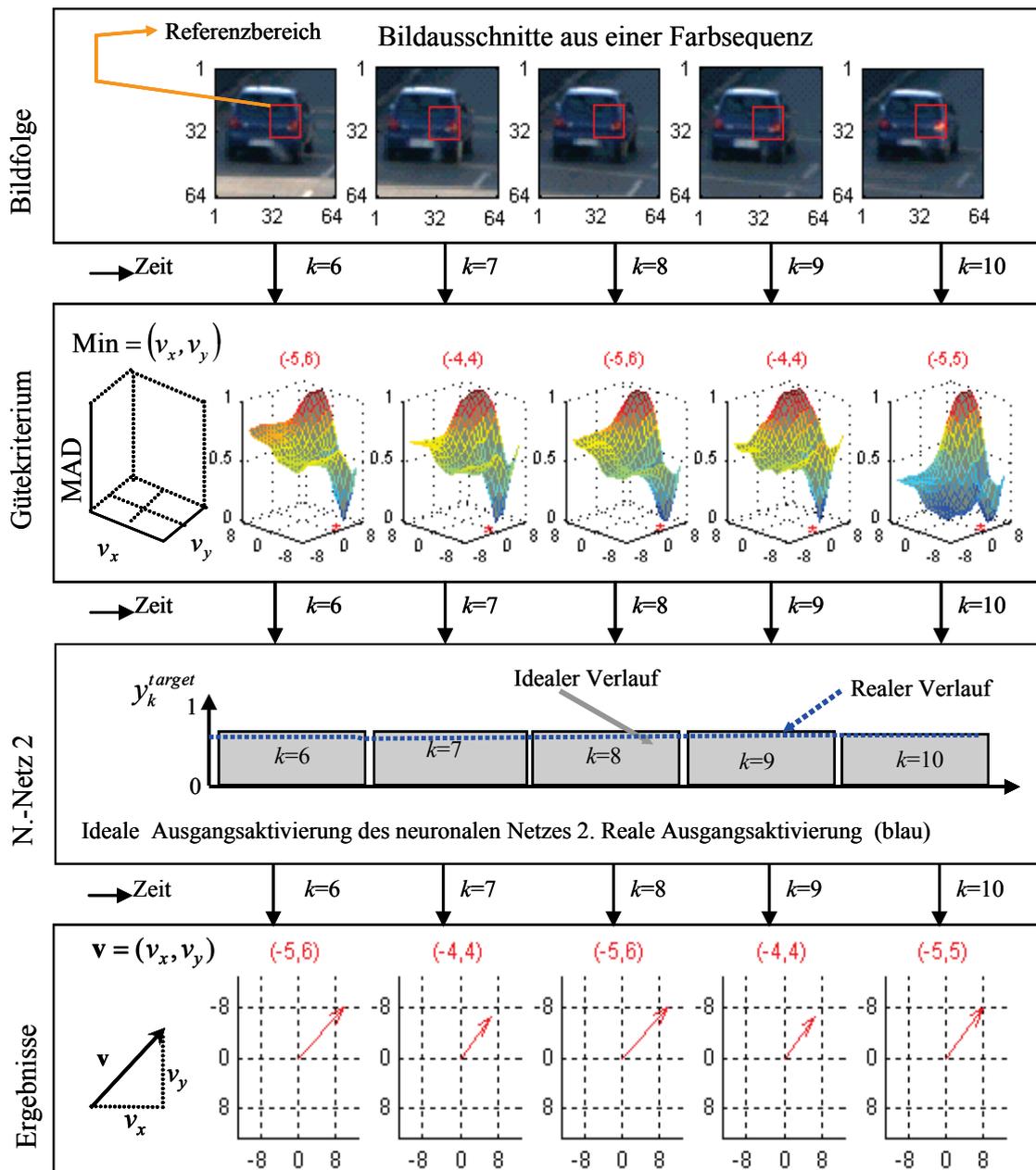


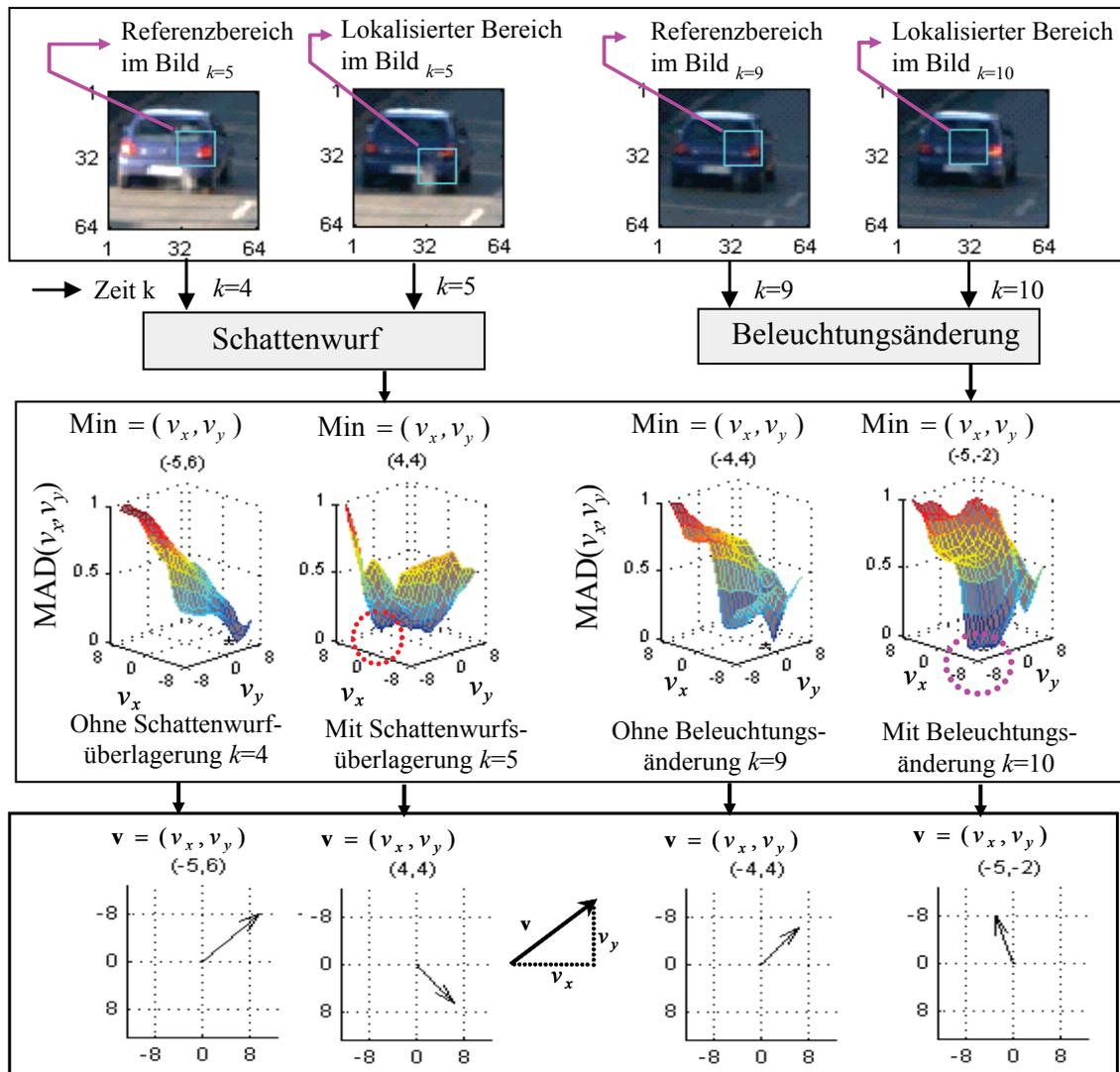
Abbildung 8-1-1A2: Darstellung des Signalflusses in der Verarbeitungskette am Beispiel einer Bildsequenz mit Schattenwurf und Beleuchtungsänderungen. Es sind die MAD-Funktionen der chromatischen Komponente für den markierten Referenzbereich in der Sequenz dargestellt. Der Referenzbereich (interessante Region gelb markiert) ist im Bild  $k=1$  bezüglich des Objektes positioniert. Das Objekt wird ab dem Bild  $k=5$  von einem Schattenwurf der Bauten in Fahrbahnnähe überlagert. Da aber die  $K_2$ -Komponente gegenüber solchen Änderungen weitgehend invariant ist, weist die MAD-Funktion einen guten und signifikanten Verlauf im Bereich des Minimums auf (Fortsetzung).

A2) Fortsetzung der Analyseergebnisse.



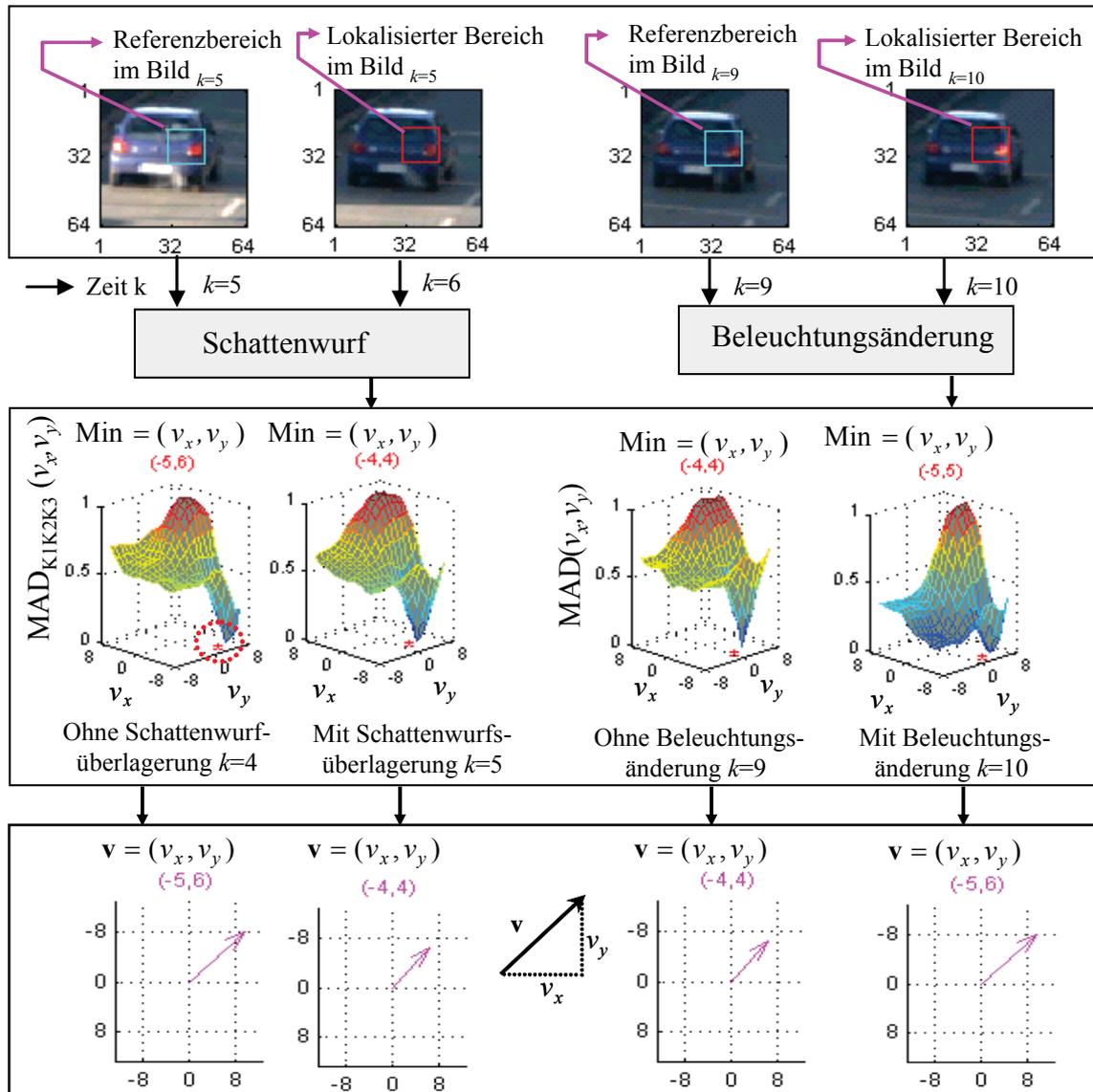
**Fortsetzung zur Abbildung 8-1-1A2:** Die MAD- Funktion hat sich bei  $k=5$  nicht deformiert, deshalb ist die Verfolgung des Objektes innerhalb der Sequenz gesichert. Das neuronale Netz 2 bewertet diese Verläufe als zuverlässige MAD-Funktionen. Die MAD-Funktion bei  $k=10$  ist gering deformiert. Die Lage des Minimums hat sich nicht verschoben. Dies ist auf die Tatsache zurückzuführen, dass das Objekt im Bild  $k=10$  von inhomogenen Beleuchtungs-änderungen überlagert wurde. Die abgeleiteten Verschiebungsvektoren sind unten angegeben, wobei man die Kontinuität des Verlaufes der Vektoren erkennen kann. Diese Vektoren beschreiben nun die tatsächliche Bewegung des Objektes. Dieses ist anhand des rot markierten Referenzbereiches innerhalb der Bilder zu erkennen.

**B1)** Der Einfluss von nicht-kooperativen Messsituation (Schattenwurf und Beleuchtungsänderung) auf die Analyseergebnisse mittels herkömmlicher Verfahren, die auf der Helligkeitsinformation basieren.



**Abbildung 8-1-1B1:** Bildausschnitte aus einer längeren Farbbildsequenz. Das Objekt (cyan markiert) ist hier von einem Schattenwurf (links) und einer Beleuchtungsänderung (rechts) überlagert. Die berechneten MAD-Funktionen für die Bilder  $k=4$  und  $k=5$  sind unten dargestellt. Eine ausschließliche Minimumsuche detektiert in diesem Fall nur das dominante Nebenminimum (rot markiert). Dieses führt zur Unterdrückung des Haupt-minimums (beim Schattenwurf). Als Vergleich ist daneben die berechnete MAD-Funktion (ohne Schattenwurfüberlagerung) dargestellt. Der von der deformierten MAD-Funktion ermittelte Verschiebungsvektor weist nun einen falschen Messwert, der zu einer falschen Objektverfolgung führt auf (s. Bild  $k=5$  oben cyan markierter Bereich). Die Bilder  $k=9$  und  $k=10$  beschreiben den Einfluss der Beleuchtungsänderung auf die Verfolgung. Daher erkennt man unten (rechts) die starke Deformation der MAD-Funktion. Das Nebenminimum (Aufgrund der Beleuchtungsänderungen) ist dominant (rot markiert). Dieses führt zu einem falschen Verschiebungsvektor, der sich anhand der oben angegebenen Positionierung des gefundenen Objekts zeigen lässt. Die MAD-Funktion ohne Beleuchtungsänderung ist daneben dargestellt.

**B2)** Der Einfluss von nicht-kooperativen Messsituation (Schattenwurf und Beleuchtungsänderung) auf die Analyseergebnisse mit Hilfe der vorgeschlagenen adaptiven Systemstruktur, die auf Auswertung der Farbinformation basiert.



**Abbildung 8-1-1B2:** Bildausschnitte aus einer längeren Farbbildsequenz. Das Objekt (cyan markiert) ist hier von einem Schattenwurf (links) und einer Beleuchtungsänderung (rechts) überlagert. Die berechneten MAD-Funktionen für die Bilder  $k=4$  und  $k=5$  sind unten dargestellt. Eine ausschließliche Minimumsuche detektiert in diesem Fall das Hauptminimum (rot markiert). Dieses führt zur Unterdrückung des Hauptminimums (bei der Schattenwurfüberlagerung). Als Vergleich ist daneben die berechnete MAD-Funktion (ohne Schattenwurfüberlagerung) dargestellt. Die beiden MAD-Funktionen weisen einen signifikanten Verlauf im Bereich des Minimums auf. Dadurch erkennt man die Stabilität der Analyse bei dem Einfluss vom Schattenwurf. Der von dieser MAD-Funktion ermittelte Verschiebungsvektor beschreibt nun die Bewegung des Objektes (s. Bild  $k=5$  oben rot markierter Bereich). Die Bildern  $k=9$  und  $k=10$  beschreiben den Einfluss des Beleuchtungsänderung auf die Verfolgung. Daher erkennt man unten (rechts) die gewisse Deformation der MAD-Funktion bei diesem extrem Fall. Erstaunlicherweise ist die Lage des Minimums stabil. Dadurch ist die Objektverfolgung auch in solchen Situationen  $k$  mittels des vorge-schlagen Systems gewährleistet.

## 8.2 Normierungskoeffizienten für den Ellipsenabstandsfehler

Der Ellipsenabstandsfehler  $\Delta_r$  wird in der Gleichung (5-3-11d) definiert und durch die Koeffizienten  $d_a$  und  $d_b$  normiert. Wird eine Gerade durch die Mittelpunkte zweier Ellipsen  $f_{\text{ell}}(A)$  und  $f_{\text{ell}}(B)$  gelegt, so schneidet diese Gerade die Ellipse  $f_{\text{ell}}(A)$  und  $f_{\text{ell}}(B)$  in ihrem Rand. Der Normierungskoeffizient  $d_a$  entspricht der Strecken vom Mittelpunkt der Ellipse A zu dem Schnittpunkt der Verbindungsstrecke AB mit dem Rand der Ellipse  $f_{\text{ell}}(A)$ . Analog verhält es sich mit der Ellipse  $f_{\text{ell}}(B)$  und dem Koeffizienten  $d_b$  (s. Abb. 8-1-1). Für die Berechnung der Ellipsen werden die folgenden Gleichungen benötigt:

- Die Vektorgleichung vom Ellipsenmittelpunkt zum Ellipsenrand,

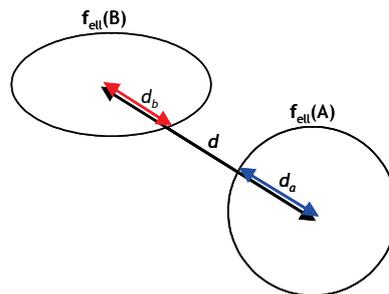
$$\begin{pmatrix} x \\ y \end{pmatrix} = t \cdot \begin{pmatrix} dx \\ dy \end{pmatrix}$$

gedreht um einen Winkel  $\beta$

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} t \cdot dx \cdot \cos \beta - t \cdot dy \cdot \sin \beta \\ t \cdot dx \cdot \sin \beta + t \cdot dy \cdot \cos \beta \end{pmatrix}$$

- und der Ellipsengleichung.

