

Universität der Bundeswehr München  
Fakultät für Luft- und Raumfahrttechnik  
Institut für Systemdynamik und Flugmechanik

# **Integrale Multisensorielle Fahrumgebungserfassung nach dem 4D-Ansatz**

Volker von Holt

Vollständiger Abdruck der von  
der Fakultät für Luft- und Raumfahrttechnik  
der Universität der Bundeswehr München  
zur Erlangung des akademischen Grades eines

Doktor-Ingenieurs (Dr.-Ing.)

genehmigten Dissertation

Vorsitzender: Prof. Dr.-Ing. C. Mundt  
1. Berichterstatter: Prof. Dr.-Ing. E. D. Dickmanns  
2. Berichterstatter: Prof. Dr.-Ing. G. Färber (Technische Universität München)

Diese Dissertation wurde am 24.06.2004 bei der Universität der Bundeswehr München, 85577 Neubiberg, eingereicht und durch die Fakultät für Luft- und Raumfahrttechnik am 30.06.2004 angenommen.

Tag der Prüfung: 12.11.2004



## Vorwort

Die vorliegende Arbeit geht zurück auf meine Tätigkeit als wissenschaftlicher Mitarbeiter am Institut für Systemdynamik und Flugmechanik der Universität der Bundeswehr München. Wesentliche Teile entstanden aber auch erst in der nachfolgenden Zeit parallel zu meiner Tätigkeit bei der AUDI AG.

Herrn Prof. Dr.-Ing. E.D. Dickmanns danke ich für die Unterstützung der Arbeit und die langjährige Geduld mit mir während deren Entstehung. Ebenso gebührt mein Dank Herrn Prof. Dr.-Ing. G. Färber für die freundliche Übernahme des Korreferats. Herrn Prof. Dr.-Ing. C. Mundt danke ich für die Übernahme des Vorsitzes der Prüfungskommission.

Den vielen ehemaligen Kollegen des Instituts, den wissenschaftlichen Mitarbeitern, wie dem Sekretariat und der Werkstatt gebührt mein Dank für die vielfältige Unterstützung, die gute Zusammenarbeit und das angenehme Arbeitsklima.

Den Kollegen bei der AUDI AG danke ich für ihre Toleranz in Zeiten, in denen die Erstellung dieser Arbeit nicht immer ganz frei war von Rückwirkungen auf meine dortige Tätigkeit. Ferner danke ich den Kollegen der Elektronikforschung der VOLKSWAGEN AG für vielfältige Anregungen, Diskussionen und die Unterstützung beim Einfahren von Fahrszenen.

Meiner Familie und meinen Freunden danke ich für ihr Verständnis und den Rückhalt, den sie mir über die Jahre gegeben haben.

Frau Margrit Meier danke ich für das sorgfältige Korrekturlesen der Arbeit.

**Meinen Eltern**

# Inhaltsverzeichnis

<b>I</b>	<b>Einführung</b>	<b>1</b>
<b>1</b>	<b>Einleitung</b>	<b>3</b>
1.1	Zivilisation und Fortschritt . . . . .	3
1.2	Der Begriff der Intelligenz . . . . .	5
1.3	Sehende, intelligente Maschinen als Ziel . . . . .	6
1.3.1	Entwicklungen bis 1996 . . . . .	7
1.3.2	Entwicklungen seit 1996 . . . . .	9
1.4	Die Gegenwart: Fahrerassistenzsysteme . . . . .	10
1.5	Zielsetzung und Beitrag der Arbeit . . . . .	11
1.6	Struktur der Arbeit . . . . .	12
<b>2</b>	<b>Stand der Technik</b>	<b>13</b>
2.1	Umfelderfassende Sensorik . . . . .	13
2.1.1	Radarsensorik . . . . .	14
2.1.2	Lasersensorik . . . . .	14
2.1.3	Videosensorik . . . . .	16
2.2	Umfelderfassung in (semi-)autonomen Fahrzeugen . . . . .	17
2.3	Informationsfusion . . . . .	20
<b>3</b>	<b>Maschinelle und Biologische Wahrnehmung</b>	<b>23</b>
3.1	Menschliche Wahrnehmung . . . . .	24
3.1.1	Das Wahrnehmungssystem . . . . .	24
3.1.2	Wahrnehmung statischer Szenen . . . . .	25
3.1.3	Wahrnehmung dynamischer Szenen . . . . .	27
3.1.4	Aufmerksamkeits-/Blickrichtungs-Steuerung . . . . .	28
3.1.5	Bewegungssteuerung . . . . .	29
3.2	Maschinelle Wahrnehmung . . . . .	29

3.2.1	Klassischer 4D-Ansatz zur visuellen Wahrnehmung . . . . .	29
3.2.2	Verallgemeinerter 4D-Ansatz zur multisensoriellen Wahrnehmung . . . . .	31
<b>II</b>	<b>Multisensorielle maschinelle Wahrnehmung</b>	<b>33</b>
<b>4</b>	<b>Sensoren</b>	<b>35</b>
4.1	Video . . . . .	35
4.2	Lidar . . . . .	37
4.3	Fahrzeugsensorik . . . . .	38
<b>5</b>	<b>Rekursive Schätzverfahren in der Umfelderkennung</b>	<b>39</b>
5.1	Kalman-Filter-Tuning . . . . .	39
5.2	Meßwertassoziation bei umfelderfassenden Sensoren . . . . .	40
<b>6</b>	<b>Umfeldmodellierung</b>	<b>43</b>
6.1	Bewegungsmodelle . . . . .	43
6.1.1	Objekte . . . . .	43
6.1.2	Eigenfahrzeug . . . . .	45
6.2	Formmodelle . . . . .	46
6.2.1	Formelemente . . . . .	47
6.2.2	Topologische Beschreibung der Gesamtform . . . . .	48
6.2.3	Geometrische Beschreibung der Gesamtform . . . . .	49
6.2.4	Photometrische Beschreibung der Formelemente . . . . .	49
6.2.5	Formmodell-Datenbank . . . . .	50
6.3	Umfeldmodell und Szenenbaum . . . . .	54
6.3.1	Szenenrepräsentation . . . . .	54
6.3.2	Zustandsschätzung auf Basis des Szenenbaums . . . . .	56
<b>7</b>	<b>Verfahren der Bildverarbeitung</b>	<b>59</b>
7.1	Konturvermessung . . . . .	59
7.1.1	Bildverarbeitungs-Operatoren . . . . .	59