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} a \cdot \cos(s) \\ b \cdot \sin(s) \end{pmatrix}, \text{ mit } s=0 \dots 2\pi$$



**Abbildung 8-2-1:** Die schematische Darstellung der Berechnung der Normierungskoeffizienten  $d_a$  und  $d_b$  für den Ellipsenabstandsfehler.

Die Bestimmung des Schnittpunktes der Verbindungsstrecke der Ellipsen mit dem Rand der Ellipse  $f_{\text{ell}}(A)$  wird durch das Gleichsetzen der Gleichungen erreicht. Es gilt:

$$\cos(s)^2 = t^2 \left( \frac{dx \cdot \cos \beta - dy \cdot \sin \beta}{a} \right)^2 \quad \text{und} \quad \sin(s)^2 = t^2 \left( \frac{dx \cdot \sin \beta + dy \cdot \cos \beta}{b} \right)^2$$

somit errechnet sich  $t$  mit

$$t_{1/2} = \pm \sqrt{\frac{1}{\left( \frac{dx \cdot \cos \beta - dy \cdot \sin \beta}{a} \right)^2 + \left( \frac{dx \cdot \sin \beta + dy \cdot \cos \beta}{b} \right)^2}}$$

wegen der Symmetrie der Ellipse wird nur ein Schnittpunkt benötigt. Der Koeffizient für die Ellipse  $f_{\text{ell}}(A)$  ergibt sich so aus

$$d_a = \sqrt{(t \cdot dx)^2 + (t \cdot dy)^2}$$

mit  $dx = x_m^B - x_m^A$  und  $dy = y_m^B - y_m^A$

$a$  : erste Halbachse der Ellipse A

$b$  : zweite Halbachse der Ellipse A und

$\beta$  : Drehung der Ellipse A.

Für die Berechnung des zweiten Koeffizienten  $d_b$  sind lediglich die Parameter der Ellipse  $f_{\text{ell}}(A)$  mit den Parametern der Ellipse  $f_{\text{ell}}(B)$  auszutauschen.

### 8.3 Untersuchungen zum Vergleich von Farbhistogrammen

In dieser Arbeit sollen Objekte anhand ihrer Farbinformation voneinander unterschieden werden. Dazu finden Farbhistogramme ihre Verwendung. Um die Farbinformation eines Objektes mit einem Histogramm darstellen zu können, muss zuerst die *Quantisierung* (Anzahl der Urnen) festgelegt werden. In diesen Versuchen wurde mit Quantisierungstufen von  $2^3$ ,  $2^4$ ,  $2^5$ ,  $2^6$  und  $2^8$  gearbeitet. Des Weiteren muss festgelegt werden, welcher Farbraum auf das Histogramm abgebildet werden soll. Dabei kann durch eine geeignete Transformation des Farbraumes nicht nur die Separabilität zweier Objekte erhöht werden, sondern auch die abzubildenden Farbkanäle reduziert werden. Da eine Reduzierung der verwendeten Farbkanäle eine Ersparnis an Rechenzeit bedeutet, wird auch in den Versuchen eine unterschiedliche Anzahl von Kanälen untersucht. Die verwendeten *Farbräume* sind:

- **RGB-Farbraum:** Dieser Farbraum enthält die aus den Bildsequenzen entnommenen Rohdaten und benötigt somit keine weitere Verarbeitung.
- **$Y C_b C_r$ -Farbraum:** Dieser Farbraum gehört zu den wahrnehmungsorientierten Farbräume und kann durch eine lineare Transformation aus dem RGB-Farbraum gewonnen werden. Er wird durch die Luminanzkomponente (Y) und die Chrominanzkomponenten ( $C_b$  und  $C_r$ ) beschrieben.
- **$C_b C_r$ -Farbraum:** Mit dem  $c_b c_r$ -Farbraum als Teilraum des  $Y C_b C_r$ -Farbraumes, soll untersucht werden, ob das Weglassen der Helligkeitsinformation (Luminanz) eine Verbesserung und Optimierung des Rechenaufwandes mit sich bringt.
- **HSV-Farbraum:** Der HSV-Farbraum wird durch eine nichtlineare Transformation aus dem RGB-Farbraum gewonnen. Der Vorteil dieses Raumes besteht darin dass Farbton (H), die Sättigung (S) und die Helligkeit (V) in drei Kanälen getrennt sind.
- **HS-Farbraum:** Als Teilraum des HSV-Farbraums lässt sich die Farb- und Sättigungskomponente trennen und untersuchen, so dass wie beim  $C_b C_r$ -Farbraum Helligkeitsschwankungen keinen Einfluss auf den Farbvergleich besitzen.
- **H-Farbraum:** Es wird zusätzlich die Sättigung außer Acht gelassen. Mit diesem Farbraum wird untersucht, ob allein die Farbigekeit ausreicht, um Objekte voneinander zu unterscheiden.

Um nun für zwei Histogramme eine Distanz  $d(\mathbf{h}_1, \mathbf{h}_2)$ , entsprechend eines Ähnlichkeitsmaßes, angeben zu können, wurden sechs *Histogrammdistanzfunktionen* angewendet:

- **min** (Minkowski-Abstand)
- **HS** (Histogrammschnitt)
- **KLD** (Kullback-Leibler-Divergenz)
- **JD** (Jeffrey-Divergenz)
- **$\chi^2$**  ( $\chi^2$ -Abstand)
- **BA** (Bhattacharya-Abstand)

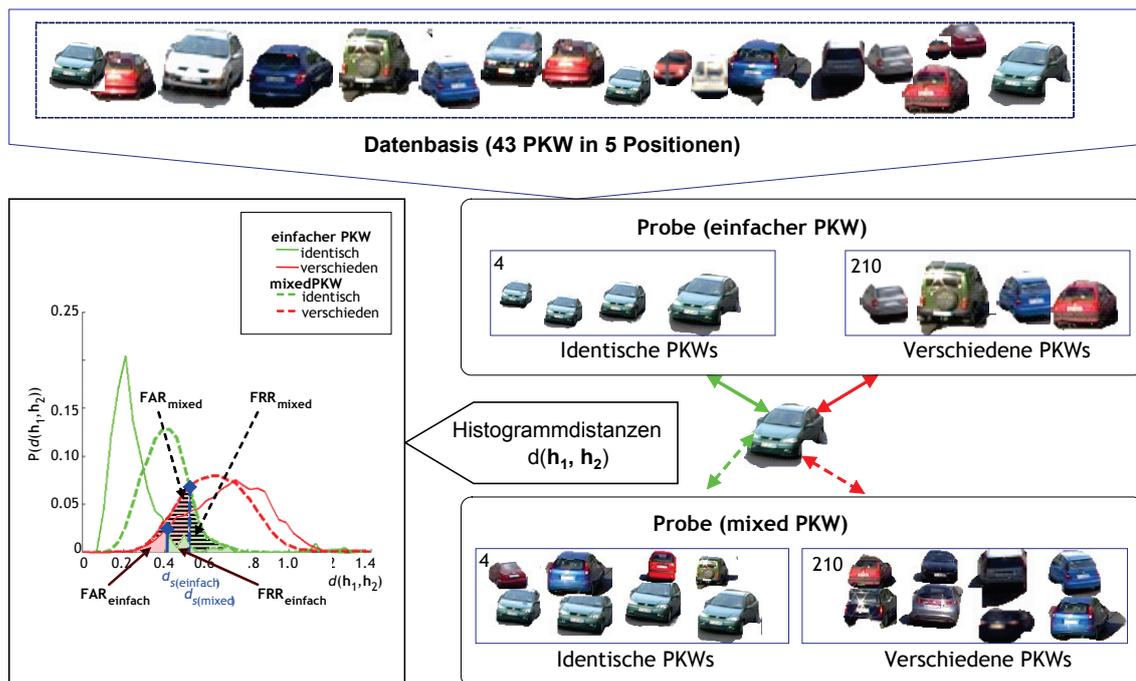
Es bestehen somit in diesem Fall 30 verschiedene Varianten, um ein Histogramm für ein Objekt zu erzeugen und sechs verschiedene Möglichkeiten, um die Ähnlichkeit zwischen zwei Objekten durch die Farbhistogrammdarstellung zu berechnen.

## 8. ANHANG

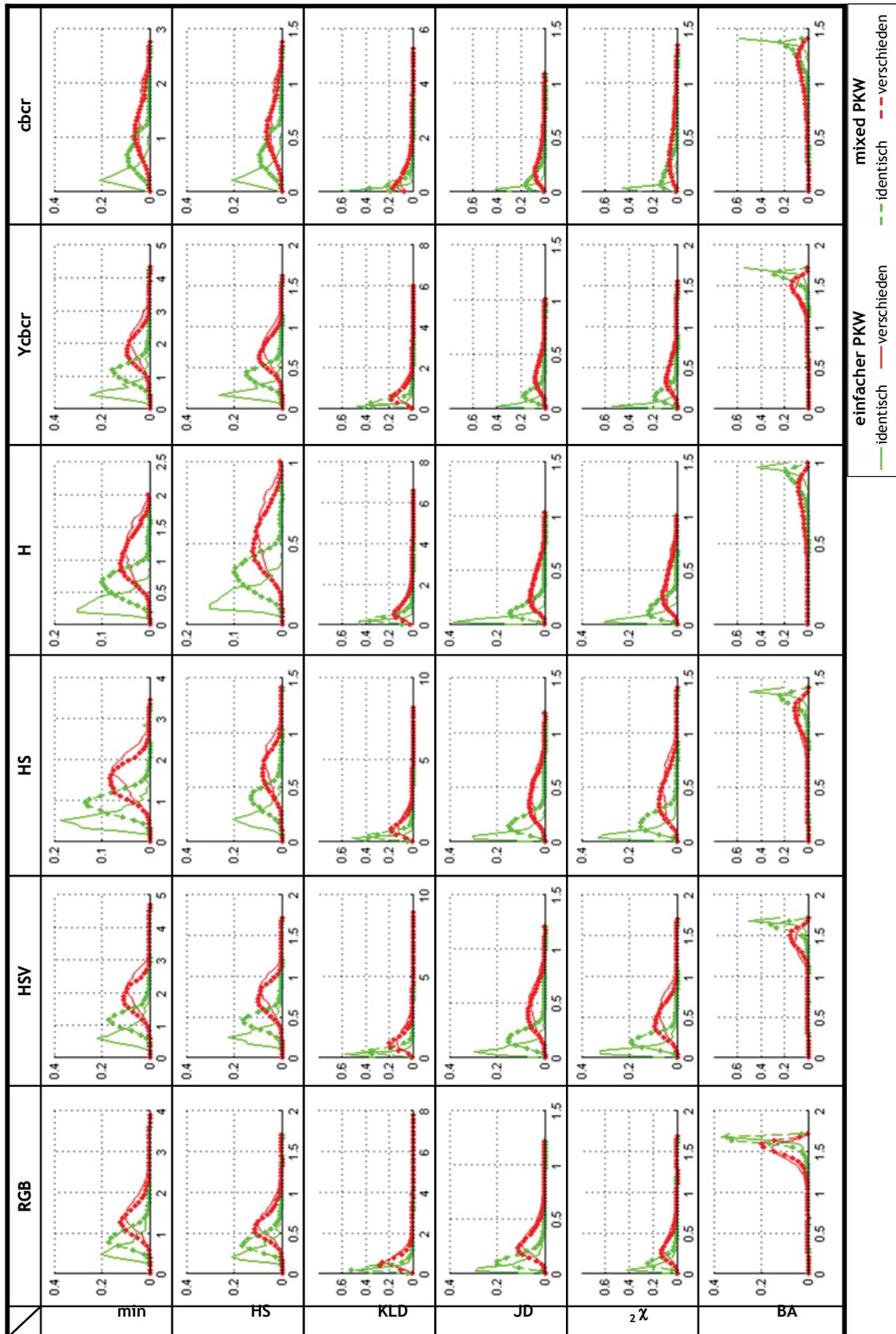
Um die geeignetste Variante zu bestimmen wurde eine Datenbasis bestehend aus 43 PKW in 5 verschiedenen Positionen ( $n=215$ ) aufgestellt. Das Ziel ist es für jede Variation die Verteilung der Histogrammdistanzen für identische und verschiedene Objekte zu ermitteln. Dies wird für zwei Fälle (s. Abb. 8-3-1) untersucht:

- **einfacher PKW:** Ein PKW wird mit einer Probe, bestehend aus einem PKW verglichen.
- **mixed PKW:** Ein PKW wird mit einer Probe, bestehend aus zwei PKW verglichen.

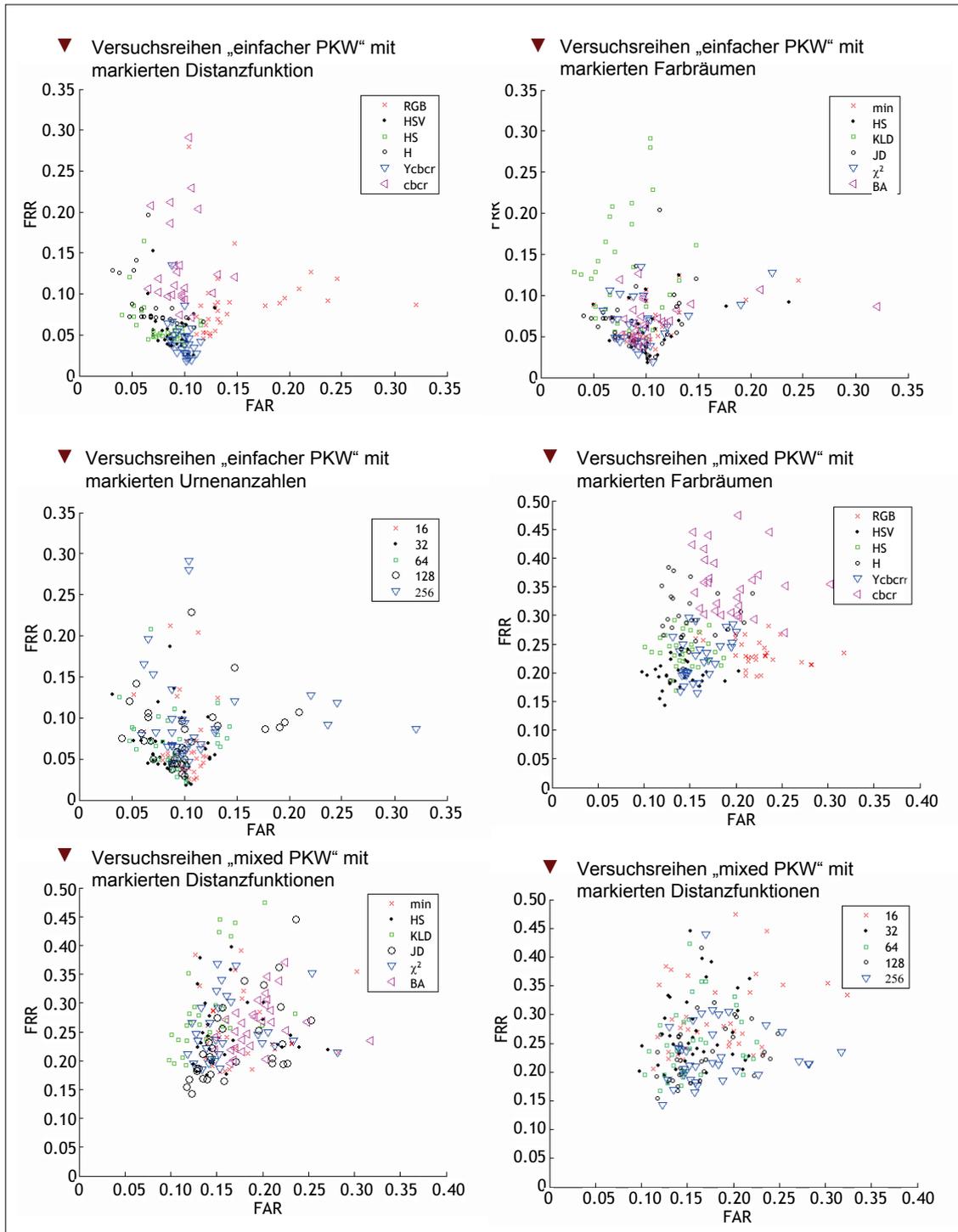
Der Grund für diese zwei Untersuchungen sind die spezifischen Anforderungen des Bewegungsmodells und des Beobachtungsmodells an die Histogrammdistanzen. Die errechneten Verteilungsfunktionen sind in den Abbildungen (8-3-2) dargestellt. Die Güte der Histogrammvariationen wird durch die FAR (false acceptance rate) und den FRR (false rejection rate) nach [Knauer 05] bestimmt. Dazu wird der Schnittpunkt  $d_s$  der Verteilungen für identische und verschiedene Proben bestimmt. Die FAR ist dann Anteil der negativ Proben für die gilt  $d(\mathbf{h}_1, \mathbf{h}_2) < d_s$  und FRR ist umgekehrt der Anteil der positiven Proben deren Distanzen größer gleich der Schwelle  $d_s$  sind. Im optimalen Falle zeigen ein FAR und ein FRR von Null an, dass die identischen und verschiedenen Häufigkeitsverteilungen keinen Überlappungsbereich besitzen. Die Resultate der Versuchsreihen sind in Abbildung (8-3-1) und Abbildung (8-3-2/3) dargestellt.



**Abbildung 8-3-1:** Die schematische Darstellung der Versuchsdurchführung für eine Variation. Für jedes Fahrzeug werden zwei Versuchsreihen „einfacher PKW“ und „mixed PKW“ betrachtet. Aus der Verteilung der ermittelten Farbhistogrammdistanzen (links) werden die FRR und der FAR als Gütemaß ermittelt.



**Abbildung 8-3-2:** Ausschnitt aus den errechneten Häufigkeitsverteilungen der Farbdistanzen für die Versuche mit einer Quantisierung von 64 Histogrammurnen unter der Variation der Farbräume und Histogrammdistanzfunktionen [Senst 09].



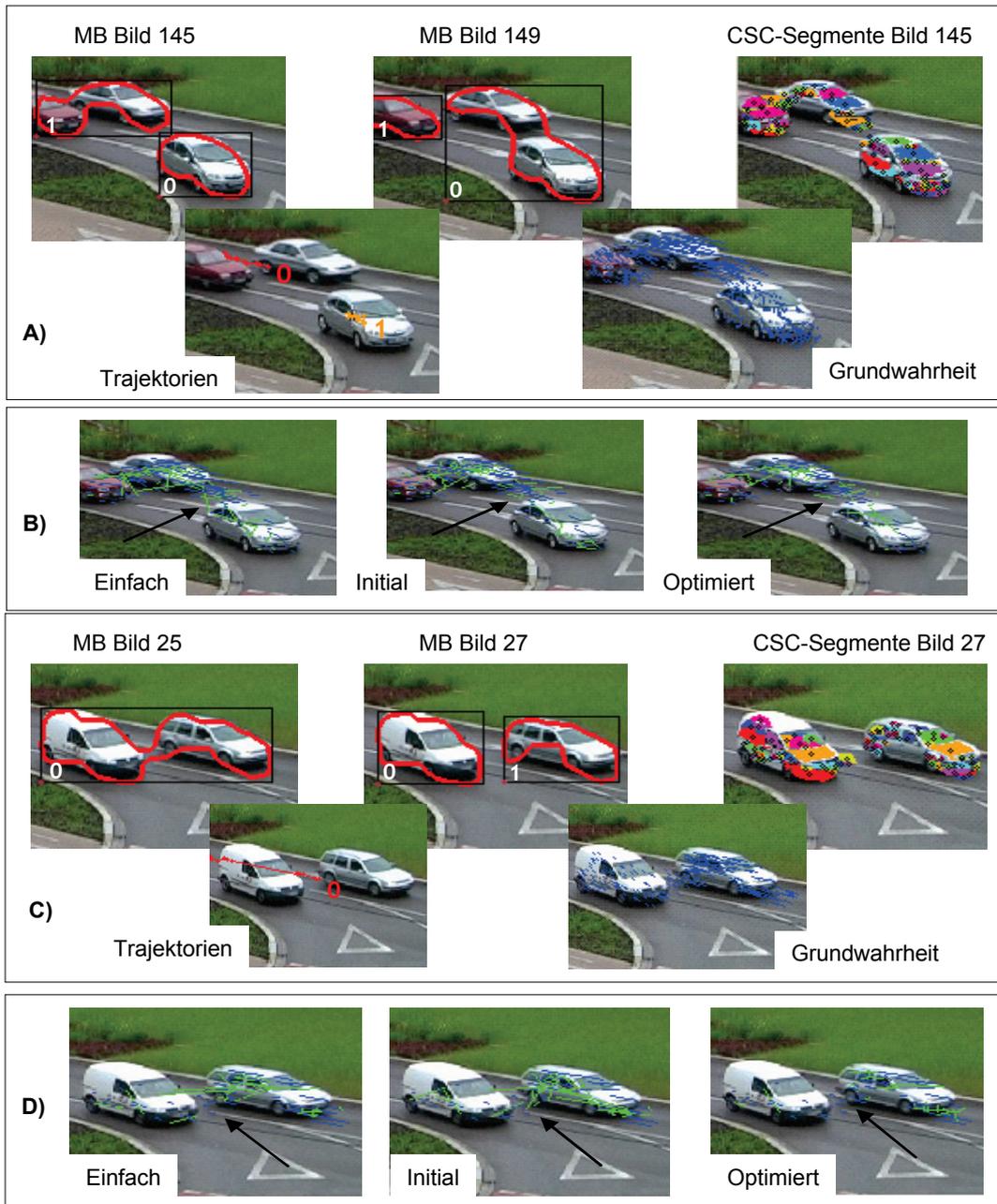
**Abbildung 8-3-3:** Darstellung der Gütemaße FAR und FRR der Versuchsreihen. Die Diagramme zeigen die FAR und die FRR der 180 Versuchsreihen in den Kategorien „einfacher PKW“ und „mixed PKW“. Die Datenpunkte sind dabei jeweils in ihrer Zugehörigkeit zum Farbraum, Urmenanzahl und Distanzfunktion markiert.

#### 8.4 Beispielsequenzen: Optimierung der Gewichte

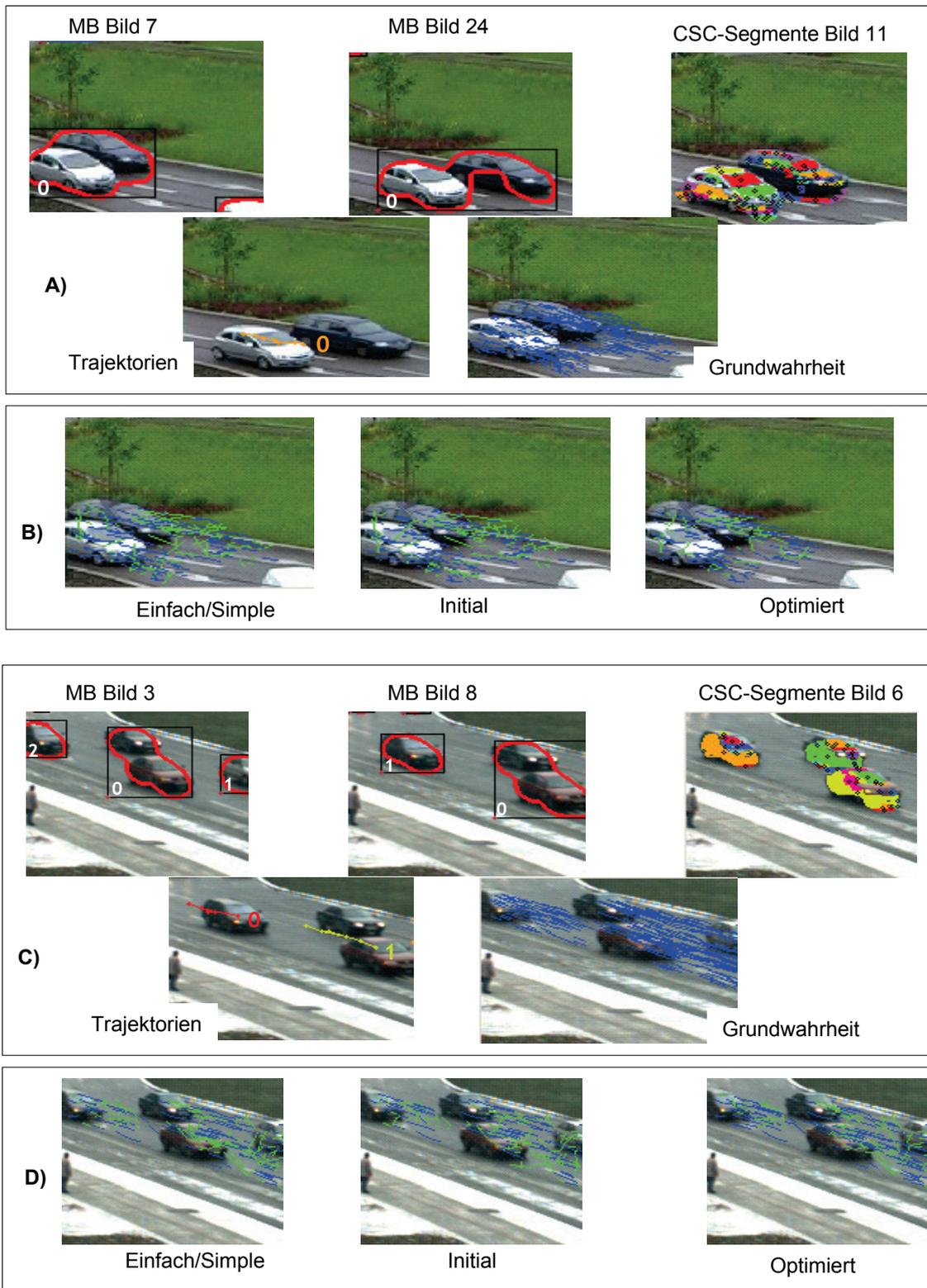
Für die Validierung der optimierten Gewichte werden die Teilszenen mit den errechneten Parametern durchlaufen. Es wird die Zuordnungsmatrix, in Abhängigkeit von dem Gewichtsvektor  $w$  und dem Algorithmus zur Auflösung mehrdeutiger Zuordnungen, errechnet. Aus dem Vergleich mit der Zuordnungsmatrix der Grundwahrheit werden die

## 8. ANHANG

*falsch* und *richtig* zugeordneten Paarungen sowie der Restfehler  $f_{SRE}$  erfasst (s. Abschnitt 3.6).



**Abbildung 8-4-1:** A) Verkehrsaufnahme „Verschmelzung“ mit der Fehlerbetrachtung des CSC-Matchings (B). Richtige CSC-Segment-Paarungen werden blau und falsche werden grün dargestellt. C) Verkehrsaufnahme „Teilung“ mit der Fehlerbetrachtung des CSC-Matchings. Da sich die Störung der Objektverfolgung durch ein ungenaues CSC-Matching ergibt, wird eine Optimierung der Gewichte durchgeführt. Dieser Störeffekt tritt während des Gesamtprozesses nur dann auf, wenn sich mehrere Objekte in einem MB befinden. Das Ähnlichkeitsmaß der relativen Position besitzt die spezielle Eigenschaft, dass es von der Größe des MBs abhängig ist. Bei der Verschmelzung oder der Teilung eines MBs wird dieses Maß gestört. Es ist deshalb notwendig die Optimierung der Gewichte bei Szenen mit Verschmelzung und Teilung und bei Szenen mit konstantem MB getrennt zu betrachten.

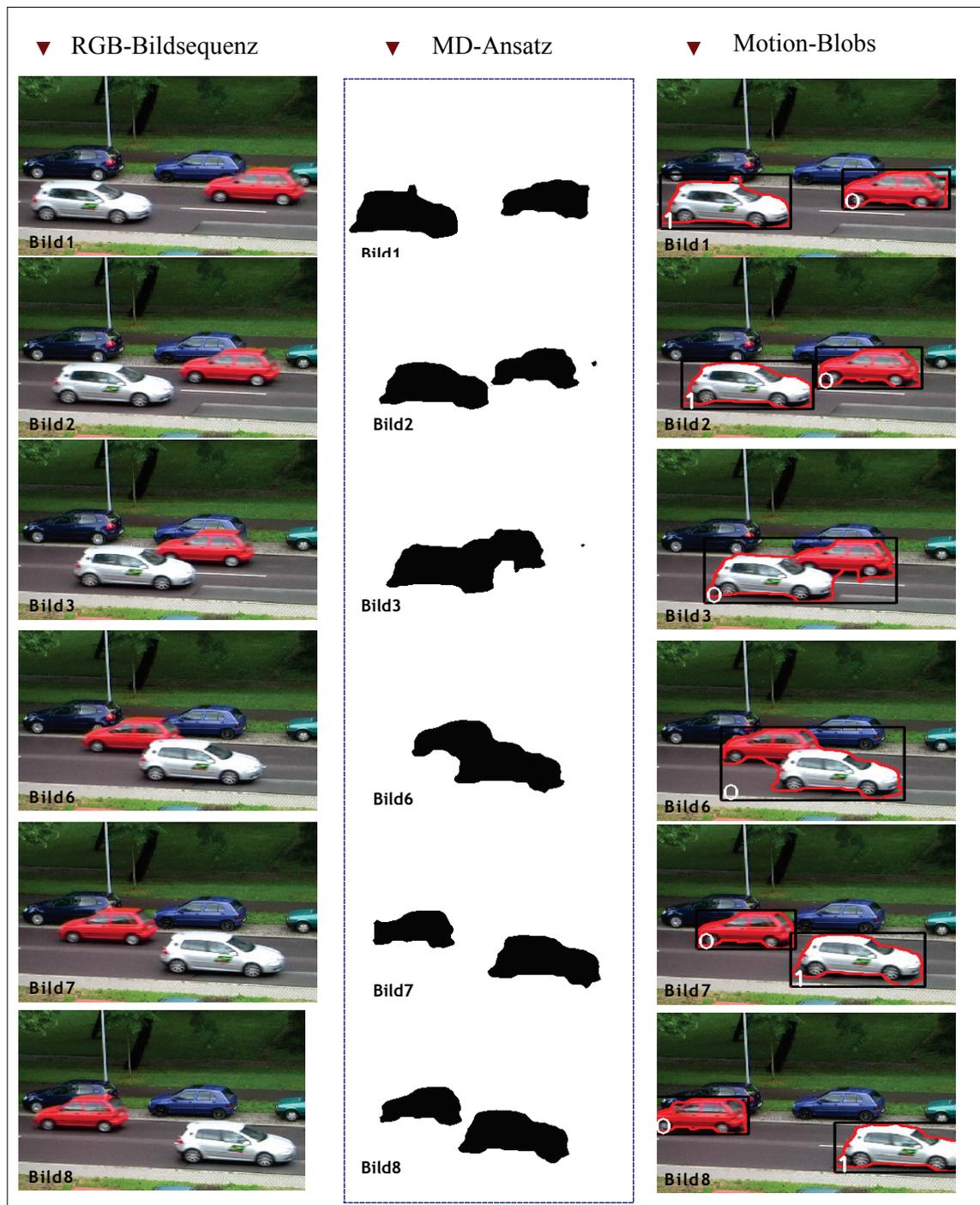


**Abbildung 8-4-2:** A) Verkehrsaufnahme „Konstanter MB1“ mit der Fehlerbetrachtung des CSC-Matchings B). Richtige CSC-Segment-Paarungen werden blau und falsche werden grün dargestellt. C) und D) Verkehrsaufnahme „Konstanter MB2“ mit der Fehlerbetrachtung des CSC-Matchings.

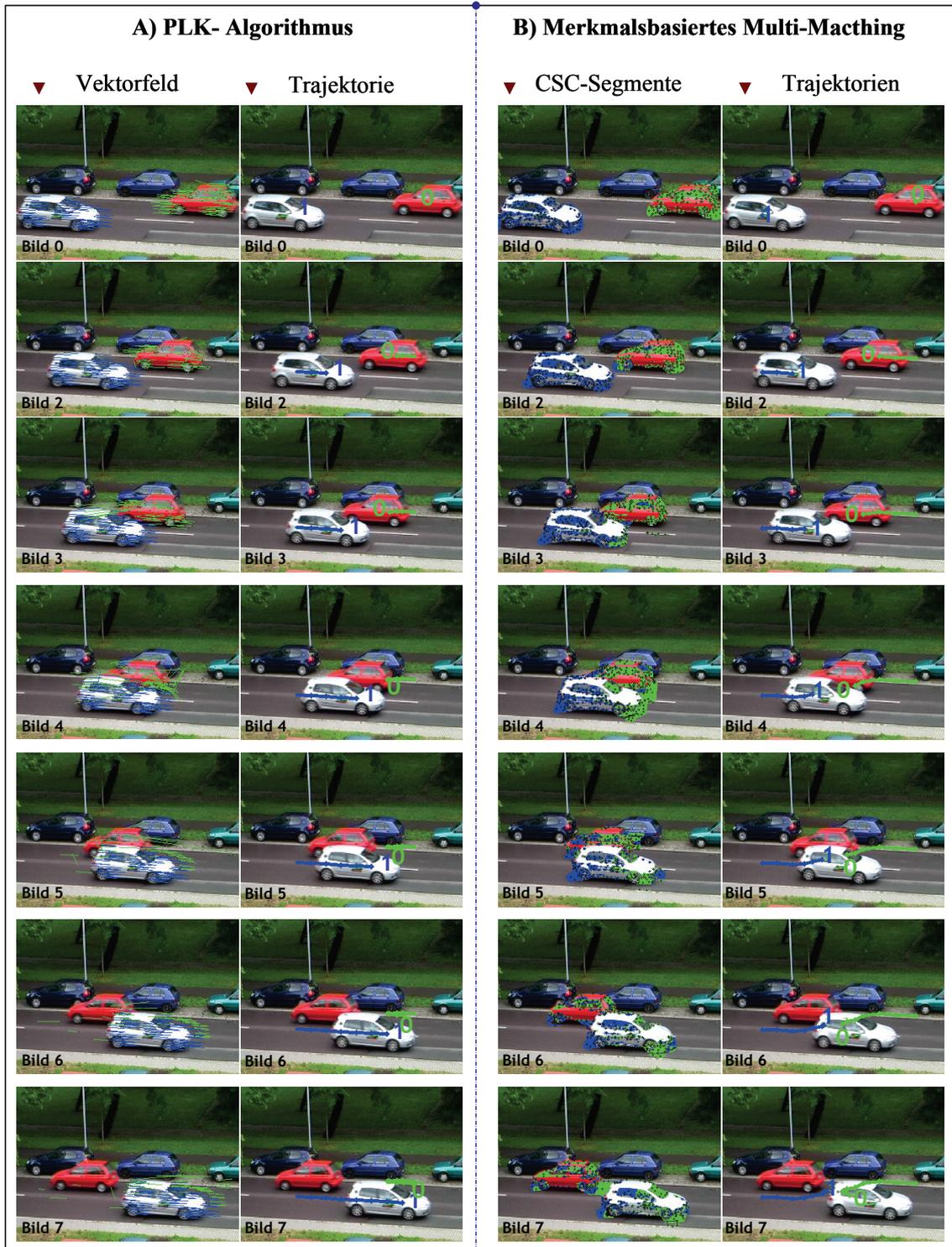
## 8.5 Beispielsequenzen: Einfache Konfliktsituationen

Die folgenden Abbildungen enthalten Ausschnitte aus mehreren Bildsequenzen, die die Aufgabe haben, die Funktion des Systems im Kapitel 5 unter dem Einfluss von von nicht-kooperativen Messsituationen zu verdeutlichen.