7.1.2	Bestimmung der Extremwerte . . . . .	61
7.1.3	Maskenoperatoren mit variablen Nullspalten . . . . .	62
7.1.4	Vermessung von Konturkanten . . . . .	63
7.2	Linienaggregation . . . . .	66
7.2.1	Lokale gerichtete Hough-Transformation . . . . .	68
7.2.2	Konturpunktverkettung . . . . .	70
7.3	Bewertung von Messung und Gruppierung . . . . .	72
7.3.1	Bewertung der Attribute eines Extremums . . . . .	72
7.3.2	Gesamtgütemaß eines Extremums . . . . .	73
7.3.3	Gesamtgütemaß einer Kontur . . . . .	75
7.4	Zusammenfassende Wertung . . . . .	75
<b>8</b>	<b>Hopfield Netze</b>	<b>77</b>
8.1	Hopfield-Netze zur Lösung von Optimierungsproblemen . . . . .	79
8.1.1	Berechnungsvorschriften für Hopfield-Netze . . . . .	81
8.1.2	Einsatzgebiete von Hopfield-Netzen . . . . .	82
8.2	Hopfield-Netze im Rahmen der relationalen Merkmalszuordnung . . . . .	83
8.2.1	Unäre Abstandsmaße . . . . .	84
8.2.2	Binäre Abstandsmaße . . . . .	85
8.2.3	Bildung eines Attributierten Relationalen Graphen am Beispiel . . .	87
8.3	Alternative Zuordnungsstrategien . . . . .	91
<b>9</b>	<b>Multisensorielle Objekterkennung</b>	<b>93</b>
9.1	Allgemeines Vorgehen . . . . .	93
9.2	Meßmodell Video . . . . .	96
9.2.1	Vorhersage der Merkmale . . . . .	96
9.2.2	Vermessung der Merkmale . . . . .	98
9.2.3	Assoziation der Merkmale . . . . .	99
9.3	Meßmodell Lidar . . . . .	102
9.3.1	Vorhersage der Merkmale . . . . .	103
9.3.2	Vermessung der Merkmale . . . . .	104

9.3.3	Assoziation der Merkmale . . . . .	106
9.4	Form- und Bewegungsschätzung . . . . .	108
9.4.1	Fahrspur und relativer Eigenzustand . . . . .	108
9.4.2	Objekte . . . . .	109
9.5	Initialisierung . . . . .	111
9.5.1	Lidar . . . . .	111
9.5.2	Video . . . . .	112
<b>10</b>	<b>Szeneninterpretation</b>	<b>115</b>
10.1	Wissensrepräsentation und Regelbasis . . . . .	116
10.2	Domäne und Einsatzzweck . . . . .	119
10.3	Referenzszenario . . . . .	121
10.4	Zeitlicher Ablauf . . . . .	123
10.5	Hypothesenbewertung . . . . .	126
10.6	Aufmerksamkeitssteuerung . . . . .	129
10.6.1	Aufmerksamkeitssteuerung bei beschränkter Rechenleistung . . . .	129
10.6.2	Aufmerksamkeitssteuerung bei beschränktem Erfassungsbereich . .	130
10.6.3	Strategien zur Aufmerksamkeitssteuerung . . . . .	131
<b>III</b>	<b>Realisierung und Experimentelle Ergebnisse</b>	<b>133</b>
<b>11</b>	<b>Architektur des Wahrnehmungssystems</b>	<b>135</b>
11.1	Historische Entwicklung . . . . .	135
11.2	Hardware . . . . .	137
11.2.1	Versuchsfahrzeuge . . . . .	137
11.2.2	Rechnerhardware . . . . .	138
11.3	Software . . . . .	138
11.3.1	Betriebssystem und Programmiersprache . . . . .	138
11.3.2	Prozeßstruktur . . . . .	139
11.3.3	Essentielle Klassen . . . . .	140
11.3.4	Beispielhafte Abläufe . . . . .	142



<b>12 Experimentelle Ergebnisse</b>	<b>145</b>
12.1 Fahrspurerkennung . . . . .	146
12.1.1 Szenarienbeschreibung . . . . .	146
12.1.2 Ergebnisse der Fahrspurschätzung . . . . .	149
12.2 Objekterkennung . . . . .	151
12.2.1 Erkennung eines PKW . . . . .	151
12.2.2 Erkennung eines Kleintransporters . . . . .	156
12.2.3 Erkennung von Querverkehr . . . . .	160
12.2.4 Vergleichende Diskussion der Ergebnisse . . . . .	163
<b>13 Zusammenfassung und Ausblick</b>	<b>167</b>
<b>IV Anhang</b>	<b>169</b>
<b>A Homogene Koordinaten</b>	<b>171</b>
<b>B Rekursive Schätzverfahren</b>	<b>173</b>
B.1 Zustandsraumdarstellung . . . . .	173
B.2 Kalman-Filter . . . . .	174
<b>Literaturverzeichnis</b>	<b>177</b>



# Abbildungsverzeichnis

3.1	Schnittdarstellung des menschlichen Auges . . . . .	24
3.2	4D-Ansatz zum dynamischen maschinellen Sehen . . . . .	30
3.3	Verallgemeinerter 4D-Ansatz zum multisensoriellen maschinellen Sehen . .	32
4.1	Erfassungsbereich der Videokamera . . . . .	36
4.2	Kamera- und Bild-Koordinatensystem . . . . .	36
4.3	Strahlgeometrie des Mehrstrahl-Lidars . . . . .	38
6.1	Hierarchische Formkonstruktion aus Formelementen an einem Beispiel . .	47
6.2	Festlegung von Zwangsbedingungen durch Formparametrisierung . . . . .	49
6.3	Formmodell Fahrspur . . . . .	51
6.4	Formmodell Quader . . . . .	52
6.5	Quaderkonstruktion aus Formelementen . . . . .	53
6.6	Beispielausprägungen des Polyedermodells . . . . .	53
6.7	Polyeder Schnittdarstellung . . . . .	54
6.8	Szenenbaum . . . . .	55
7.1	Standardmaskenoperatoren zur Gradientenbestimmung . . . . .	60
7.2	Allgemeiner Korrelationsoperator . . . . .	61
7.3	Diskontinuitätenbestimmung mit unterschiedlichen Maskenbesetzungen . .	64
7.4	Konturen mit Aufsetzpunkten und Suchpfaden von Maskenoperatoren . .	65
7.5	Kantenorientierte Bildvermessung an einem Beispiel bei unterschiedlicher Parametrierung des Maskenoperators . . . . .	67
7.6	Linienaggregation mittels lokaler, gerichteter Hough-Transformation . . .	69
7.7	Linienaggregation durch Konturpunktverkettung . . . . .	70
7.8	Kantenorientierte Bildvermessung und Konturverkettung an einem Beispiel	71
7.9	“Zuordnungsrauschen“ aufgrund ambivalenter Extrema auf Suchpfaden . .	74
8.1	Neuron eines Hopfield-Netzes . . . . .	80

8.2	Rückgekoppelte Struktur eines Hopfield-Netzes . . . . .	81
8.3	Umsetzung einer Vertex-Relation in Kanten des ARG . . . . .	86
8.4	Notation einer Vertex-Relation . . . . .	86
8.5	Notation einer Parallel-Relation . . . . .	87
8.6	Bildung eines ARG anhand der Topologie des Formmodell . . . . .	88
8.7	Abbilden der gemessenen Daten-Merkmale auf den ARG . . . . .	89
8.8	Beispielhafter Ablauf der Iterationen eines ARG . . . . .	90
9.1	Module der Informationsfusion für eine 1-Sensor-Anordnung . . . . .	94
9.2	Module der Informationsfusion für eine 2-Sensor-Anordnung . . . . .	95
9.3	Reduktion der Verdeckungsanalyse auf ein ebenes Problem . . . . .	97
9.4	Beispielhafter Ablauf der “manuellen“ Zuordnung . . . . .	101
9.5	Zuordnung von Merkmals-Kandidaten zu Modell-Merkmalen . . . . .	101
9.6	Verdeckungsanalyse aus Sicht des Mehrstrahllasers in der Draufsicht . . .	104
9.7	Lidar-Rohdaten im Videobild und aus der Draufsicht . . . . .	107
9.8	Lidar-Segmente im Videobild und aus der Draufsicht . . . . .	107
9.9	Ablaufdiagramm der Fahrspur- und Eigenzustands-Schätzung . . . . .	109
9.10	Möglichkeiten zur Form- und Bewegungsschätzung von Objekten . . . . .	110
9.11	Lidarbasierte Instanziierung von Objekthypothesen . . . . .	112
9.12	Vorgehen bei der videobasierten Instanziierung von Objekthypothesen . .	114
10.1	Repräsentation von Wissen auf mehreren Ebenene . . . . .	117
10.2	Expertennetzwerk und unterlagerte Wissenssebenen . . . . .	118
10.3	Video-Umfeldexperte . . . . .	119
10.4	Lidar-Umfeldexperte . . . . .	119
10.5	Darstellung eines Referenzszenarios in der Draufsicht . . . . .	121
10.6	Ablaufdiagramm der Umfeldwahrnehmung mit Videosensor . . . . .	124
10.7	Ablaufdiagramm der Umfeldwahrnehmung mit Lidarsensor . . . . .	125
10.8	Beispiel einer Aufmerksamkeitssteuerung bei mehreren, unterschiedlich priorisierten Funktionen . . . . .	132
11.1	Erste Transputerstruktur zum autonomen Fahren am ISF . . . . .	136

11.2	Versuchsfahrzeug VaMoRs des ISF . . . . .	137
11.3	Versuchsfahrzeug AUDI A8 der AUDI AG . . . . .	137
11.4	Hardware-Anordnung des Versuchssystems . . . . .	138
11.5	Verteilungsdiagramm mit Komponenten . . . . .	139
11.6	Prozeßstruktur zur Umfeldwahrnehmung . . . . .	140
11.7	Hierarchie essentieller Klassen der Wahrnehmungskomponente . . . . .	141
11.8	Sequenzdiagramm für die Einbringung von Lidarmeßdaten . . . . .	143
11.9	Sequenzdiagramm für die Einbringung von Videomeßdaten . . . . .	143
12.1	Beipielszenario auf der A 39 bei Wolfsburg, Bild 1-6 . . . . .	147
12.2	Beipielszenario auf der A 39 bei Wolfsburg, Bild 7-12 . . . . .	148
12.3	Zustandsgrößen der Fahrspur im A 39 Szenario . . . . .	150
12.4	Erstmalige Detektion des PKW durch den Lidarsensor . . . . .	152
12.5	Erstmalige Detektion des PKW durch den Videosensor . . . . .	152
12.6	PKW mit “Fehl“-Segmentierung bei reiner Lidarvermessung . . . . .	153
12.7	PKW mit Stabilisierung der Schätzung durch den Videosensor . . . . .	153
12.8	Zustandsgrößenverlauf des PKW bei reiner Lidarvermessung . . . . .	154
12.9	Zustandsgrößenverlauf des PKW bei Lidar- und Videovermessung . . . . .	155
12.10	Erfassung des Transporters durch den Lidarsensor . . . . .	157
12.11	Erfassung des Transporters durch Lidar- und Videosensor . . . . .	157
12.12	Zustandsgrößenverlauf des Transporters bei reiner Lidarvermessung . . . . .	158
12.13	Zustandsgrößenverlauf des Transporters bei Lidar- und Videovermessung . . . . .	159
12.14	Querender PKW kurz nach dem Eintritt in den Sichtbereich . . . . .	161
12.15	Querender PKW frontal vor dem Eigenfahrzeug . . . . .	161
12.16	Zustandsgrößenverlauf eines querenden PKW . . . . .	162
12.17	Vermessung der Fahrbahninfrastruktur mit dem Lidarsensor am Beispiel von Leitplanken . . . . .	165
12.18	Lidarrohdaten mit den zugehörigen Entfernungswerten als Basis der Segmentierung des Videobildes . . . . .	166



# Kurzfassung

In der vorliegenden Arbeit wird eine Mehrsensoranordnung zur Wahrnehmung der Fahrumgebung in Strassenfahrzeugen vorgestellt. Als Sensoren gelangen eine monochrome Videokamera sowie ein Mehrstrahllaser zum Einsatz. Der Mehrstrahllaser hat die Aufgabe der Detektion von Objekten im Fernbereich und im peripheren Blickfeld, ist aufgrund seiner Eigenschaften aber nicht für die Feinvermessung oder Klassifikation von Objekten geeignet. Die Videokamera besitzt demgegenüber ein eingeschränkteres Blickfeld als der Mehrstrahllaser und liefert in diesem Bereich Informationen über den Fahrspurverlauf, sowie eine redundante, höher aufgelöste Vermessung der Objekte.

Im Gegensatz zu anderen Ansätzen erfolgt die Interpretation der Szene und die Objektbildung nicht getrennt für jeden Sensor, gefolgt von einer anschließenden Informationsfusion auf Objektebene. Vielmehr gehen die Rohdaten beider Sensoren in eine zentrale Einheit zur Szeneninterpretation und Objektbildung ein. Durch diese enge Verkopplung der Sensoren ergibt sich eine einschrittige Szeneninterpretation auf einer gemeinsamen Wissensbasis im Gegensatz zur vorgenannten zweischrittigen Verarbeitung mit getrennten Wissensbasen für jeden Sensor.

Im Rahmen der Arbeit werden aus dem Bereich Bildverarbeitung Verfahren zur kantenbasierten Bildvermessung, Merkmalsgruppierung und -assoziation vorgestellt. Aus dem Bereich der Lasersensorik werden Segmentierungs- und Assoziationsverfahren für Lasermeßdaten entwickelt. Im Zentrum der Interpretations- und Informationsfusionseinheit wird auf rekursive Schätzverfahren nach dem 4D-Ansatz zurückgegriffen.

Die Tragfähigkeit des verfolgten Systemansatzes wird in einer Reihe von Versuchen sowohl in Autobahnszenarien wie in innerstädtischen Szenarien nachgewiesen und diskutiert.

# Teil I

## Einführung





# Kapitel 1

## Einleitung

### 1.1 Zivilisation und Fortschritt

Der Begriff der “Zivilisation“ und der heutige Stand der mit diesem Attribut verbundenen zivilen Welt ist untrennbar mit den Begriffen Werkzeug und Maschine verbunden. Im Laufe der Evolution hat die Gattung Mensch das Bestreben und die Fähigkeit entwickelt, sich die “Arbeit“ im engeren wie im weiteren Sinn zu erleichtern mit dem Ziel, die eigene Macht über die Natur, aber auch über andere Menschen zu steigern, um so zu besseren Überlebenschancen zu gelangen. Sei es durch die Erfindung einfacher Werkzeuge für die Landwirtschaft oder durch die Erfindung von Waffen, haben diese Entwicklungen dazu geführt, daß die am besten angepaßten Menschen überlebt haben. Mit jeder neuen Erfindung war nicht nur einerseits eine Entlastung des Menschen - gleichbedeutend mit einer Produktivitätssteigerung - verbunden, sondern andererseits auch die Möglichkeit gegeben, die freigewordenen Kapazitäten für neue Ideen und Erfindungen zu nutzen. Dies mußten nicht zwangsläufig materielle Dinge sein, sondern es konnte sich auch um geistige Werke jeder Art handeln, wie man sie im westlichen Zivilisationskreis insbesondere mit dem antiken Ägypten, Griechenland oder dem Römischen Reich verbindet. Alle diese Kulturkreise waren durch gleichzeitiges Voranschreiten in technischer wie in geistiger Hinsicht gekennzeichnet.