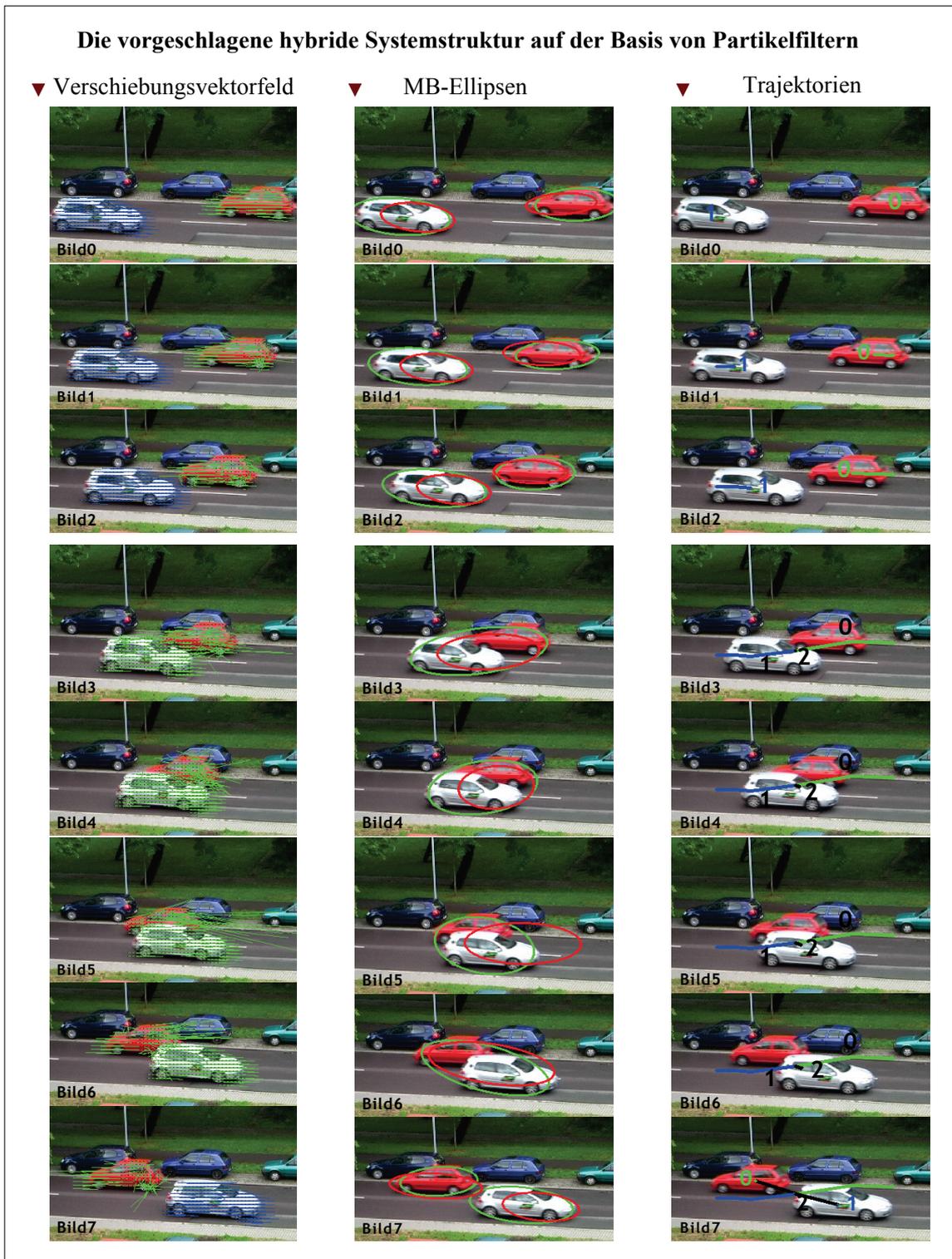
### 8.5.1 Szene 1



**Abbildung 8-5-1a:** Verlauf der MB-Extrahierung durch den vorgeschlagenen MD-Ansatz in Verbindung mit dem Nachverarbeitungsalgorithmus für die Szene 1. Die Konfliktsituation in dieser Szene besteht darin, dass im Bild 3 die MB des roten und weißen PKW zu einem MB verschmelzen, der dann im Bild 7 wieder getrennt werden muss.



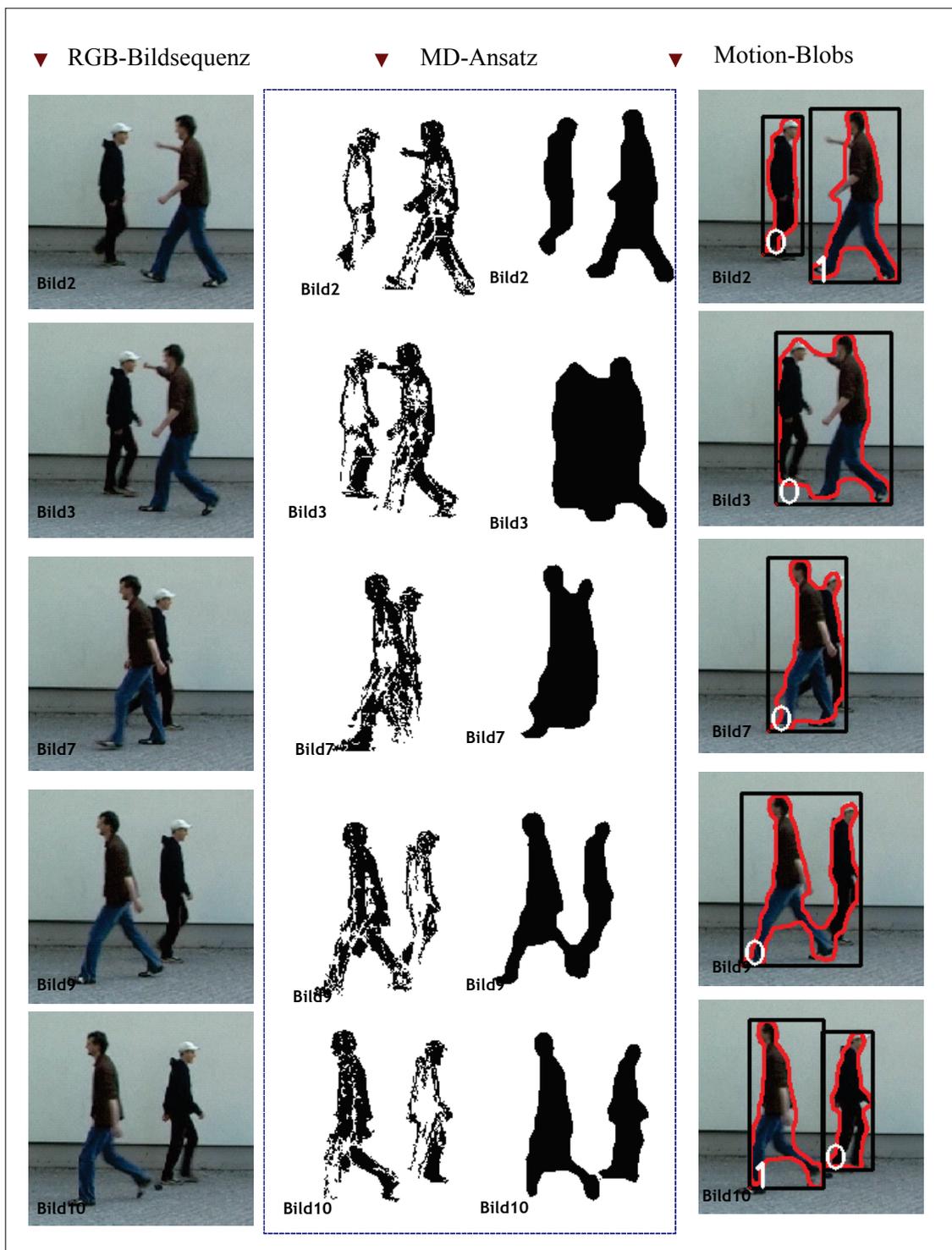
**Abbildung 8-5-1b:** Multi-Objektverfolgung der Szene 1 mit dem PLK- Algorithmus (A) und mit dem merkmalsbasierten Verfahren auf der Grundlage von CSC-Segmenten. Durch die Teilverdeckung des roten PKW mit dem weißen PKW werden beide Verfahren gestört. Die Bewegungsvektoren des roten PKW werden im Laufe der Objektverfolgung durch den weißen PKW abgelenkt und es kommt zu einer Migration der Bewegungsvektoren. Das CSC-Matching zeigt anfänglich wenige falsch zugeordnete Segmente (Bild 2), deren Anzahl im Laufe der Objektverfolgung zunimmt bis die PKW nicht mehr durch ihre CSC-Segmente unterscheidbar sind.



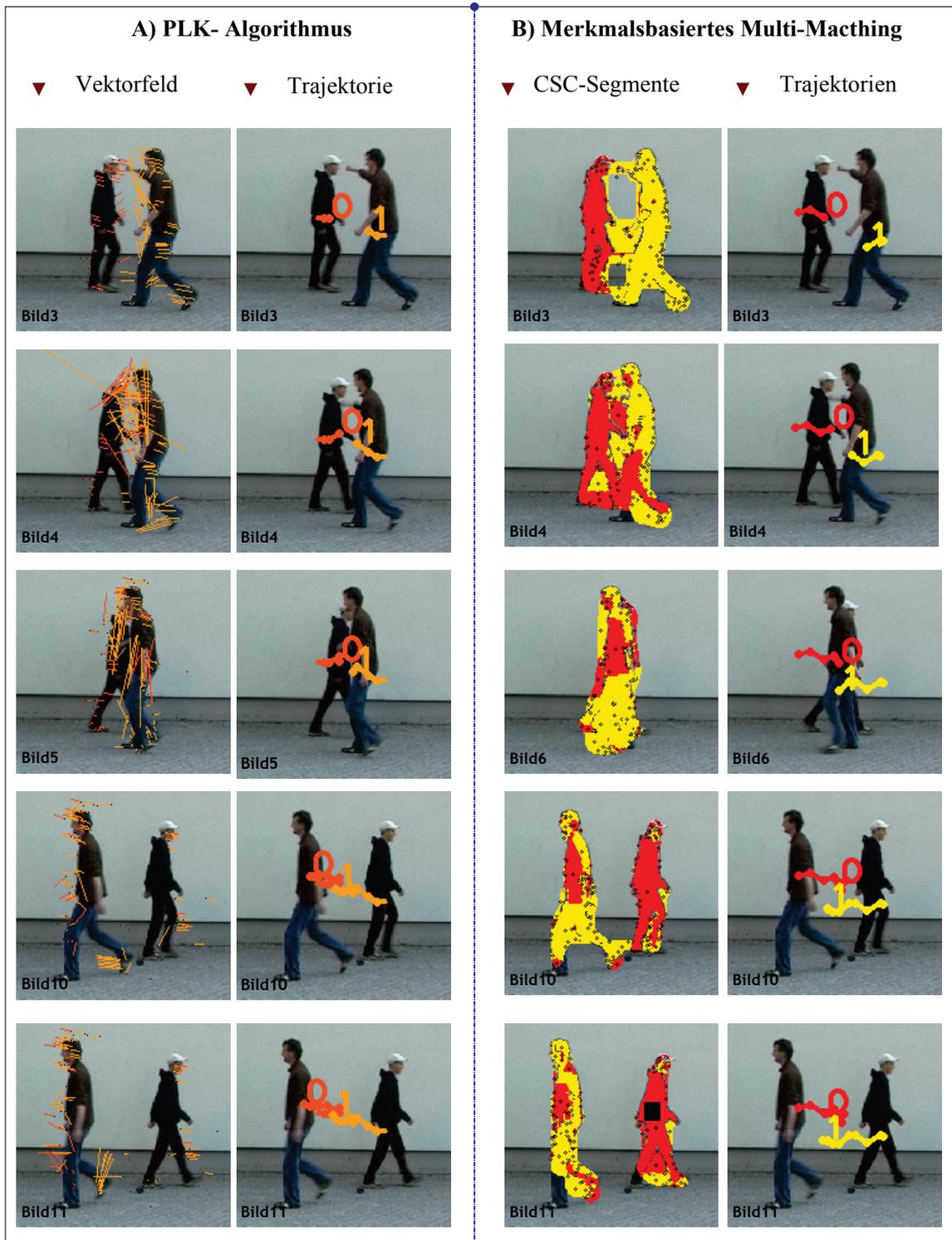
**Abbildung 8-5-1c:** Multi-Objektverfolgung der Szene 1 mit der vorgeschlagenen hybriden Systemstruktur auf der Basis von Partikelfiltern. Durch die Datenaufbereitung werden Merkmale extrahiert, die keine explizite Repräsentation der Region benötigen und sich direkt aus dem MB berechnen. Hier sind die Verschiebungsvektoren, die umgebenden und prädiktieren Ellipsen dargestellt.

Im Bild 3 verschmelzen die Trajektorien, die die PKW repräsentieren und es wird die Trajektorie 2 erzeugt, mit dem Vermerk, dass 0 und 1 Elterntrajektorien von 2 sind. Im Bild 7 wird die Konfliktsituation erfolgreich gelöst, indem die Trajektorien 0 und 1 wieder korrekt den separat Segmentierten PKW zugeordnet wird.

## 8.5.2 Szene 2

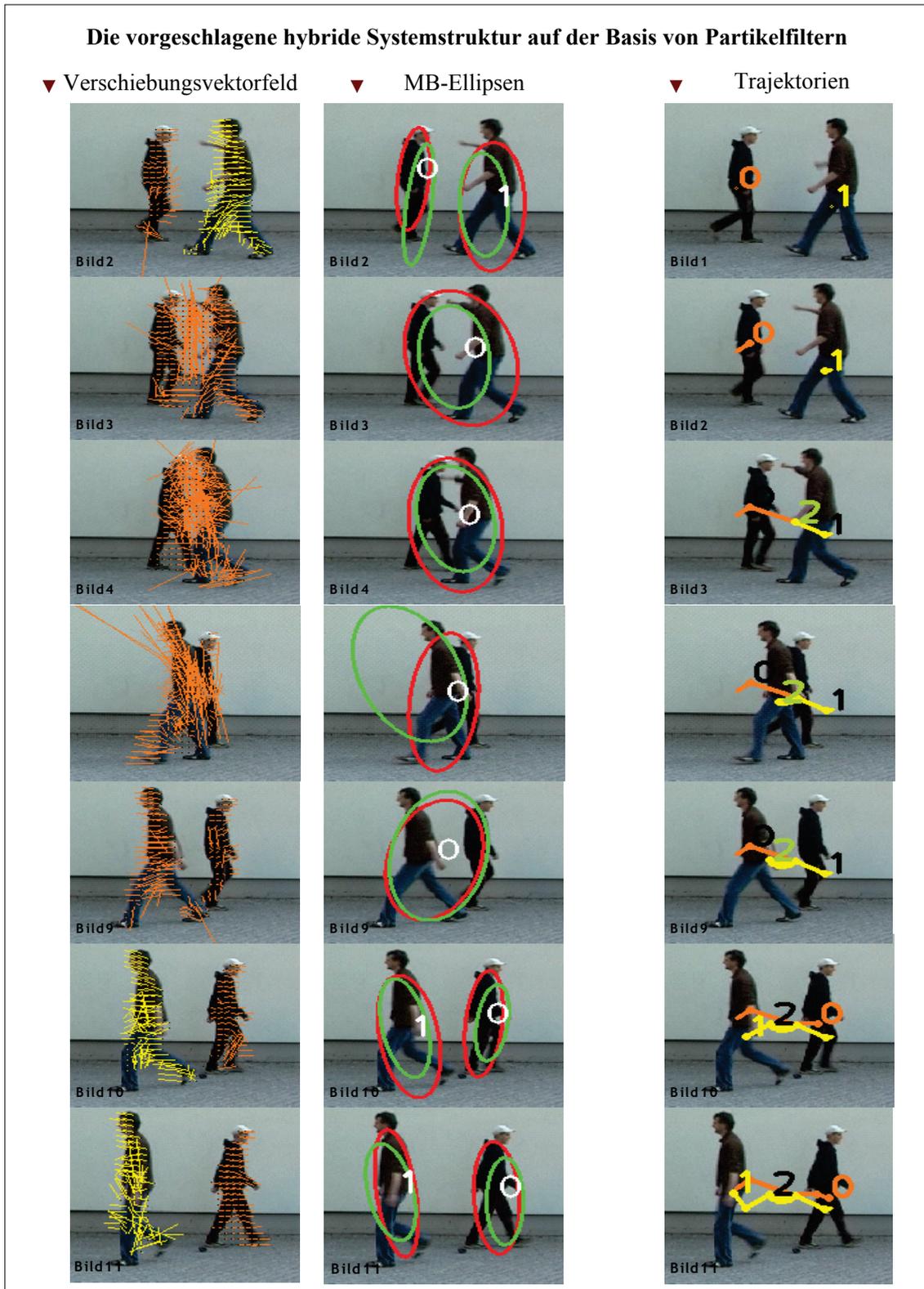


**Abbildung 8-5-2a:** Verlauf der MB-Extrahierung durch den vorgeschlagenen MD-Ansatz in Verbindung mit dem Nachverarbeitungsalgorithmus für die Szene 2. Die Konfliktsituation in dieser Szene besteht darin, dass im Bild 3 die MB beider Personen zu einem MB verschmelzen, der dann im Bild 10 wieder getrennt werden muss.



**Abbildung 8-5-2b:** Multi-Objektverfolgung der Szene 2 mit dem PLK- Algorithmus (A) und mit dem merkmalsbasierten Verfahren auf der Grundlage von CSC-Segmenten. Durch die totale Verdeckung der hinteren Personen durch die vordere Person liefern die beiden Verfahren nicht zuverlässige Bewegungsvektoren.

Die Bewegungsvektoren der Personen werden im Laufe der Objektverfolgung durch die gegenseitige Beeinflussung abgelenkt und es kommt zu einer Migration der Bewegungsvektoren. Das CSC-Matching zeigt anfänglich wenige falsch zugeordnete Segmente (Bild 3), deren Anzahl im Laufe der Objektverfolgung zunimmt bis die Personen nicht mehr durch ihre CSC-Segmente unterscheidbar sind.



**Abbildung 8-5-2c:** Multi-Objektverfolgung der Szene 2 mit der vorgeschlagenen hybriden Systemstruktur auf der Basis von Partikelfiltern. Durch die Datenaufbereitung werden Merkmale extrahiert, die keine explizite Repräsentation der Region benötigen und sich direkt aus dem MB berechnen. Hier sind die Verschiebungsvektoren, die umgebenden und prädiktierten Ellipsen dargestellt.

Im Bild 3 verschmelzen die Trajektorien 0 und 1, der Personen und erzeugen die Trajektorie 2. Im Bild 7 wird die Konfliktsituation erfolgreich gelöst, indem die Trajektorien 0 und 1 wieder korrekt der separat segmentierten Person zugeordnet werden.