Offensichtlich ist es aber so, daß die Stabilität eines Kulturkreises einerseits bedingt durch äußere Herausforderungen anderer Kulturkreise aber auch durch innere Instabilität an einer gewissen Stufe ins Wanken gerät. Dies scheint immer dann der Fall zu sein, wenn die Innovationskraft abnimmt, weil die Menschen eines Kulturkreises in ihren Ideen und Theorien gefangen sind. So wurde mit der Zerstörung des Römischen Reichs eine Zivilisationsstufe verlassen, die aus vielfältigen Gründen erst über 1000 Jahre danach im späten Mittelalter wieder in anderer Form erreicht wurde - vielfach durch Wiederaufgreifen antiken Gedanktums (Demokratisierungsbestrebungen, Abschaffung der Leibeigenschaft/Sklaverei).

Der Beginn der Neuzeit läßt sich in diesem Sinn als der Beginn der Industrialisierung verknüpfen mit der Erfindung der Niederdruck-Dampfmaschine durch James Watt im Jahr 1765, die mit einem Schlag den Wechsel von durch menschliche oder tierische Arbeits-

kraft getriebenen Werkzeugen und Maschinen auf einen Antrieb durch maschinell erzeugte Energie darstellte. In Verbindung mit der Entdeckung der Elektrizität und die heutzutage nahezu uneingeschränkte Mobilität durch Eisenbahn-, Automobil- und Flugverkehr stellt dies eine einzigartige Emanzipation des Menschen von der Natur dar. Diese gab den Menschen einerseits eine noch vor wenigen Generationen undenkbbare Freiheit, die ihn andererseits aber auch der Natur und den natürlichen Lebensgrundlagen entfremdet hat und die angesichts der heutigen Weltbevölkerungszahl neue Probleme und Abhängigkeiten gebracht hat.

Vielleicht wäre die westliche Welt, welche diese Zivilisationsstufe im wesentlichen geprägt hat, ebenso an dem Punkt angelangt, wo Ihre Schöpfungskraft abnimmt und würde wie frühere Hochkulturen untergehen. Anders als bei vorangegangenen Kulturen hat es aber neben der Dampfmaschine noch eine weitere revolutionäre Erfindung gegeben: den Computer. Dieser hat in seiner Anfangsausprägung als Rechenmaschine rein mechanischer Art in verschiedenen Ausführungen begonnen, ehe er im Jahr 1941 von Konrad Zuse in elektrischer Gestalt erfunden wurde. Aber erst gepaart mit der Erfindung des Transistors im Jahre 1947 durch Bardeen, Brattain und Shockley [Bardeen, Brattain, Shockley 47] und die im Verlauf der letzten Jahrzehnte erfolgte Miniaturisierung im Bereich der Elektronik rücken neue Ideen in greifbare Nähe. Sollte es, nachdem uns die Dampfmaschine mit der von ihr erzeugten Energie in die Lage versetzt hat, nahezu alle mechanischen Arbeiten von Menschen auf Maschinen zu übertragen, mithilfe von Computern möglich sein, auch geistige, intelligente Tätigkeiten auf Maschinen zu übertragen ? Das wäre ganz sicher ein gewaltiger Schritt, der sofort die Frage aufwirft, was dann noch für den Menschen zu tun bleibt, wenn Maschinen intelligent werden und sich am Ende womöglich gar selbst erfinden.

Aller "Science-Fiction" zum Trotz wird es aber so schnell wohl kaum soweit kommen. Geht man davon aus, daß der Mensch als das am höchsten entwickelte Lebewesen die Maßstäbe für eine solche Maschine setzt, so sprechen eine ganze Reihe technischer Gründe gegen solche Visionen :

- Die Anzahl der Neuronen des menschlichen Gehirns im Vergleich zur Anzahl der Transistoren auf einem Mikroprozessor.
- Die Art der assoziativen Speicherung von Wissensinhalten und die weitestgehend ungeklärte immense Lernfähigkeit des menschlichen Gehirns.

Vor allem der Umstand der völlig anders gearteten Informationsspeicherung und -verarbeitung beim Menschen und die relative Unkenntnis darüber sprechen gegen derartige Phantasien. Zwar gibt es eine Vielzahl sowohl physiologischer als auch psychologischer Untersuchungen an Lebewesen, ebenso wie es in technischer Hinsicht Bestrebungen mit dem Anspruch gibt, die menschliche Verarbeitungs- und Verhaltensmuster nachahmen sollen (Beispiele: Künstliche Intelligenz, Künstliche Neuronale Netze, Fuzzy Logik). Doch das Problem wird im allgemeinen nur am Rande gestreift und gerade im Hinblick auf die genannten technischen Verfahren ist die Namensgebung oft schon die größte Gemeinsamkeit mit menschlichen Verarbeitungsmustern.

Folgt man vielmehr einer Politik der kleinen Schritte, so stellt sich zunächst die Frage, auf welchen Gebieten Ansatzpunkte bestehen, intelligentes Verhalten von Maschinen sinnvoll

zu demonstrieren, um so zu weiteren Erkenntnissen der Übertragung intelligenten Verhaltens auf Maschinen zu gelangen. Um eine Antwort auf diese Frage zu geben, bedarf es jedoch zunächst noch einer genauen Übereinkunft darüber, was denn eigentlich unter dem Begriff “Intelligenz“ zu verstehen sei.

## 1.2 Der Begriff der Intelligenz

Der Begriff der Intelligenz wird, da er dem Alltagsvokabular entstammt, mit den verschiedensten vagen Bedeutungen und Assoziationen belegt. Diese reichen von der Bindung an bestimmte Noten in der Schule oder bei sog. Intelligenztests bis hin zum Vorhandensein bestimmter (oft einseitiger) Begabungen und Anlagen. In einer Reihe einschlägiger Literatur- und Nachschlagewerke findet man zu diesem Begriff folgende Definitionen:

### **Philosophisches Wörterbuch** [Schischkoff 82]

(Intelligenz ist) ... die dem Menschen eigene geistig-verstandesmäßige Begabung, im engeren Sinn die Fähigkeit, sich in ungewohnten Situationen schnell zurechtzufinden, das Wesentliche eines Sachverhalts oder eines Vorgangs richtig und schnell zu erfassen, überhaupt geistige Beweglichkeit, Anpassungsfähigkeit, Neugierde, die Fähigkeit raschen Denkens und Urteilens, ....

### **Klinisches Wörterbuch** [Psyhyrembel 82]

Die Intelligenz ist eine zusammengesetzte Fähigkeit, nämlich die individuelle geistige Fähigkeit, richtig und weitgehend zu abstrahieren, kombinieren, transformieren und implizieren.

### **Brockhaus Lexikon** [Brockhaus]

(Intelligenz ist) ... Klugheit, Fähigkeit der Auffassungsgabe, des Begriffs, Urteilens, geistige Anpassungsfähigkeit an neue Aufgaben ...

Diese mit Absicht aus verschiedenen Fachgebieten stammenden Definitionen zeigen zunächst einen doch recht breiten Konsens darüber, was in etwa unter dem Begriff “Intelligenz“ zu verstehen ist. Doch auch die zur Begriffsdefinition herangezogenen Ausdrücke sind an sich wieder vage gehalten. So zeigt etwa die Aussage über den Intelligenzquotienten (IQ) aus [Psyhyrembel]: “Der IQ gibt nur bedingt das Leistungsniveau eines Kindes oder Erwachsenen wider, da er wesentliche Persönlichkeitsmerkmale nicht erfasst.“ Im gleichen Sinn trifft [Zimbardo 83] seine knappe Definition der Intelligenz, da diese im Sinne der Psychologie “mit sämtlichen Persönlichkeitsmerkmalen verwoben ist“. “Dies kann sowohl die Feinmotorik betreffen ebenso wie das soziale Verhalten“.

Es lassen sich aber dennoch mehrere wesentliche Kennzeichen des Begriffs “Intelligenz“ aus diesen Definitionen ableiten:

- **Auffassungsgabe**, d.h. die Kopplung an die Fähigkeit, Dinge auf- oder wahrzunehmen.

- **Lernen**, d.h. die Fähigkeit, Dinge zu beurteilen und sein eigenes Verhalten und Urteil zu adaptieren im Sinne von Abstrahieren (neues Wissen bilden) und Trainieren (bekanntes Wissen schnell und sicher anwenden).
- (Angepaßtes) **Verhalten**, d.h. die Kopplung an die Fähigkeit, die aus dem Erlernten gezogenen Schlüsse zu äußern oder anzuwenden.

In jedem Fall ist die Intelligenz und deren Ausbildung also untrennbar mit der Kognition verbunden. Wir nehmen Reize aus unserer Umwelt über unsere Sinnesorgane auf, abstrahieren daraus Wissen über unsere Umwelt - der Vorgang, den man üblicherweise als Lernen bezeichnet - speichern dieses erlernte Wissen, und können so später in ähnlichen Situationen durch Äußerungen oder Handlungen (Interaktion) angepaßt(er) auf diese Situationen reagieren. Ein solches Verhaltensschema auf eine Maschine zu übertragen, bedeutet nicht weniger als daß eine so geartete Maschine über Sinnesorgane (Sensoren) verfügen muß um ihre "Intelligenz" zu entwickeln, und daß sie über Handlungsorgane (Aktuatoren) verfügen muß, um ihre Intelligenz zu äußern. Das heißt aber nichts anderes als daß die kognitiven Aspekte ein ganz wesentliches, wenn nicht das wesentliche Arbeitsgebiet zur Erzielung von Fortschritten auf dem Weg zu intelligenten Maschinen sind.

### 1.3 Sehende, intelligente Maschinen als Ziel

Der Sinn, welcher beim Menschen (aber auch bei vielen anderen, insbesondere höheren Lebewesen) eine überragende Bedeutung besitzt, ist der Gesichtssinn. Die im Laufe der Evolutionsgeschichte relativ spät erfolgte Entwicklung dieses Sinnes steht beim Menschen in direktem Zusammenhang mit der Bevölkering der Bäume ("Klettern") sowie der Entwicklung des aufrechten Gangs. Da sowohl die "natürliche" Umwelt, als auch in noch größerem Ausmaß die vom Menschen geschaffene "künstliche" Umwelt auf die Dominanz und Leistungsfähigkeit des menschlichen Sehsystems ausgerichtet sind, bietet sich eine Nutzbarmachung dieses Sinnes auch im Rahmen der Robotik mit dem Ziel sehender, intelligenter Maschinen an [Nilsson 84], [Moravec 83]. Dieser Weg ist am Institut für Systemdynamik und Flugmechanik (ISF) der Universität der Bundeswehr München (UniBwM) in der Arbeitsgruppe von Prof. Dickmanns beschritten worden, wo sich das Gebiet des Rechnersehens im Bereich der Robotik als Hauptarbeitsgebiet herausgebildet hat.

Den Kern des am ISF verfolgten Ansatzes bildet die Annahme eines internen Weltmodells, welches auf dem Einsatz dynamischer Modelle basiert. Auf der Basis dieser Modelle kann der erwartete Zustand der realen Welt für zukünftige Zeitpunkte vorhergesagt werden. Auf der Basis eines Vergleichs dieser Erwartungen mit Messungen aus der realen Welt folgt dann, getrieben durch die auftretenden Abweichungen, eine Verbesserung der internen Modellvorstellung. Durch diese Art der Verarbeitung kann einerseits die nichteindeutige Inversion der Abbildungsgleichungen umgangen werden, andererseits ist der benötigte Rechenaufwand aufgrund des rekursiven Ansatzes sehr gering, was die Echtzeitfähigkeit des Verfahrens begründet. Auf diesen rekursiven Ansatz wird in Kapitel 3 ausführlicher eingegangen. An dieser Stelle soll zunächst die historische Entwicklung dargestellt werden.

### 1.3.1 Entwicklungen bis 1996

Ausgehend von der Idee, dynamische Systeme auf der Basis visueller mit einer Kamera aufgenommener Informationen zu regeln [Dickmanns 80], erfolgten am ISF Anfang der 80er Jahre erste Versuche mit einem Stab-Wagen-System (invertiertes Pendel) [Meissner 82]. In dem von einer Kamera gelieferten Bild wurde die Stabposition im Bild mittels einfacher Merkmalsoperatoren zur Kantendetektion vermessen und mit der aus dem dynamischen Modell des Stab-Wagen-Systems prädizierten verglichen. Auf der Basis der daraus folgenden Zustandsbeobachtung des Systems konnte der Stellmotor entsprechend gesteuert und der Stab in Balance gehalten werden. Durch Hinzunahme eines einfachen Formmodells für den Stab konnten bereits damals Fehlmessungen im Bild als solche erkannt werden.

Als zweites erfolgreiches Anwendungsbeispiel demonstrierte [Wünsche 87] ein Satellitenandockmanöver, bei dem ein "Servicesatellit" seinen Rendezvouspartner unter mehreren in der Szene enthaltenen Objekten zunächst erkennen und anschließend (mit einem Stachel) an diesen andocken mußte. Im Rahmen dieser Arbeit kam erstmals der sog. 4D-Ansatz [Dickmanns 87] zur Interpretation dynamischer Szenen in seiner vollen Ausprägung zur Geltung, da das System zur Erfüllung seiner Mission ein Szenenmodell und ein dynamisches Modell der Sonde beinhaltete und die Rückkopplung der visuellen Informationen mittels eines Kalman-Filters [Kalman 60] erfolgte.