## 8.6 Beispielsequenzen: Gehäufte Konfliktsituationen

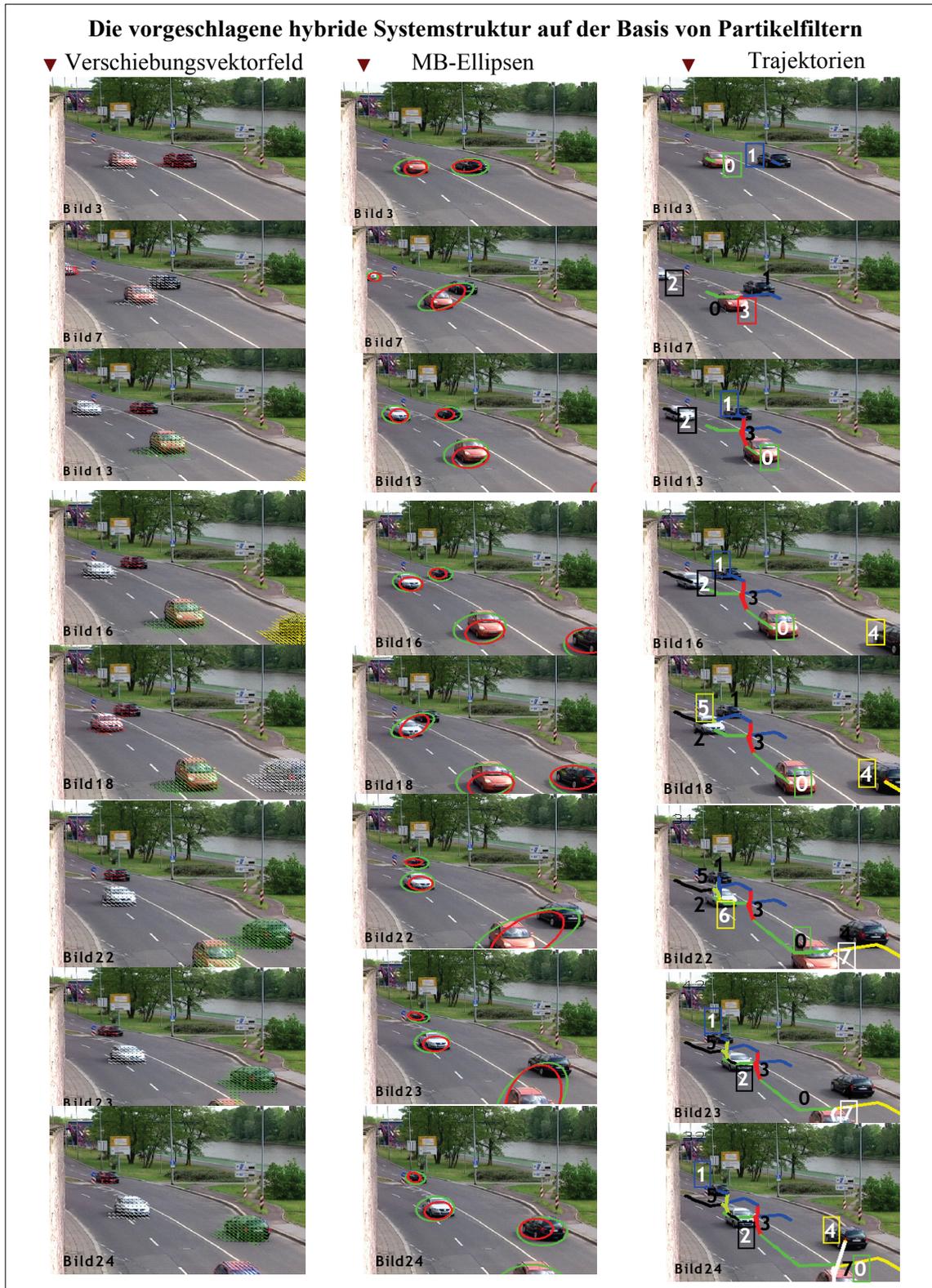
Die folgenden Abbildungen enthalten Ausschnitte aus mehreren Bildsequenzen, die die Aufgabe haben, die Funktion des Systems im Kapitel 5 unter dem Einfluss von nicht-kooperativen Situationen zu verdeutlichen.

### 8.6.1 Szene 3



**Abbildung 8-6-1a:** Die Konfliktsituationen in dieser Szene bestehen darin, dass die MB der entgegenkommenden PKW verschmelzen und wieder getrennt werden müssen. Multi-Objektverfolgung der Szene 3 mit dem PLK-Algorithmus (A) und mit dem merkmalsbasierten Verfahren auf der Grundlage von CSC-Segmenten.

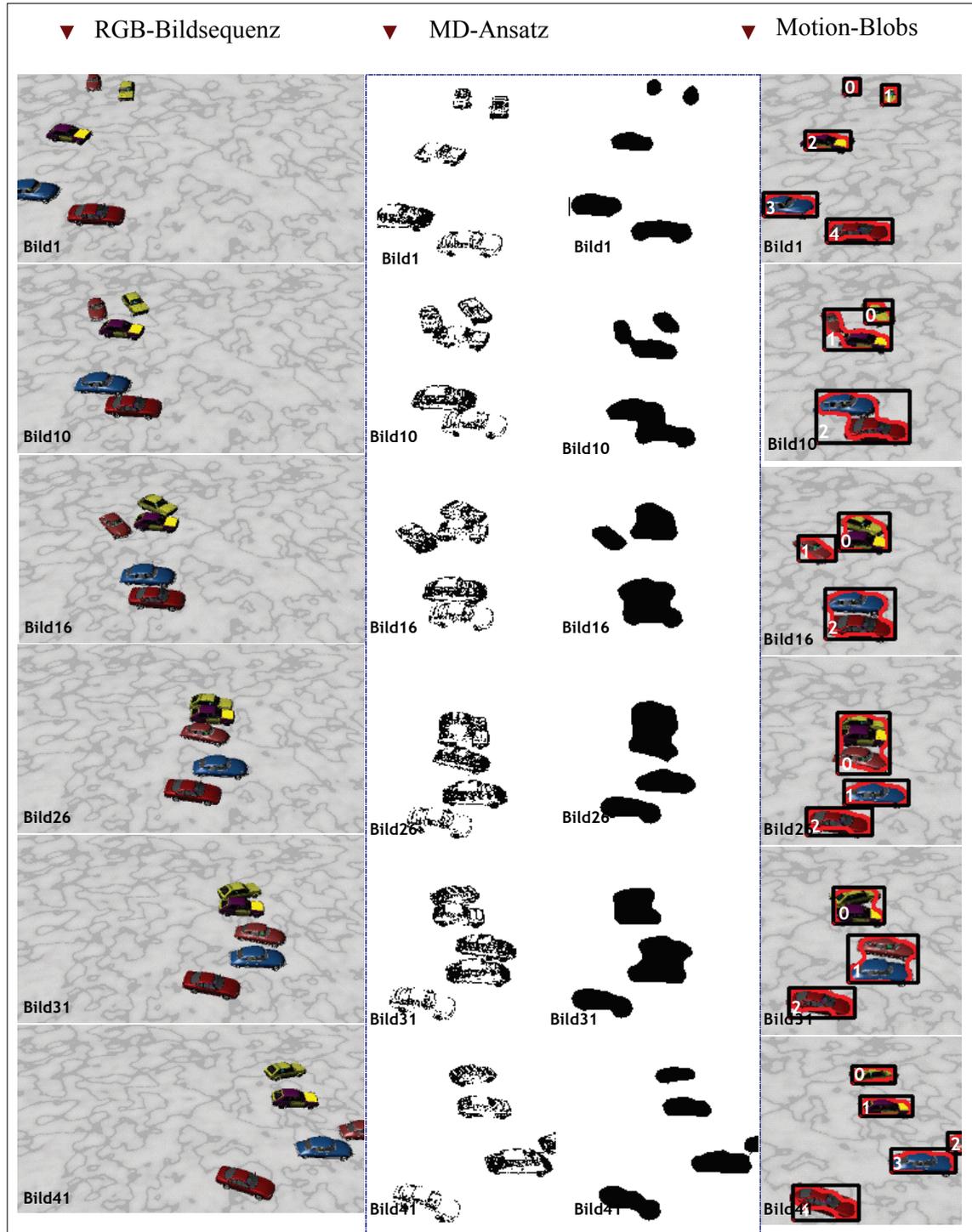
Durch die sehr geringe Verdeckung der PKW wird das PLK-Verfahren nicht gestört. Es werden einige wenige Bewegungsvektoren abgelenkt, was die Objektverfolgung nicht beeinflusst (Bild 7). Das CSC-Matching zeigt zu Beginn der Verschmelzung gute Resultate auch hier werden ab Anfang sehr wenige falsch zugeordnete Segmente (Bild 7) erkannt, deren Anzahl im Laufe der Objektverfolgung nicht so drastisch, wie in Szene 2 zunimmt (Bild 18). Es kommt aber dennoch zu einer Segmentvermischung, so dass der schwarze PKW nicht mehr durch seine CSC-Segmente zu lokalisieren ist.



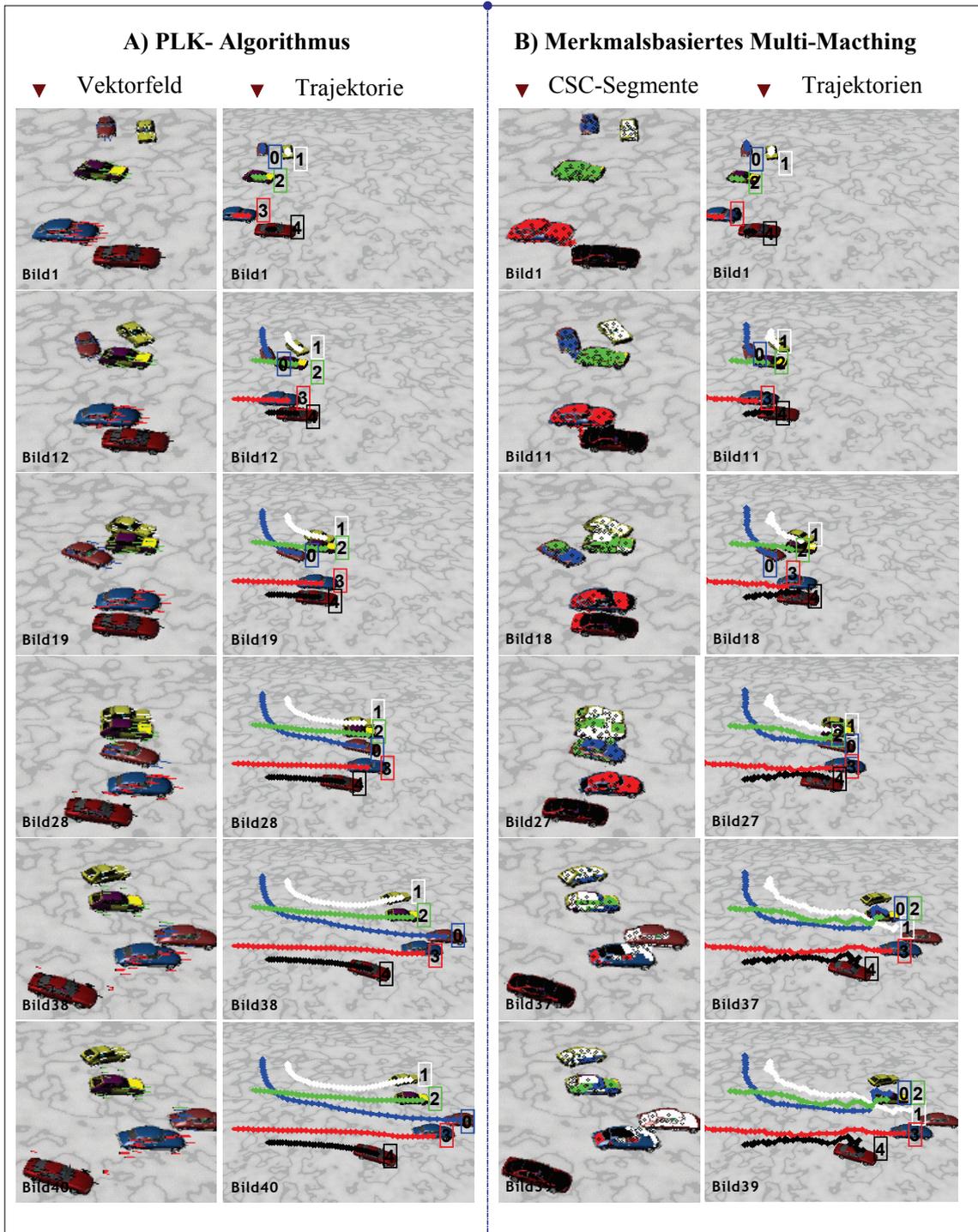
**Abbildung 8-6-1b:** Multi-Objektverfolgung der Szene 3 mit der vorgeschlagenen hybriden Systemstruktur auf der Basis von Partikelfiltern. Durch die Datenaufbereitung werden Merkmale dafür extrahiert, die keine explizite Repräsentation der Region benötigen und sich direkt aus dem MB berechnen. Hier sind die Verschiebungsvektoren, die umgebenden und prädizierten Ellipsen dargestellt.

Es kommt zu der gehäuft verschmelzen in dem zwei PKW in einem MB enthalten sind (Bild 7/22). Alle Konfliktsituation konnten erfolgreich gelöst werden, so dass die PKW ihrer ursprünglichen Trajektorie im Bild 24 zugeordnet sind.

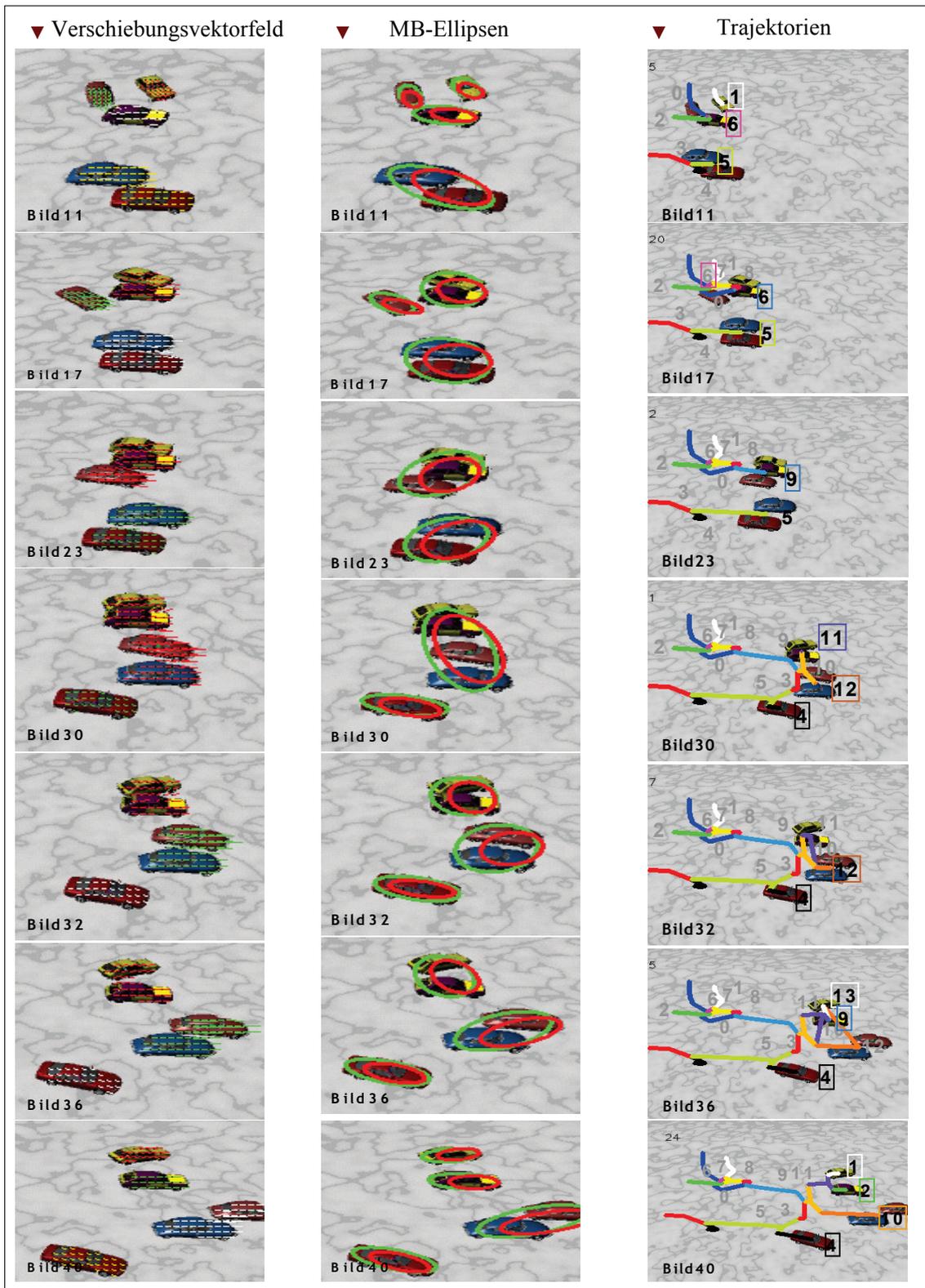
## 8.6.2 Szene 4



**Abbildung 8-6-2a:** Verlauf der MB-Extrahierung durch den vorgeschlagenen MD-Ansatz in Verbindung mit dem Nachverarbeitungsalgorithmus für die Szene 4. Die Konfliktsituationen in dieser Szene bestehen aus komplexen Verschmelzungs- und Teilungsereignissen. Dabei sind in einem MB bis zu drei PKW enthalten (Bild 26), die sich gegenseitig teilweise verdecken. Um die Komplexität und Auflösbarkeit der PKW in einer Bildsequenz zu erhalten wurde diese Szene künstlich mit dem 3D-Animationsprogramm Maya<sup>TM</sup> von Alias erzeugt.