Ungefähr zeitgleich konnten 1986 auch erste autonome Fahrten mit dem dafür speziell ausgerüsteten Versuchsfahrzeug VaMoRs (Versuchsfahrzeug für autonome Mobilität und Rechnersehen) durchgeführt werden [Zapp 88]. Beim VaMoRs handelt es sich um einen 5t Kastenwagen der Firma Daimler-Benz, der mit einem Generator für die Stromversorgung, Sensoren und Aktuatoren für Gas, Bremse und Lenkung sowie einer in zwei Achsen beweglichen Plattform zur Aufnahme mehrerer Kameras ausgestattet ist. Ziel der Aktivitäten in diesem Bereich war zunächst die autonome quergeregelte Fahrt entlang markierter Fahrspuren. Ausgehend von einem generischen räumlichen Modell für den Fahrspurverlauf und einem dynamischen Modell für die Fahrzeugbewegung erfolgt die Schätzung der betreffenden Zustände aus dem Vergleich des prädizierten Verlaufs der Fahrspurmarkierungen und den tatsächlich im Videobild gemessenen Markierungen. Während das anfängliche Modell lediglich horizontale Krümmungen des Fahrspurverlaufs berücksichtigte, wurde dies später durch [Mysliwetz 90] auch auf vertikale Krümmungen ausgedehnt und durch [Behringer 96] zu einem sehr robusten Verfahren weiterentwickelt, das auch in weiteren autonomen Versuchsfahrzeugen von Industriepartnern (Daimler-Benz, Dornier) eingesetzt wurde.

Um über die autonome Querführung auf Basis der Fahrspurerkennung hinaus auch die autonome Längsführung von Straßenfahrzeugen zu beherrschen, müssen darüberhinaus Hindernisse erkannt werden. Im Rahmen der Arbeiten von [Thomanek 96] wurde hierfür eine Komponente am ISF entwickelt und erfolgreich in den Versuchsfahrzeugen eingesetzt. Die hierfür eingesetzten Verfahren basierten aufgrund der beschränkten Rechenleistung auf sehr einfachen aber dennoch robusten Repräsentationen (einhüllende Rechtecke) und Verfahren (kantenbasierte Bildverarbeitung). Über diesen Ansatz hinaus gingen die Arbeiten von [Schick 92] und [Schmid 94], bei denen Fahrzeuge durch dreidimensionale Formen beschrieben werden. [Schick 92] legte das Hauptgewicht auf die Schätzverfahren und arbeitete in der Simulation, während [Schmid 94] auch auf realen Bildfolgen Versuche unternommen

hat. Neben diesen Arbeiten, die auf die Fahrzeugerkennung abzielen, fanden durch [Kinzel 94] auch erste Arbeiten statt, um den am ISF verfolgten Ansatz auch auf die Erkennung von Lebewesen auszudehnen.

Im Bereich der Handlungsableitung wurden neben den rein reflexiven Verhaltensmodi des Spurfahrens und Abstandhaltens auch Arbeiten an komplexeren Manövern wie etwa dem Konvoifahren und Überholmanövern begonnen [Brüdigam 94].

Im Jahr 1993 wurde dann mit dem VaMP (VaMoRs PKW) ein ebenfalls zu autonomen Versuchsfahrten umgerüsteter Mercedes 500 SEL in Betrieb genommen. Dieser ist im Gegensatz zum VaMoRs nur mit Kameraplattformen, die in einer Achse beweglich sind, ausgerüstet, erhielt aber eine zusätzliche Plattform für die Rückschau. Aufgrund seiner bauartbedingten höheren Endgeschwindigkeit wird das VaMP seitdem für Untersuchungen in Autobahnszenarien eingesetzt, während das VaMoRs für die Aufgaben der Navigation auf Wegen niedriger Ordnung und in leichtem Gelände verwendet wird. Mit dem VaMP einher geht auch der Nachweis, die benötigte Rechenleistung für autonome Fahrten auch in einem PKW unterbringen zu können.

Mit dem VaMP konnte 1994 das autonome Fahren im mittleren und dichten Verkehr auf Autobahnen nachgewiesen werden [Dickmanns et. al. 94]. Im Jahr 1995 wurde im Rahmen einer Langstreckenfahrt [Behringer 96] die Robustheit der Verfahren demonstriert. Mit dem VaMoRs wurden seit 1994 sowohl die Landmarkennavigation [Hock 94] als auch Abbiegemanöver [Müller 96] erfolgreich bewältigt, die beide eine zwingende Vorbedingung für das autonome Navigieren auf Wegenetzen und das autonome Fahren auf Straßen niedriger Ordnung sind.

Neben den Aktivitäten im Straßenfahrzeugbereich, die in den letzten Jahren den Hauptarbeitsbereich darstellten, wurde die Tragfähigkeit des 4D-Ansatzes im Bereich des dynamischen maschinellen Sehens aber auch durch Anwendungen im Luftfahrtbereich unter Beweis gestellt. So konnte [Eberl 87] erstmals 1987 einen autonomen Landeanflug von Flugzeugen ohne Störungen im Simulationskreis demonstrieren, den [Schell 92] durch Hinzunahme von inertialen Meßgrößen auch mit Windstörungen zur Anwendung brachte. Durch [Werner 97] wurden diese Ansätze in der Simulation auch auf den Bereich Hubschrauberflug ausgedehnt mit dem Schwerpunkt auf der Fusion von bildverarbeitungs- und inertialbasierter Eigenzustandschätzung. Ebenfalls in diesen Bereich läßt sich das im Rahmen der Space Shuttle D2-Mission 1993 erfolgte Experiment des "Greifens eines freifliegenden Körpers im Weltall" einreihen, bei dem es erstmals gelang, einen freifliegenden Körper durch Auswertung von Videobildfolgen (mit großer Totzeit) und gezieltes Steuern eines Greifarmer einzufangen [Fagerer 96], [D. Dickmanns 97].

**Kameraplattformen:** In den Bereichen der autonomen Land- und Luftfahrzeuge kommen wie o. a. Plattformen [Schiehlen 95] mit einer oder zwei beweglichen Achsen als Kameraträger zum Einsatz. Dadurch können die wahrnehmbaren Szenenbereiche gegenüber den Blickfeldern der Kameras deutlich erweitert werden, ohne daß die Anzahl der Kameras erhöht werden muß. Dies kann sowohl in Form eines langsamen Folgens oder Vorsteuerns - etwa beim Abbiegemanöver durch [Müller 96] - oder in Form schneller, sakkadischer Bewegungen - zur kurzzeitigen Blickfixierung relevanter Szenenbereiche - geschehen.

**Rechnerhardware:** Bis zum Jahr 1990 kamen als Rechnerhardware Intel x86-basierte Systeme zum Einsatz. Diese waren vom Institut für Meßtechnik (IFM) der UniBwM für diesen Einsatzzweck um einen Videobus erweitert worden. Auf den Systemen waren die bildverarbeitenden Prozesse plziert, während die Fahrzeuge auf dem Umweg über einen PC angesteuert wurden. Im Jahr 1990 begann der Umstieg auf Transputer und damit auf Standard-Hardware. Im Jahr 1994 waren dann alle Softwarekomponenten, sowohl die der Bildverarbeitung als auch die der Fahrzeug- und Plattformansteuerung, auf Transputern lauffähig. Mit den Transputern war ein sehr kompaktes und robustes System gefunden.

**Software:** Bis 1990 kam eine Vielfalt verschiedenster Programmiersprachen zum Einsatz. Diese reichten von Assemblersprachen über Fortran bis hin zu Pascal. Mit den Transputern erfolgte der Übergang auf C bzw. in Spezialfällen OCCAM als Programmiersprache. Da diese Programmiersprachen die Wiederverwendung vom Konzept her nur bedingt unterstützen und auch die Entwicklung auf eingebetteten Systemen der Transparenz wenig zuträglich ist, erfolgte die Entwicklung der meisten Funktionalitäten getrennt voneinander.

### 1.3.2 Entwicklungen seit 1996

Mit dem Ende der Entwicklung fundamentaler Fähigkeiten für autonome Straßenfahrzeuge Ende 1995 durch die Arbeiten von [Behringer 96], [Thomanek 96] und [Müller 96] wurde im Jahr 1996 ein neues Systemkonzept zum erwartungsbasierten **multifokalen sakkadischen Sehen (EMS-Vision)** ins Leben gerufen [Dickmanns 99]. Dessen Ziel besteht in der Hauptsache in der Integration der bestehenden Fähigkeiten in ein neues systematisches Gesamtkonzept, der Erweiterung bestehender Funktionen und der Handhabung komplexerer Szenarien und Missionen. Mit dem Neuentwurf einher sollte auch eine verbesserte Lern- und Adaptionsfähigkeit aller Systemkomponenten sowohl durch eine andere Softwarestruktur als auch durch die Integration datenbankähnlicher Konzepte gehen.

Einen ersten Entwurf in dieser Richtung bildet [D. Dickmanns 97], der für die Wahrnehmungsseite eine hierarchische (baumartige) Szenenrepräsentation und eine modifizierbare Objektrepräsentation entwarf, die u. a. auch von [Fagerer 96] und [Werner 97] eingesetzt wurde. Eine Implementation dieses Systems lag in der Programmiersprache C auf einem PC unter dem Betriebssystem DOS vor, wobei für die Bildmeßverfahren weiterhin ein Transputersubsystem zum Einsatz kommt.

Für das neue **EMS-Vision** Konzept fiel, dem damaligen Stand der Entwicklung in Hard- und Software entsprechend, die Entscheidung, auf PC-Systeme als Hardwarebasis, Windows NT als Betriebssystem und C++ als Programmiersprache zu gehen. Aus diesen Entscheidungen leiteten sich weitere Problempunkte ab, für die es eine Lösung zu finden galt. Dazu zählen einerseits Systemaspekte aufgrund der neuen Hard- und Software, andererseits aber auch die gesamte Systemarchitektur und funktionale Struktur, die auf dieser Basis unter der gegebenen Zielsetzung gänzlich neuen Anforderungen gerecht werden muß.

Im Rahmen des **EMS-Vision** Konzepts erfolgten Arbeiten, welche die Architektur des Handlungszeuges und damit die Verhaltensentscheidung für die Fahrzeugeigenbewegung



auf eine breitere Basis stellten [Maurer 00] und diese um die explizite Repräsentation von Fähigkeiten in Form eines Fähigkeitsnetzes erweiterten [Siedersberger 03]. Im Bereich der Verhaltensentscheidung und Blickrichtungssteuerung wurden wesentliche Beiträge von [Pellkofer 03] geliefert. Die von [Hock 94] begonnenen Arbeiten zur Landmarkennavigation wurden von [Gregor 02] im Rahmen von komplexeren Szenarien auch für Fahrten jenseits befestigter Wege ausgedehnt. Auf Seiten der Wahrnehmung wurde die Objekterkennung durch die Arbeit von [Rieder 00] weiterentwickelt; [Lützel 02] erweiterte die Fahrspurerkennung in Richtung der Auswertung von Farbbildern und des Fahrens auf Feldwegen.

## 1.4 Die Gegenwart: Fahrerassistenzsysteme

Einen evolutionären Schritt - im Sinne des Grades der Autonomie - vor der autonomen Fahrzeugführung liegen die sog. Fahrerassistenzsysteme. Diese können sich zwar möglicherweise eines Tages bis hin zu einer vollständigen Entlastung des Fahrers durch autonomes Fahren entwickeln, den Schwerpunkt der gegenwärtigen Entwicklungen bilden aber den Fahrer unterstützende, eben “assistierende“ Systeme, wie etwa das sog. “Adaptive Cruise Control“ (ACC) zur Abstandsregelung zum vorausfahrenden Fahrzeug mit aktivem Eingriff in die Längsführung oder das sog. “Lane Departure Warning“ (LDW) zur Spurverlassenswarnung. Diese beiden sind Vertreter der Systeme mit aktivem Eingriff (ACC) bzw. mit reiner Warn-/Informationsfunktion (LDW). Zur ersten Gruppe der aktiv eingreifenden Fahrerassistenzsysteme zählen auch das mittlerweile weitverbreitete “Elektronische Stabilitätsprogramm“ (ESP) oder die verschiedenen Formen von “Bremsassistenten“. Letztgenannte Fahrerassistenzsysteme benötigen (bislang) keine umfelderfassende Sensorik und werden daher im Rahmen dieser Arbeit nicht weiter betrachtet. Desweiteren unterscheidet man Fahrerassistenzsysteme (FAS) in Abhängigkeit der jeweiligen Funktionsausprägung in Komfort- und Sicherheitssysteme. Diese Unterscheidung hat auch ganz wesentlich mit der Frage der Verantwortung zu tun. So ist ACC als Komfortsystem definiert - die Verantwortung liegt weiterhin voll beim Fahrer - während beispielsweise ein System zur Automatischen Notbremsung (ANB) [Kopischke 00], [von Holt, Maurer 04] als Sicherheitssystem definiert ist und in fest definierten Bereichen damit auch die “Verantwortung“ übernimmt, mit dementsprechenden Forderungen an die Robustheit und Zuverlässigkeit der Umfelderfassung. Beiden gemeinsam ist im Gegensatz zu den autonomen Systemen aber immer die Berücksichtigung des Fahrers durch das System, auch wenn diese je nach System stark variieren kann. Zu den genannten Aspekten kommt als weiterer wichtiger Einflußfaktor die Frage der Wirtschaftlichkeit, da sich Fahrerassistenzsysteme dem Markt unterwerfen müssen, anders als das beim Forschungsziel des Autonomen Fahrens der Fall ist.

Die Herausforderung bei der Entwicklung derartiger Systeme stellt also nicht in erster Linie die möglichst vollständige Nachahmung des Verhaltens eines Fahrers dar, sondern die klar umrissene Definition eines den Fahrer unterstützenden Systems. Als eine Konsequenz daraus haben sich als umfelderfassende Sensoren neben der Videosensorik mehrere andere Sensorarten für verschiedene Systeme etabliert. Diese können - bezogen auf die jeweilige Aufgabe - dem Videosensor überlegen sein, erfordern weniger Rechenleistung, oder sind

einfach robuster, weniger komplex und damit kostengünstiger zu fertigen. Als wesentliche Sensoren seien in diesem Zusammenhang die Radar- und Lidar-Sensoren genannt, die in verschiedensten Ausprägungen zur Erfassung des Fern- und des Nahbereichs um das Fahrzeug dienen. Die Hauptvorteile bestehen in der - unter den gegebenen technischen Randbedingungen - häufig größeren Reichweite, der weitgehenden Witterungsunabhängigkeit (Radar) und der direkten Entfernungsmessung bzw. Geschwindigkeitsmessung (Radar). Im Gegensatz zur breit angelegten Herangehensweise beim EMS-Vision-System fokussieren diese Systeme viel stärker auf die zur Lösung der jeweiligen Aufgabe notwendigen Aspekte, haben dafür aber die Tendenz, spezialisierte Nischenlösungen zu bilden, auf denen Weiterentwicklungen nur schwer aufsetzen können.