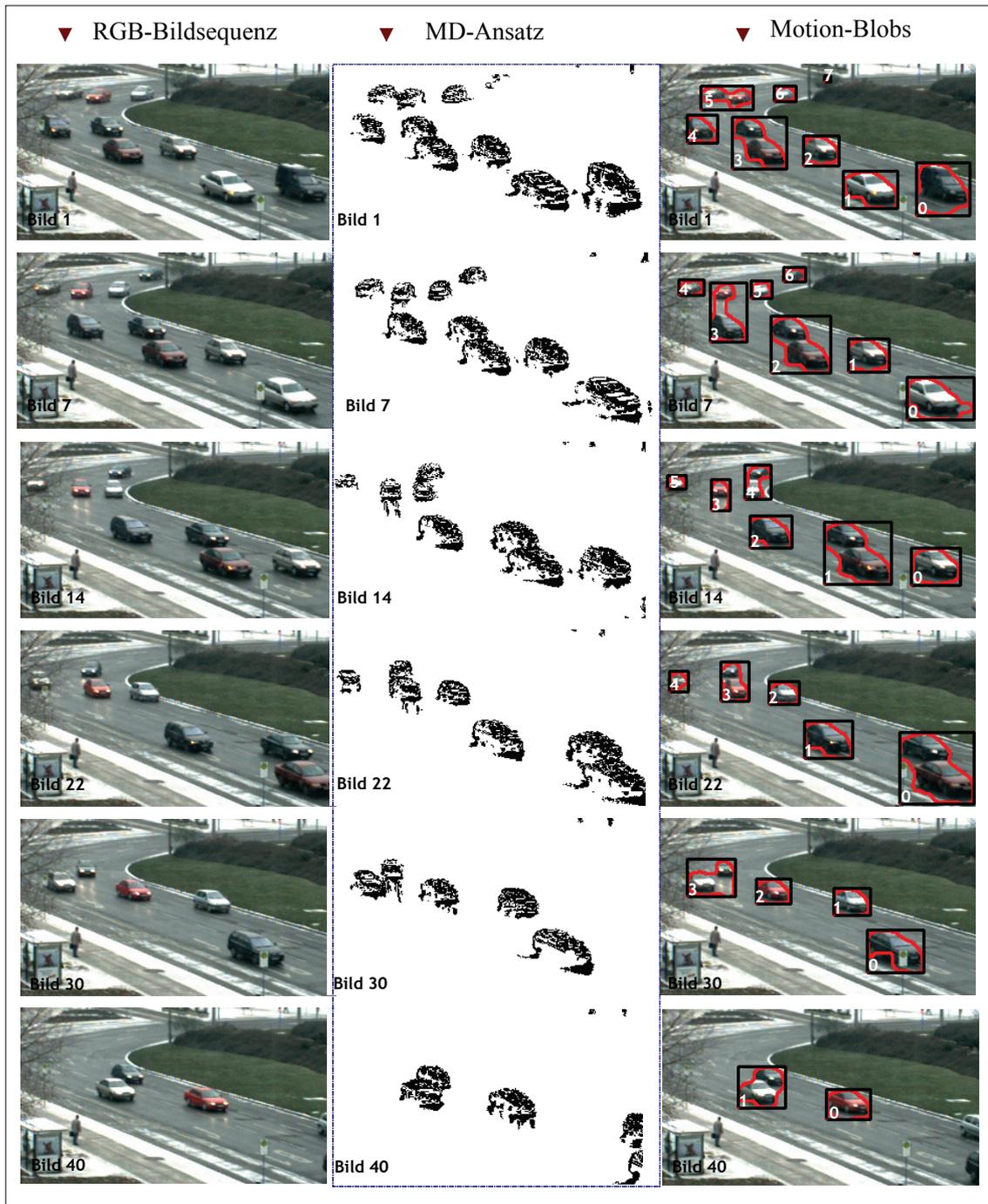
**Abbildung 8-6-2b:** Multi-Objektverfolgung der Szene 4 mit dem PLK- Algorithmus (A) und mit dem merkmalsbasierten Verfahren auf der Grundlage von CSC-Segmenten. Durch die sehr geringe Verdeckung und die hohe Texturierung der PKW erzeugt das PLK-Verfahren gute Ergebnisse. Es werden nur wenige Bewegungsvektoren abgelenkt, die die Objektverfolgung nicht beeinflusst haben (Bild 38). Das CSC-Matching zeigt zu Beginn der Verschmelzung gute Resultate (Bild 11). Die PKW können durch die CSC-Segmente gut getrennt werden. Im Laufe der Objektverfolgung kommt es aber dennoch zu einer Segmentvermischung, die Fehler verursacht.



**Abbildung 8-6-2c:** Multi-Objektverfolgung mit der vorgeschlagenen hybriden Systemstruktur auf der Basis von Partikelfiltern. Durch die Datenaufbereitung werden Merkmale extrahiert. Hier sind die Verschiebungsvektoren, die umgebenden und prädiktierten Ellipsen dargestellt.

Durch die Verschmelzung entsteht eine Vielzahl von Trajektorien. Im Bild 32/36 ist zu erkennen, dass die Bewegungsverfolgung teilweise fehlschlägt. Das Gewinnerpartikel entspricht nicht den realen Gegebenheiten. Jedoch sind in der Partikelmenge, Partikel mit korrekten Zuständen enthalten. So dass diese Störung der Bewegungsprädiktion durch den Partikelfilter im Verlauf der Verfolgung herausgefiltert werden kann (Bild 40). Die Trajektorie 10 beinhaltet als Elterntrajektorien Trajektorie 3 und 0.

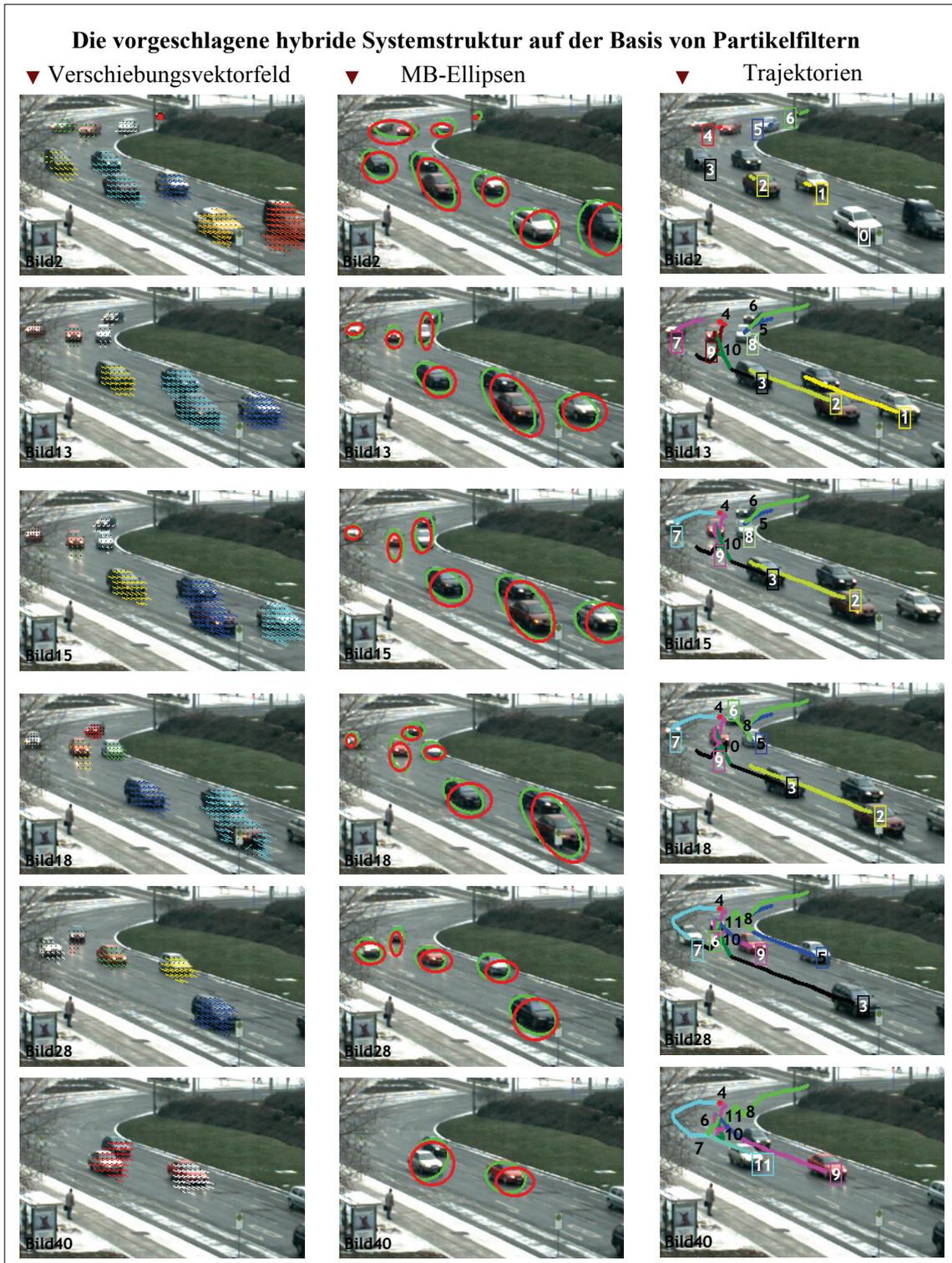
## 8.6.3 Szene 5



**Abbildung 8-6-3a:** Verlauf der MB-Extrahierung durch den vorgeschlagenen MD-Ansatz in Verbindung mit dem Nachverarbeitungsalgorithmus für die Szene 2. Die Konfliktsituationen in dieser Szene bestehen aus gehäuften Verschmelzungs- und Teilungsereignissen. Es treten zusätzlich Teilverdeckungen durch einen Baum im Vordergrund auf.



**Abbildung 8-6-3b:** Multi-Objektverfolgung der Szene 5 mit dem PLK- Algorithmus (A) und mit dem merkmalsbasierten Verfahren auf der Grundlage von CSC-Segmenten. Durch die sehr geringe Verdeckung und die hohe Textuierung der PKW erzeugt das PLK-Verfahren gute Ergebnisse. Es werden jedoch durch die Teilverdeckung durch den Baum Bewegungsvektoren abgelenkt (Bild 22), so dass sie eine Bewegung von null erhalten. Das merkmalsbasierte Verfahren liefert robuste Resultate, die die Bewegungstrajektorien die tatsächliche Bewegung beschreiben. Die PKW können gut durch die CSC-Segmente getrennt werden (Bild 8/23) und die Trajektorien bleiben stabil nur gegen Ende vermischen sich einige CSC-Segmente, was die Bewegungsverfolgung jedoch nicht beeinflusst.



**Abbildung 8-6-3c:** Multi-Objektverfolgung der Szene 5 mit der vorgeschlagenen hybriden Systemstruktur auf der Basis von Partikelfiltern. Durch die Datenaufbereitung werden Merkmale extrahiert, die keine explizite Repräsentation der Region benötigen und sich direkt aus dem MB berechnen. Hier sind die Verschiebungsvektoren, die umgebenden und prädiktierten Ellipsen dargestellt.

Im Bild 12 wird durch ein ungenaues Bewegungsmodell die Trajektorie 4 nicht richtig mit dem MB der beiden PKW verbunden und so die Trajektorie 10 falsch erzeugt. Diese Störung wird im Verlauf der Verfolgung durch den Partikelfilter herausgefiltert und die Multi-Objektverfolgung mit dem Partikelfiltersystem erzeugt korrekte Trajektorien.



### 9. Literaturverzeichnis

- [Ahmad 07a] Ahmad, M.; Al-Hamadi, A.; Krell, G.; and Michaelis, B.: *Very Low Bit Rate Video CODEC Based on Wavelet Filters*, International Conference on Computational Science and Its Applications, 2007, Malaysia, (published by IEEE-CS), pp. 65-71.
- [Ahmad 07b] Ahmad, M.; Al-Hamadi, A.; Krell, G.; and Michaelis, B.: *Enhancing the visual quality in hybrid filters wavelet-based low bit-rate video codec*. In Proceedings of 6th International Conference of Measurement of Audio and Video Quality in Networks- MESAQIN 2007, Prague, CD, ISBN: [978-80-01-03734-8].
- [Ahmad 08] Ahmad, M.; Krell, G.; Al-Hamadi, A.; Hadhoud, M.; Michaelis, B.: *Spatially scalable wavelet-based intra-frame video codec using subband interpolation*. 4th International Symposium on Image/Video Communications (ISIVC), 2008 Spain, pp 55-60.
- [Al-Hamadi 99] Al-Hamadi, A.; Mecke, R.; Michaelis, B.: *Verbesserte Bewegungsbestimmung durch adaptive Farbraumtransformation*; Proc. 5th German Workshop on Colour Image Processing; Ilmenau, 1999, pp. 75-82; [ISSN 1432-3346].
- [Al-Hamadi 00a] Al-Hamadi, A.; Michaelis, B.; Mecke, R.: *Verbesserte Störsicherheit bei der Bewegungsbestimmung durch adaptive Farbraumtransformation*; Proceedings in Artificial Intelligente 9; Dynamische Perzeption; Ulm, Nov. 2000; infix; pp.167-170.
- [Al-Hamadi 00b] Al-Hamadi, A.; Hofmeister, H.; Michaelis, B.; Brückner, B.; Mecke, R.: *Automatisierte Objektselektion und -verfolgung in einem adaptiven Farbraum*; Proc. 6th German Workshop on Colour Image Processing; 2000, GFaI Berlin; pp.27-34.
- [Al-Hamadi 01a] Al-Hamadi A.: *Verbesserte Störsicherheit bei der Bewegungsanalyse in monokularen Farbbildsequenzen durch adaptive Farbraumtransformation*, Dissertation; 166 Seiten, Universität Magdeburg; September 2001.
- [Al-Hamadi 01b] Al-Hamadi, A.; Michaelis, B.; Mecke, R.: *Bewegungsanalyse in Farbbildsequenzen unter Berücksichtigung von nichtkooperativen Messsituationen*; Fachtagung Informationstechnik, März 2001, Magdeburg, [ISBN 3-930385-29-5], pp. 147-152.
- [Al-Hamadi 02a] Al-Hamadi, A.; Michaelis, B.: *Adaptive Tracking of Moving Objects in Colour Video Sequences*. 12<sup>th</sup> Portuguese Conference on Pattern Recognition, RECPAD 2002, Aveiro, Portugal, P.05; [ISBN 972-780-067-9].
- [Al-Hamadi 02b] Al-Hamadi, A.; Michaelis, B.: *Fulfilment of the Demand for Robustness and Flexibility in the Motion Analysis Systems by Adaptive Color Space Transformation*. First European Conference on Color in Graphics, Imaging and Vision, CGIV 2002, Poitiers, France, April, 2002; pp. 540-544.
- [Al-Hamadi 02c] Al-Hamadi, A.; Michaelis, B.: *Intensity-based method for tracking of objects in colour video sequences under the influence of non-cooperative situations*.

- IASTED, International Conference on Signal Processing, Pattern Recognition and Applications; Crete, Greece, June, 2002; pp. 62-67.
- [Al-Hamadi 02d] Al-Hamadi, A.; Michaelis, B.: *Robust Method for Detection and Motion Estimation of Objects in Colour Video Sequences using Multi-Channel-Matching*. IASTED, International Conference on Visualization, Imaging and Image Processing; Malaga, Spain, 2002; pp. 633-638.
- [Al-Hamadi 03a] Al-Hamadi, A.; Michaelis, B.; Niese, R.: *Towards robust automatic segmentation and tracking analysis of objects in video sequences*. 3rd IEEE-EURASIP; 3rd International Symposium on Image and Signal Processing and Analysis ISPA 03, 2003, Rome, Italy, pp. 645-651.
- [Al-Hamadi 03b] Al-Hamadi, A.; Niese, R. ; Michaelis, B: *A robust approach for contour extraction and tracking of moving objects in video sequences*; The IASTED International Conference on Signal Processing, Pattern Recognition, and Applications; 2003, Rhodes/Greece; pp. 336-341.
- [Al-Hamadi 03d] Al-Hamadi, A.; Niese, R.; Michaelis, B.: *Hierarchische merkmalsbasierte Bewegungsanalysemethode für Videosequenzen*. Proc. 9th German Workshop on Color Image Processing, FarbBV '03, Esslingen, 2003, pp. 27-34.
- [Al-Hamadi 03e] Al-Hamadi, A.; Niese R.; Michaelis, B.: *Another paradigm for the solution of the correspondence problem in motion analysis*; CIARP '2003 8th Iberoamerican Conference on Pattern Recognition; Springer Verlag, Havana; pp.95-103.
- [Al-Hamadi 03f] Al-Hamadi, A.; Michaelis, B.; Mecke R.: *Improvement of fail safe characteristics in motion analysis using colour information*; CIARP '2003 8th Iberoamerican Conference on Pattern Recognition; Springer Verlag, Havana; pp.104-112.
- [Al-Hamadi 03g] Al-Hamadi, A.; Michaelis, B. : *Object-adapted intensity-based motion analysis in video*; ICAPR '2003 Fifth International Conference on Advances in Pattern Recognition; Dec, 2003; Calcutta, India; pp. 410-413.
- [Al-Hamadi 04a] Al-Hamadi, A.; Niese, R.; Michaelis, B: *Feature-based Correspondence Analysis in Color Image Sequences*; International Conference on Computer Vision and Graphics 2004; September, Warsaw, Poland; Springer Verlag, pp. 179-187.
- [Al-Hamadi 04b] Al-Hamadi, A.; Niese R. ; Michaelis, B: *Multi-Object Tracking In Video Using A Trisection Paradigm*; 7<sup>th</sup> International Conference on Pattern Recognition and Image Analysis: New information technologies; PRIA-7-2004; St. Petersburg; Russian Federation; pp. 599-602.
- [Al-Hamadi 04c] Al-Hamadi, A.; Niese R. ; Michaelis, B: *A Fast and Robust Approach for the Segmentation of Moving objects*; International Conference on Computer Vision and Graphics 2004; September 22-24, Warsaw, Poland; pp. 13-20.
- [Al-Hamadi 05a] Al-Hamadi, A.; Niese, R. and Michaelis, B.: *Multi-Object Tracking in Video Using a Trisection Paradigm*; International journal "Pattern Recognition and Image Analysis": 2005, Vol. 15, No 2; 2005, pp.492-495.
- [Al-Hamadi 05b] Al-Hamadi, Ayoub, Michaelis, Bernd: *Suppression of Cast Shadows for Improved Object Discrimination in Video Sequences*; IEEE International Conference on Computational Intelligence for Modelling Control and Automation-CIMCA'2005, Vienna, pp.518-524.

## 9. LITERATURVERZEICHNIS

---

- [Al-Hamadi 05c] Al-Hamadi, Ayoub; Michaelis, Bernd: *Robust Multi-Object Tracking Under a Wide Range of Real-World Conditions*; IEEE International Conference on Computational Intelligence for Modelling Control and Automation -CIMCA'2005, Vienna, Austria, pp. 815-821.
- [Al-Hamadi 05d] Al-Hamadi, Ayoub K.; Michaelis, Bernd; Kube, Karsten: *A Shadow Detection Filter on the Basis of Colour-Structure-Code Algorithm*, International Congress for global Science and Technology, Conference on Graphics, Vision and Image Processing (GVIP-05), Cairo 2005, ISSN [1687-398X], pp. 428-433.
- [Al-Hamadi 05e] Al-Hamadi, Ayoub K.; Michaelis, Bernd: *Hierarchical Feature-based Paradigm for Multi-object Tracking*, International Congress for Global Science and Technology, Conference on Graphics, Conference on Graphics, Vision and Image Processing (GVIP-05), Cairo 2005, pp. 434-439.
- [Al-Hamadi 06a] Al-Hamadi, A.; Niese, R. ; Michaelis, B: *A Fast and Robust Approach for the Segmentation of moving Objects*; Computational Imaging and Vision; Book of Computer Vision and Graphics, Springer Verlag, pp. 13-20, 2006, [ISBN 10-1-4020-4179 (e-book)].
- [Al-Hamadi 06b] Al-Hamadi, A.; and Michaelis B.: *An intelligent Paradigm for Multi-Objects Tracking in Crowded Environment*, Journal of Digital Information Management (JDIM) , Volume 4, Number 3, September 2006, pp. 183-190.
- [Al-Hamadi 06c] Al-Hamadi, A.; Michaelis, B.: *A Shadow Detection Filter on the Basis of Colour-Structure-Code Algorithm*, International Journal on Graphics, Vision and Image Processing, Special issue articles on: "Object Tracking and Recognition", ISSN [1687-398X], March 2006, pp. 51-57.
- [Al-Hamadi 06d] Al-Hamadi, A.; Michaelis, B.: *Hierarchical Feature-based Paradigm for Multi-object Tracking*, International Journal on Graphics, Vision and Image Processing, Special issue articles on: "Object Tracking and Recognition", ISSN [1687-398X], March 2006, pp. 39-44, 2006.
- [Al-Hamadi 06e] Al-Hamadi, A.; Niese, R.; Panning, A.; and Michaelis, B.: *Toward Robust Face Analysis Method of Non-Cooperative Persons in Stereo Color Image Sequences*, Journal of Machine GRAPHICS and VISION, Vol. 15, No.3/ 4, 2006, pp. 245-254.
- [Al-Hamadi 06f] Al-Hamadi, A.; Niesem R.; Michaelis, B: *Feature-based Correspondence Analysis in Color Image Sequences*; Computational Imaging and Vision; Book of Computer Vision and Graphics, Springer Verlag, pp. 179-187, 2006, [ISBN 10-1-4020-4179 (e-book)].
- [Al-Hamadi 06g] Al-Hamadi, A.;Panning, A.; Niese, R.; and Michaelis, B.: *A Model-based Image Analysis Method for Extraction and Tracking of Facial Features in Video Sequences*, 4th International Multi- Conference on Computer Science and Information Technology CSIT 2006, Amman, Vol. (3), pp. 499-509.
- [Al-Hamadi 06h] Al-Hamadi, A.; Michaelis, B.: *An Intelligent Paradigm for Multi-Objects Tracking in Crowded Environment*, The 4th International Multi-Conference on Computer Science and Information Technology CSIT 2006, , Amman, Vol. (3), pp. 488-498.

## 9. LITERATURVERZEICHNIS

---

- [Al-Hamadi 07a] Al-Hamadi, A.; Homberg, U.; Niese, R.; and Michaelis, B.: *Efficient Tracking Approach of Multiple Interacting Objects Using Data Association*. 22<sup>nd</sup> Int. Symp. on Computer and Information Sciences, IEEE, 2007, pp. 1-6 Ankara, (CD).
- [Al-Hamadi 07b] Al-Hamadi, A.; Niese, R.; Michaelis, B.: *Multi-Object Tracking In Video Using A Trisection Paradigm*; Special issues of the international journal "Pattern Recognition and Image Analysis", Vol. 17, No. 4, 2007, pp. 493-507.
- [Al-Hamadi 08a] Al-Hamadi, A.; Pathan, S.; Michaelis, B.: *Multi-Object Tracking Based on Particle Filter and Data Association in Color Image Sequences*; International Conference on Computer Vision and Graphics 2008; Nov. 10-12, Warsaw, Poland; Springer Verlag. pp. 133-143.
- [Al-Hamadi 08b] Al-Hamadi, A.; Niese R.; Pathan S.; Michaelis, B.: *Geometric and Optical Flow Based Method for Facial Expression Recognition in Color Image Sequences*; International Conference on Computer Vision and Graphics 2008; Nov. 10-12, Warsaw, Poland; Springer Verlag. pp. 228-239.
- [Arts 91] Arts, E; Korst J.: *Simulated Annealing and Boltzmann Machines. A Stochastic Approach to Combinatorial Optimization and Neural Computing*. Wiley 1991.
- [Aschwanden 93] Aschwanden, P. F.: *Experimenteller Vergleich von Korrelationskriterien in der Bildanalyse*; 1. Aufl.; Hartung-Gorre Verlag, 1993. Dissertation
- [Aziz 99] Aziz, A. M.: *New Data Fusion Algorithms for Distributed Multi-Sensor Multi-Target Environments*. Naval Postgraduate School, Monterey, Dissertation, 1999.
- [Bandopadhyay 86] Bandopadhyay, A.; Dutta, R.: *Measuring Image Motion in Dynamic Images*. Proc. Workshop on Motion: Representation and Analysis, Washington: IEEE Computer Society Press 1986; pp. 67-72.
- [Barron 94] Barron, J.L.; Fleet, D. J.; Beauchemin, S.: *Performance of Optical Flow Techniques*, 1994.
- [Bar-Shalom 95] Bar-Shalom, Y.; Li, X.: *Multitarget-Multisensor Tracking: Principles and Techniques*. 3. Ausgabe, Storrs: YBS, 1995.
- [Bertozzi 00] Bertozzi, M; Broggi, A; Fascioli, A.: *Vision-based intelligent vehicles: State of the art and perspectives*, Elsevier 2000, Robotics and Autonomous Systems 32 1–16.
- [Bevilacqua 02] Bevilacqua, A. *Effective Object Segmentation in a Traffic Monitoring Application*, ARCES – DEIS (Department of Electronics, Computer Science and Systems) University of Bologna, Italy, 2002.
- [Bevilacqua 03] Bevilacqua, A.: *Effective Shadow Detection in Traffic Monitoring Applications*, Journal of WSCG, Vol.11, No. 1, 2003, Plzen, Czech Republic.
- [Blacke 93b] Blake, A.; Curwen, R.; Zisserman, A: *A framework for spatio-temporal control in the tracking of visual contours*; International Journal of CV; 11(2):127-145; 1993.
- [Blackman 99] Blackman, S.; Popoli, R.: *Design and Analysis of Modern Tracking Systems*. Reihe: Artech House Radar Library, Boston: Artech House, 1999.
- [Blake 93a] Blake, A.; Curwen, R.; Zisserman, A.: *Affine-invariant contour tracking with automatic control of spatiotemporal scale*; Proc. Fourth Int. Conference on Computer vision, Berlin, May 1993, IEEE Computer Society Press, pp. 66-75.

## 9. LITERATURVERZEICHNIS

---

- [Blake 98] Blake, A.; Isard, M.: *Active Contours*, Springer Verlag London, 1998, ISBN 3-540-76217-5.
- [Böhme 93] Böhme, G.: *Fuzzy-Logic, Einführung in die algebraischen und logischen Grundlagen*, Springer-Verlag 1993.pp.43-60.
- [Borgefors 88] Borgefors, G.: *Hierarchical Chamfer Matching: A Parametric Edge Matching Algorithm*; IEEE Transactions on Pattern Analysis and Machine Intelligence; Vol. 10, No. 6, Nov. 1988. S. 849-865.
- [Bothe 93] Bothe, H.: *Fuzzy Logic, Einführung in Theorie und Anwendung*, Springer-Verlag 1993.
- [Bouguet 00] Bouguet, J-Y.: *Pyramidal Implementation of the Lucas Kanade Feature Tracker. Description of the algorithm*. Intel Corporation Microprocessor Research Labs. 2000.
- [Brandner 02] Brandner, M.; Prinz, A.: *Real-Time Tracking of complex objects using dynamic interpretation tree*; DAGM 2002; pp. 9-19; 2002; Springer Verlag.
- [Bronstein 01] Bronstein, I.N.; Semendjajew, K.A.; Musiol, G.; Mühling, H.: *Taschenbuch der Mathematik*, Verlag Harri Deutsch, 2001.
- [Brox 04] Brox, T.; Bruhn, A.; Papenber, N.; Weickert J.: *High Accuracy Optical Flow Estimation Based on a Theory for Warping*. 8<sup>th</sup> European Conference on Computer Vision. 2004.
- [Bruhn 04] Bruhn, A.; Weickert, J.; Schnörr, C.: *Lucas/Kanade Meets Horn/Schunck: Combining Local and Global Optic*. First online version published in October, 2004.
- [Bruhn 05] Bruhn A.; Weickert J.; Schnörr C.: *Lucas/Kanade Meets Horn/Schunck: Combining Local and Global Optic Flow Methods*. International Journal of Computer Vision. 2005, Heft 61(3), pp.221-231.
- [Calow 03] Calow, R.; Michaelis, B.; Al-Hamadi, A.: *Solutions for model-based analysis of human gait*; 25<sup>th</sup> DAGM Pattern Recognition Symposium, 2003, Magdeburg; pp.540-547; Springer Verlag; LNCS2781.
- [Casella 96] Casella, G.; Robert, C. P.: *Rao-Blackwellisation of Sampling Schemes*. In: Biometrika. 1996, Heft 83(1), pp. 81–94.
- [Chang 03] Chang, Ch.; Ansari, R.: *Kernel Particle Filter: iterative sampling for efficient visual tracking*. In: Proc. of IEEE International Conference of Image Processing (ICIP). Piscataway: 2003, Vol. 3, pp. 977–980
- [Chang 05] Chang, Ch.; Ansari, R.: *Kernel Particle Filter for Visual Tracking*. In: IEEE Signal Processing Letters. Piscataway: IEEE Signal Processing Society, 2005, Vol. 12(3), pp. 242–245.
- [Coifman 98] Coifman, B. (corresponding autor): *A real time computer vision system for vehicle Tracking and traffic surveillance*, submitted for publication in Transportation Research-C, Revised December 1, 1998; [www.ceegs.ohio-state.edu/~coifman/documents/TR-Crw.pdf](http://www.ceegs.ohio-state.edu/~coifman/documents/TR-Crw.pdf) -

## 9. LITERATURVERZEICHNIS

---

- [Corradini 2000] Corradini, A.; Gross, H.M.: *Camera-based Gesture Recognition for Robot Control*. In *Proceedings of IJCNN'2000*, vol. IV, pp. 133-138, July 24-27, Como (Italy), 2000.
- [Cox 94] Cox, I.J.; Hingorani, S. L.: *An Efficient Implementation and Evaluation of Reid's Multiple Hypothesis Tracking Algorithm for Visual Tracking*. In: Proc. of the 12th IAPR International Conference on Pattern Recognition (ICPR). Los Alamitos: IEEE Computer Society Press, 1994, Vol. 1, pp. 437-442.
- [Daniilidies 92] Daniilidis, K.: *Zur Fehlerempfindlichkeit in der Ermittlung von Objektbeschreibungen und relativen Bewegungen aus monokularen Bildfolgen*. Dissertation, Universität Karlsruhe, Fakultät Informatik, Juli 1992.
- [David 03] David, A.; Forsyth; Ponce J.: *Computer Vision: A Modern Approach*; Prentice Hall, 2003.
- [Deriche 90] Deriche, R.; Faugeras, O.: *Tracking line segments*; Image and Vision Computing; 8(4): 261-270, 1990.
- [Dickmanns 92] Dickmanns, E.D; Mysliwetz, B.D.: *Recursive 3-D Road and Relative Ego-State Recognition*. IEEE Trans. On Pattern Analysis and Machine Intelligence, Vol. 14, No. 2, February 1992.
- [Doucet 00] Doucet, A.; de Freitas, N.; Murphy, K. P.; Russell, S. J.: *Rao-Blackwellised Particle Filtering for Dynamic Bayesian Networks*. In: Proc. of the 16th Conference on Uncertainty in Artificial Intelligence. San Francisco: MK Publishers Inc., 2000, pp. 176-183.
- [Doucet 03] Doucet, A.; Godsill, S.; Andrieu, C.: *On sequential Monte Carlo sampling methods for Bayesian filtering*, Statistics and Computing. Volume 10, Issue 3, S.197-208, 2000.
- [Doucet 98] Doucet, A.: *On sequential simulation-based methods for Bayesian filtering*, Dept. Eng., Univ. Cambridge, UK, Tech. Rep., 1998
- [Dror 95] Dror, Itiel E.; Zagaeski, Mark; Moss, Cynthia F.: *Three-Dimensional Target Recognition via Sonar: A Neural Network Model*. Neural Networks. Vol. 8, No. 1, pp. 149-160, 1995.
- [Dubuisson 93] Dubuisson, M.P.; Jain, A.K.: *Object contour extraction using color and motion*. In Proc. of the Conf. on Computer Vision and Pattern Recognition, pp. 471-476, IEEE CSP, New York, 1993
- [Dueck 90] Dueck, G; Scheuer, T.: *Threshold Accepting: A General Purpose: Optimization Algorithm Superior to Simulated Annealing*, Journal of Computational Physics, pp.161-175, 1990.
- [Elmezain 08a] Elmezain, M.; Al-Hamadi, A.; Michaelis, B.: *Real-Time Capable System for Hand Gesture Recognition Using Hidden Markov Models in Stereo Color Image Sequences*. Journal of WSCG, Vol.16, [ISSN: 1213-6972], pp. 65-72, Feb. 2008.
- [Elmezain 08b] Elmezain, M.; Al-Hamadi, A.; Appenrodt, J.; Michaelis, B.: *A Hidden Markov Model-Based Continuous Gesture Recognition System for Hand Motion Trajectory*, 19th International Conference on Pattern Recognition (ICPR 2008), Dec. 2008. – Tampa, Florida, USA. CD.

## 9. LITERATURVERZEICHNIS

---

- [Elmezain 09a] Elmezain, M.; Al-Hamadi, A.; Michaelis, B.: *A Novel System for Automatic Hand Gesture Spotting and Recognition in Stereo Color Image Sequences*; 17-th International Conference on Computer Graphics, Visualization and Computer Vision'2009, February 2 - 5, 2009. - Plzen.
- [Fox 03] Fox, D.; Hightower, J.; Liao, L.; Schulz, D.; Borriello, G.: *Bayesian filtering for location estimation*. In: IEEE Pervasive Computing. 2003, Heft 2(3), S. 24–33.
- [Gnedenko 65] Gnedenko, B. W.; Hans-Joachim Roßberg (Hrsg.): *Lehrbuch der Wahrscheinlichkeitsrechnung*. 4. Ausgabe, Reihe: Mathematische Lehrbücher und Monographien, Berlin: Akademie-Verlag, 1965.
- [Gordon 93] Gordon, N; Salmond, D.;Smith, A. F. M.: *Novel approach to nonlinear and non-Gaussian Bayesian state estimation*, Proc. Inst. Elect.Eng., F, Volume. 140, S. 107–113, 1993.
- [Guse 92] Guse, W.: *Objektorientierte Bewegungsanalyse in Bildfolgen*, VDI-Fortschrittberichte, Reihe 10, Nr. 223. Düsseldorf: VDI-Verlag, 1992.
- [Halir 98] Halir, R.; Flusser, J.: *Numerically Stable Direct Least Squares Fitting of Ellipses*, 1998.
- [Helmick 00] Helmick, R.: *IMM Estimator with Nearest-Neighbor Joint Probabilistic Data Association*. In: Yaakov Bar-Shalom (Hrsg.), *Multitarget-Multisensor Tracking: Applications and Advances*. Boston: AH Publishers, 2000, Chapter 3, pp. 161–198.
- [Hofmann 03] Hofmann Marc: *Intentionsbasierte maschinelle Interpretation von Benutzeraktionen*, PhD. Thesis 2003
- [Homberg 07] Homberg, U.; *Mutli-Objektverfolgung in Farbbildsequenzen auf der Basis von Partikelfiltern*. Diplomarbeit Universität Magdeburg. 2007.
- [Horn 81] Horn, B. K. P.; Schunck, B. G.: *Determining Optical Flow*; artificial intelligence 17, 1981; S. 185-203.
- [Huwer 98] Huwer, S.; Niemann, H.: *2D-Objekt Tracking Based on Projection-Histograms*. In: Burkhardt, H.; Neumann, B. (Hrsg.): *Computer Vision - ECCV'98*, LNCS, Springer-Verlag 1998, Vol. I, S. 961-976.
- [Isard 96] Isard, M.; Blake, A.: *Visual Tracking by Stochastic Propagation of Conditional Density*. In: Proc. 4<sup>th</sup> European Conference Computer Vision. Cambridge, England, April 1996, pp. 343-356.
- [Isard 98a] Isard ,M.; Blake, A.: *CONDENSATION-Conditional Density Propagation for Visual Tracking*. In: International Journal of Computer. 1998, Heft 29(1), pp. 5-28.
- [Isard 98b] Isard, M.; Blake, A.: *ICondensation: Unifying low-level and high-level tracking in a stochastic framework*. In: Proceedings of the 5<sup>th</sup> European Conference on Computer Vision. LNCS, London: Springer-Verlag, 1998, Vol.1, pp. 893–908.
- [Jähne 97] Jähne, B.: *Digitale Bildverarbeitung*; 4. Auflage, Springer-Verlag, 1997.
- [Jähne 98] Jähne, B.; u. a.: *Study of Dynamical Processes with Tensor-Based Spatiotemporal Image Processing Techniques*; In: 5<sup>th</sup> European Conference on Computer Vision, LNCS 1407, Springer-Verlag, 1998. Vol. II, S. 322-336.
- [Jähne 02] Jähne, B.: *Digitale Bildverarbeitung*. Springer-Verlag, Berlin 2002.

## 9. LITERATURVERZEICHNIS

---

- [Jain 81] Jain, J.R.; Jain, A.K.: *Displacement Measurement and its Application in Interframe Image Coding*, IEEE Transactions on Communications, Vol. Com-29, No. 12, 1981, S. 1799-1806.
- [Kalman 60] Kalman, R. E.: *A New Approach to Linear Filtering and Prediction Problems*, 1960, S. 16-26, In: Sorenson, H. W. (Hrsg.): *Kalman Filtering: Theory and Application*. IEEE Press, 1985.
- [Kalman 85] Kalman, R. E.: *A New Approach to Linear Filtering and Prediction Problems*, 1960, S. 16-26, In: Sorenson, H. W. (Hrsg.): *Kalman Filtering: Theory and Application*. IEEE Press, 1985.
- [Karlsson 01] Karlsson, R.; Gustafsson, F.: *Monte Carlo data association for multiple target tracking*. In: IEE International Seminar on Target Tracking: Algorithms and Applications. London: IEE, 2001, Vol. 1, pp. 13/1–13/5.
- [Kirubarajan 00] Kirubarajan, T.; Bar-Shalom, Y.; Krishna, R. Pattipati: *Multiassignment for Tracking a Large Number of Overlapping Objects*. In: *Multi-target-Multisensor Tracking: Applications and Advances*. Boston: Artech House Publishers, 2000, Kapitel 4, pp. 199–231.
- [Klette 96] Klette, R.; Koschan, A.; Schluens, K.: *Computer Vision*; Vieweg Technik; [ISBN 3-528-06625-3]; 1996
- [Knauer 05] Knauer, U.; Reulke, R.; Meffert, B.: *Fahrzeugdetektion und -erkennung mittels mehrdimensionaler Farbhistogrammanalyse*, Humboldt-Universität Berlin, 2005.
- [Knöppel 01] Knöppel, C.: *Stereobasierte und spurgenaue Erkennung von Straßenfahrzeugen im Rückraum eines Straßenfahrzeuges*, Dissertation. OvG-Universität Magdeburg, Fakultät für Elektrotechnik und Informationstechnik, 2001
- [Koga 81] Koga, T.; Iinuma, K., Hirano A.; Iijima, Y.; Ishiguro, T.: *Motion-Compensated Interframe Coding for Video Conferencing*. In: NTC 81, Proc. ; New Orleans, LA, Dec. 1981, S. G5.3.1-G5.3.5
- [Köhler 05] Köhler, B.-U.: *Konzepte der statistischen Signalverarbeitung*; Berlin: Springer, 2005.
- [Kories 84] Kories, R.; Zimmermann, G.: *Eine Familie von Bildmerkmalen für die Bewegungsbestimmung in Bildfolgen*. 6. DAGM-Symp. Mustererkennung, 1984. In: *Informatik-Fachberichte 87*. Hrsg. Kropatsch, W.. Berlin-Heidelberg-New York-Tokio: Springer, 1984. S. 147-153.
- [Kories 89] Kories, R.; Zimmermann, G.: *Eine Familie von nichtlinearen Operatoren zur robusten Auswertung von Bildfolgen*. In: *Ausgewählte Verfahren der Mustererkennung und Bildverarbeitung*; VDI-Fortschrittberichte, Reihe 10, Nr. 114. Düsseldorf: 1989. S. 96-119.
- [Kreucher 04] Kreucher, C.; Hero, A.O.; Kastella, K.: *Multiple Model Particle Filtering for Multitarget Tracking*; In: *Proc. of the 12th Annual Conference on Adaptive Sensor Array Processing (ASAP)*, 2004.
- [Liu 93] Liu, J.; Skerjanc, R.: *Stereo and Motion Correspondence in a Sequence of Stereo Images*. *Signal Processing: Image Communication*, Volume 5, No. 4, October 1993, Elsevier Science Publishers, S. 305-318.

## 9. LITERATURVERZEICHNIS

---

- [Lucas 81] Lucas, B.; Kanade, T.: *An Iterative Image Registration Technique with an Application to Stereo Vision*, 1981.
- [Malik 97] Malik, J.; Russell, S.: *Traffic Surveillance and Detection Technology Development: New Traffic Sensor Technology Final Report*. Institute of Transportation Studies, California Partners for Advanced Transit and Highways, University of California, Berkeley, Forschungsbericht (UCB-ITSPRR-97-6), 1997.
- [Martnez 04] Martnez, S. V.; Knebel, J.-F.; Thiran, J.-P.: *Multi-Object Tracking using the Particle Filter Algorithm on the Top-View Plan*. [http://lts1pc19.epfl.ch/repository/Venegas2004\\_730.pdf](http://lts1pc19.epfl.ch/repository/Venegas2004_730.pdf)
- [Maskell 01] Maskell, S.; Gordon, N.: *A Tutorial on Particle Filters for On-line Nonlinear/Non-Gaussian Bayesian Tracking*; In: IEE Target Tracking: Algorithms and Applications. London: IEE, 2001, Band 2, pp. 2/1–2/15.
- [Mecke 98a] Mecke, R.; Al-Hamadi, A.; Michaelis, B.: *A Robust Method for Block-Based Motion Estimation in RGB-Image Sequences*. Proceedings of the 14th International Conference on Pattern Recognition; ICPR 98, Volume I, Brisbane, Australia, 1998, pp. 663-667.
- [Mecke 98b] Mecke, R.; Al Hamadi, A.; Michaelis, B.: *A Robust Method for Block-Based Motion Estimation in RGB-Image Sequences*. Proceedings of the 14<sup>th</sup> International Conference on Pattern Recognition (ICPR '98), Brisbane, Australien, 17.-20. August 1998, Volume I, S. 663-667.
- [Mecke 99] Mecke, R.: *Grauwertbasierte Bewegungsschätzung in monokularen Bildsequenzen unter besonderer Berücksichtigung bildspezifischer Störungen*, Dissertation, Universität Magdeburg, Fakultät Elektrotechnik, 1999.
- [Mecke 01] Mecke, R.; Michaelis, B.: *A combined method for 3-d motion analysis in image sequences*. IASTED International conference on Signal and image processing, -proceedings. Anaheim, CA: ACTA Press, 2001, pp. 328 - 332
- [Mecke 03] Mecke, R.; Michaelis, B.: *3-D Motion and Shape from Multiple Image Sequences*. 4th International Conference on 3-D Digital Imaging and Modeling, Banff, Alberta, Canada, 2003, pp 155-162
- [Melzer 02] Melzer, B., Miene, A., Hermes, Th.: *Bewegungsanalyse in Bildfolgen auf Basis eines n:m-Matchings von Farbregionen*, Universität Bremen, 2002.
- [Meyberg 01] Meyberg, K.; Vachenaer, G.: *Höhere Mathematik 2 Funktionentheorie, Fourier-Analyse, Variationsrechnung*; Springer Verlag, Berlin-Heidelberg, Germany, 2001.
- [Michels 03] Michels, K.; Klawonn, F.; Kruse, R.; Nürnberger, A.: *Fuzzy Regelung, Grundlagen, Entwurf, Analyse*, Springer-Verlag, 2003.
- [Möller 01] Möller, B.; Posch S.: *Detection and tracking of moving objects for mosaic image generation*; 23<sup>rd</sup> DAGM, Munich, Germany, 2001, pp.208-214.
- [Murphy 01] Murphy, K.; Russel, S.: *Rao-Blackwellised Particle Filtering for Dynamic Bayesian Networks*. In: Arnaud Doucet (Hrsg.): *Sequential Monte Carlo Methods in Practice*. New York: Springer, 2001, chapter 24, pp. 499–515.
- [Musmann 85] Musmann, H. G.; Pirsch, P.; Grallert, H. J.: *Advances in Picture Coding*; Proc. Of the IEEE, 1985, Volume 73, No. 4; pp. 523-548.

## 9. LITERATURVERZEICHNIS

---

- [Musmann 85] Musmann, H. G.; Pirsch, P.; Grallert, H. J.: *Advances in Picture Coding*. Proc. of the IEEE, 1985, Volume 73, No. 4. S. 523-548.
- [Mysliwetz 86] Mysliwetz, B.D.; Dickmann, E.D: *A Vision System With Active Gaze Control For Real-Time Interpretation Of Well Structured Dynamic Szenes*. Conf. Intelligent Automomous Systems, Amsterdam, 1986.
- [Mysliwetz 95] Mysliwetz, B.D.: *Parallelrechnerbasierte Bildfolgen-Interpretation zur autonomen Fahrzeugsteuerung*. Universität der Bundeswehr München, Dissertation, 1995.
- [Nagel 85] Nagel, H. H.: *Analyse und Interpretation von Bildfolgen*, Teil I und II. Informatik Spektrum, Band 8, 1985; S. 178-200 und 312-327.
- [Neri 98] Neri, A.; Colonnese, S.; Russo, G.; Talone, P.: *Automatic moving object and background separation.*; Signal Processing, Vol. 66, No. 2, pp. 219-232, 1998.
- [Niese 03] Niese, R.; *Merkmalsorientierte Bewegungsanalyseverfahren auf der Grundlage von Farbinformation*; Diplomarbeit Universität Magdeburg; 2003.
- [Niese 07a] Niese, R.; Al-Hamadi, A.; and Michaelis, B.: *A Novel Method for 3D Face Detection and Normalization*, JOURNAL OF MULTIMEDIA, ISSN: [1796-2048, 2007], Vol. 2, No. 5, September 2007, ACADEMY PUBLISHER.
- [Niese 07b] Niese, R.; Al-Hamadi, A.; and Michaelis B.: *Real-time Capable Method for Facial Expression Recognition in Color and Stereo Vision*, International Conference on Computational Science and Its Applications, Malalaysia, LNCS 4705, Springer Verlag, 2007.pp.397-408.
- [Niese 07c] Niese, R.; Al-Hamadi, A.; and Michaelis B.: *Nearest Neighbor Classification for Emotion Recognition in Stereo Image Sequences*, ISAST Transactions on Electronics and Signal Processing, International Society for Advanced Science and Technology, No. 1, Vol. 1, 2007, ISSN 1797-2329, pp.88-94.
- [Niese 08a] Niese, R.; Al-Hamadi, A.; Aziz, F.; Michaelis, B.: *Robust Facial Expression Recognition Based on 3-d Supported Feature Extraction and SVM Classification*. In IEEE International Conference on Face and Gesture Recognition (FG2008), 2008, CD.
- [Nummiaro 02] Nummiaro, K.; Koller-Meier, E.; Van Gool, L.: *Object tracking with adaptive colour-based particle filter*; DAGM 2002; Zurich, 2002, Springer-Verlag; pp.353-360
- [Ohashi 97] Ohashi, K.; Kimachi, M.; Kanama, K.: *Vehicle Recognioon from Parking Images Using Stereo Method*; Proc. of the 4<sup>th</sup> World Congress on Intelligent Transport System; 1997.
- [Ohm 95] Ohm, J.-R.: *Digitale Bildcodierung*. Springer-Verlag Berlin-Heidelberg-New York, 1995.
- [Pathan 09a] Pathan, S.; Al-Hamadi, A.; and Michaelis: *Feature-supported Multi-hypothesis Framework for Multi-object Tracking using Kalman Filter*, The 17-th International Conference on Computer Graphics, Visualization and Computer Vision'2009, pp. 197-202, February 2 - 5, 2009. – Plzen.

- [Pathan 09b] Pathan S., Al-Hamadi A., Michaelis B.: *Multi-Object Tracking Using Semantic Analysis and Kalman Filter*, Proc. Of the 6<sup>th</sup> International Symposium on Image and Signal Processing and Analysis (2009), IEEE, Salzburg Austria 2009; pp. 271-275.
- [Pathan 09c] Pathan, S.; Al-Hamadi, A.; and Michaelis: *OIF-An Online Inferential Framework for Multi-object Tracking with Kalman Filter*. Computer Analysis of Image and Patterns, CAIP 2009, LNCS 5702, Münster, Germany, pp. 1087-1095.
- [Paulus 01] Paulus D.: *Aktives Bildverstehen*, Universität Erlangen, Nürnberg 2001, pp.114-135.
- [Pavlina 03] Pavlina K., Udvarov A., Semerdjiev T.: *A Study of a Target Tracking Algorithm Using Global Nearest Neighbor Approach*. Proceedings of the 4th international conference on Computer systems and technologies: e-Learning (CompSysTech). New York: ACM Press, 2003, pp. 290–295.
- [Pham 00] Pham, D.T.; Karaboga, D.: *Intelligent Optimisation Techniques*, Springer, 2000.
- [Prati 01] Prati, A.; Mikic, I.; Grana, C.; Trivedi, M.: *Shadow detection algorithms for traffic flow analysis: a comparative study*; Proc. of IEEE Int. Conference on ITS, pp. 340-345, 2001.
- [Priebe 93] Priebe, Lutz; Rehrmann, Volker: *A Fast Hybrid Color Segmentation Method*. In: Proceedings "Mustererkennung 1993", 15. DAGM Symposium, Lübeck. S. 297-304.
- [Priebe 98] Priebe, L.: *Vergleich von Farbsegmentierungstechniken*, Labor Bilderkennen Universität Koblenz, [www.uni-koblenz.de/~lb](http://www.uni-koblenz.de/~lb), 1998.
- [Rath 01] Rath, T. M.: *Spezialisierung generischer Modelle von Straßenverkehrsszenen für die Bildfolgenauswertung*, Diplomarbeit, Fraunhofer-Institut für Informations- und Datenverarbeitung, Karlsruhe 2001.
- [Rehrmann 94] Rehrmann, V.: *Stabile, echtzeitfähige Farbbildauswertung*. PhD Thesis ,Verlag Fölbach, Koblenz, 1994
- [Rehrmann 96] Rehrmann, V., Rothhaar, M.: *Objektorientierte Bewegungsschätzung in Farbbildfolgen*. In: 2. Workshop Farbbildverarbeitung, Seiten 69-72, 1996. Ilmenau.
- [Rehrmann 98a] Rehrmann, V.: *Object-Oriented Motion Estimation in Color Image Sequences*. In: Proc. of the 5<sup>th</sup> European Conf. on Computer Vision. Springer-Verlag, 1998, pp. 704-719
- [Rehrmann 98b] Rehrmann, V.: *Korrespondenzanalyse in Farbbildern*. Workshop Dynamische Perzeption, Bielefeld, 18-19th June 1998, pp. 23-30.
- [Roecker 93] Roecker, J, A., Phillies, G, L.: *Suboptimal joint probabilistic data association*, IEEE Transaction on Aerospace and Electronic Systems Bd. 29, April 1993, pp. 510-517, ISSN 0018-9251.
- [Saneyoshi 96] Saneyoshi, K.: *Drive assist using stereo image recognition*, Proceedings of the 1996 IEEE Intelligent Vehicles Symposium, September 19-20, 1996 Tokyo, S. 230-235

## 9. LITERATURVERZEICHNIS

---

- [Schnelting 98] Schnelting, O.: *Schätzung von Bewegungsvektorfeldern in gestörter Umgebung unter Anwendung künstlicher neuronaler Netze*; Dissertation; OvG-Universität, 1997, Reihe 10, Nr. 523, Düsseldorf, VDI-Verlag, 1998.
- [Schwarzenberg 05] Schwarzenberg, G.: *Objektverfolgung mit Partikelfilter*, Universität Karlsruhe (TH), Diplomarbeit, 2005.
- [Schuster 94] Schuster, R.: *Adaptive modeling in color image sequences*. In: Mustererkennung 1994, ed. Kropatsch, W., pp. 161-169, Springer-Verlag, 1994
- [Schuster 95] Schuster, R.: *Objektverfolgung in Farbbildfolgen*. Dissertation, TU München, 1995.
- [Sens 09] Sens, T.: *Multi-Objekttracking durch eine Hybridlösung in Farbbildsequenzen*; Diplomarbeit Universität Magdeburg; IESK, 2009.
- [Shi 94] Shi, J.; Tomasi, C.: *Good Features to Track*, IEEE Conference on Computer Vision and Pattern Recognition 1994.
- [Sidenbladh 03] Sidenbladh, H.: *Multi-target particle filtering for the probability hypothesis density*. In: Proc. of the 6th International Conference on Information Fusion 2003. Cairns, 2003, Vol. 2, pp. 800–806.
- [Singh 91] Singh, A.: *Optic Flow Computation*. IEEE Computer Society Press. Los Alamitos, CA, 1991.
- [Smith 04] Smith, K; Perez D.G.: *Order Matters: A Distributed Sampling Method for Multi-Object Tracking*. In: Proc. BMVC, Schweiz. 2004.
- [Smith 96] Smith, J.R., Chang, S.: *VisualSEEk: A fully automated content-based image query system*, Department of Electrical Engineering and Center for Image Technology for New Media, Columbia University, 1996.
- [Smith 97] Smith, S.M.: *Reviews of Optic Flow, Motion Segmentation, Edge finding and Corner Finding*, Technical Report TR97SMS1, Oxford Centre for Functional Magnetic Resonance Imaging of the Brain (FMRIB), 1997.
- [Srinivasan 84] Srinivasan, R.; Rao, K.R.: *Predictive Coding based on Efficient Motion Estimation*. In: ICC 1984, Proc., May 1984, S. 521-526
- [Stiller 97] Stiller, C.; Pöschmüller W.; Hürtgen B.: *Stereo Vision in Driver Assistance Systems*, Intelligent Transportation Systems 1997.
- [Strauder 99] Strauder, J.; Mech, R.; Ostermann, J.: *Detection of moving cast shadow for object segmentation*; IEEE Transaction on Multimedia, 1999, 1(1): pp. 65-76.
- [Stüker 04] Stüker D.: *Heterogene Sensordatenfusion zur robusten Objektverfolgung im automobilen Straßenverkehr*, Dissertation, Universität Oldenburg, 2004.
- [Theodoridis 03] Theodoridis, S., Koutroumbas, K.: *Pattern Recognition*; Second Edition. Academic Press; Elsevier (USA); 2003.
- [Torma 01] Torma, P.: *Improving Monte-Carlo Bayesian Filtering by Means of Local Search: Theory and Applications in Visual Tracking*, <http://www.lateral.hu/tdk.pdf>, Thesis, 15th of January, 2001

## 9. LITERATURVERZEICHNIS

---

- [Tornow 03] Tornow, M.; Michaelis, B.; Kuhn, R.; Calow, R.; Mecke, R.: *Hierarchical method for stereophotogrammetric multi-object-position measurement*; 25th DAGM Symposium, September 10-12; 2003, Magdeburg, Germany
- [Torres 00] Torres, L.; Delp, E.J.: *New trends in image and video compression*, in X European Signal Processing Conference, Tampere, Finland, September 4-8, 2000.
- [Ulmer 94] Ulmer, B.: *Autonomous Automated Driving in real Traffic*. In: Proceedings of the First World Congress on Applications of Transport, Telematics and Intelligent Vehicle Highway Systems, 1994, S. 36–41
- [Vermaak 03] Vermaak, J.; Doucet, A.; Pérez, P.: *Maintaining Multi-Modality through Mixture Tracking*. In: Proc. of the 9th IEEE International Conference on Computer Vision (ICCV). Los Alamitos: IEEE Computer Society, 2003, Vol. 2, pp. 1110–1116.
- [Vermaak 05] Vermaak, J.; Godsill, S. J.; Pérez, P.: *Monte Carlo filtering for multi-target tracking and data association*. In: IEEE Transaction on Aerospace and Electronic Systems. 2005 Heft 41(1), pp. 309–332.
- [Weickert 00] Weickert, J.; Schnörr, C.: *Variational optic flow computation with a spatio-temporal smoothness constraint*, Technical Report / Department for Mathematics and Computer Science, University of Mannheim ; TR-2000-015, 68131 Mannheim, Germany, 2000.
- [Yang 02] Yang, M.-H.; Ahuja, N.; Tabb, M.: *Extraction of 2D Motion Trajectories and its Application to Hand Gesture Recognition*. In IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 24, pp. 1061–1074. 2002.
- [Zhang 92] Zhang, Z.; Faugeras, O.: *3D Dynamic Scene Analysis*. Springer Verlag Berlin-Heidelberg-New York, [ISBN 3-540-55429-7], 1992