## 1.5 Zielsetzung und Beitrag der Arbeit

Die im Abschnitt 1.3 dargestellten Arbeiten am ISF bestätigen durch ihre praktischen Erfolge die Richtigkeit des eingeschlagenen Weges. Allerdings ist die Lücke zwischen den Forschungsarbeiten in Richtung des “Autonomen Navigierens“ zu den wesentlich aktuelleren Zielsetzungen im Bereich der Umfelderkennung für Fahrerassistenzsysteme immer noch sehr groß. Gleichwohl besteht aus Sicht des Autors kein Zweifel, daß die Videosensorik sich in nächster Zeit auch im Automobilbereich zum wesentlichen Sensor der Umfelderkennung entwickeln wird. Aufgrund der geforderten Rahmenbedingungen, was die Robustheit und Zuverlässigkeit der Umfelderkennung betrifft, als auch was die Kosten und den Bauraum - insbesondere bei Mehrkamera-Anordnungen - für Videosensorik im Fahrzeug bedeutet, wird es aber im ersten Schritt sicher eine Kombination mit im Automobilbereich “etablierten“ Sensoren mit Radar- oder Lasertechnik geben.

In dieser Arbeit wird der Weg einer Kombination einer monokularen, monochromen Kameraanordnung mit einer Mehrstrahl-Lidarsensorik untersucht. Die beiden Sensoren sind zwar nicht in dem Maße komplementär wie dies eine Kombination aus Video- und Radarsensorik ist, dafür bietet das Lidarsystem an sich eine höhere laterale Auflösung und kann gut auf verschiedenen Ebenen mit den Videodaten fusioniert werden. Die Kombination beider Sensoren zur Umfelderkennung soll in einem allgemeinen Rahmenwerk zur Erkennung von Verkehrsszenen erfolgen, das modular aufgebaut und damit sowohl um zusätzliche Sensoren bzw. Sensorkategorien als auch um zusätzliche Objektmodelle erweitert werden kann. Im Rahmen der Arbeit wird damit eine allgemeine Szenenrepräsentation, verbunden mit entsprechenden Repräsentationen der Form und Bewegung von Objekten und Fahrbahnen entwickelt. Die Modelle und Methoden gehen dabei zum Teil bereits über das Ziel dieser Arbeit hinaus und sind - basierend auf früheren Arbeiten [von Holt, Baten 98] - für zukünftige Weiterentwicklungen vorgehalten. Ein wesentliches Augenmerk liegt in dem Einfluß und der Berücksichtigung der unterschiedlichen Eigenschaften der verwendeten Sensoren in Form entsprechender Sensor- und Meßmodelle.

Ein weiterer Beitrag der Arbeit ist im Entwurf der aus den genannten Anforderungen resultierenden Softwarestruktur in objektorientierter Sichtweise zu sehen. Diese ist als eine Hierarchie von Klassen, der Methoden dieser Klassen und ihrer Interaktion erfolgt. Nicht zuletzt wird aber auch der Nachweis der Machbarkeit eines solchen Ansatzes mit heutiger

Standard-PC-Hardware und Software und die Perspektive einer Portierung auf automotivetaugliche Steuergeräte als Ziel gesehen.

## 1.6 Struktur der Arbeit

Die Arbeit gliedert sich in insgesamt drei Teile. Den ersten einführenden Teil bildet diese Einleitung, gefolgt von der Darstellung des Standes der sensorischen - vornehmlich visuellen - Szenenerkennung in Abschnitt 2, welche sowohl auf den Bereich der autonomen Systeme, als auch auf reine Wahrnehmungssysteme ohne Handlungsweig eingeht. Im 3. Kapitel wird - anknüpfend an die in diesem Kapitel gegebene Hinführung zum Thema maschinelle Intelligenz - der Gedanke der visuellen Wahrnehmung beim Menschen herausgearbeitet. In Anlehnung daran erfolgt dann der Systementwurf zur maschinellen visuellen bzw. multisensoriellen Wahrnehmung.

Der zweite Teil der Arbeit ist den Bausteinen zur multisensoriellen maschinellen Wahrnehmung gewidmet. Zu diesen zählen neben der Darstellung der eingesetzten Sensoren in Kapitel 4 und einigen Anmerkungen zum Einsatz rekursiver Schätzverfahren im Rahmen der Umfelderkennung in Kapitel 5 vor allem die Szenen- und Objektrepräsentation im Rahmen der Umfeldmodellierung (Kapitel 6), die Methoden zur Bildvermessung (Kapitel 7) und das Zuordnungsverfahren für Bildmerkmale in Kapitel 8. Die Zusammenschau dieser "Bausteine" im Hinblick auf die multisensorielle Objekterkennung erfolgt in Kapitel 9. Abgeschlossen wird der zweite Teil mit dem Kapitel 10 über die Vorgehensweise und den Ablauf der Szeneninterpretation.

Der dritte und letzte Teil enthält die Darstellung der Realisierung in Form der Hardware- und Softwarearchitektur (Kapitel 11) und den eigentlichen Ergebnisteil in Kapitel 12. Die Arbeit schließt in Kapitel 13 mit einer Zusammenfassung und einem Ausblick.

# Kapitel 2

## Stand der Technik

Die Forschungsaktivitäten im Bereich autonomer bzw. semiautonomer Fahrzeuge mit Umfelderkennung, die Mitte der 80er Jahre begannen, haben sich von überschaubaren Aktivitäten einzelner Forscher in den Anfangszeiten hin zur Beschäftigung einer breitgefächerten Forschungsgemeinde mit dem Thema entwickelt. Es scheint daher nahezu unmöglich, einen kompletten Überblick über das Gebiet zu geben. Es sei daher zum einen auf die regelmäßig stattfindenden Konferenzen auf diesem Gebiet verwiesen. Dies sind in der Hauptsache die IEEE Conference on Intelligent Vehicles (IV), die IEEE Conference on Intelligent Transportation Systems (ITSC) sowie als Vertreter der sensororientierten Konferenzen die International Conference on Computer Vision (ICCV) - insbesondere letztgenannte ICCV deckt dabei natürlich auch viele Gebiete jenseits der Fahrumgebungserfassung mit ab. Zum anderen sei auf Übersichtsveröffentlichungen zu bestimmten Bereichen verwiesen, etwa auf [Kastrinaki et. al. 03] für einen neueren Überblick zum Thema Rechnersehen.

An dieser Stelle soll stattdessen ein mehr exemplarischer Überblick über die im Rahmen der Umfelderkennung eingesetzten Sensoren, eine Reihe prominenter Vertreter bzw. Aktivitäten auf dem Gebiet (semi-)autonomer Fahrzeuge mit Umfelderkennung gegeben werden, und zur Einordnung dieser Arbeit in das Themenfeld Sensorfusion werden schließlich einige aktuelle Arbeiten aus diesem Themengebiet angeführt.

### 2.1 Umfelderkennende Sensorik

Für die Erfassung der Fahrumgebung haben sich im automobilen Bereich vor allem vier Sensorprinzipien etabliert: Radarsensoren, Lasersensoren, Ultraschallsensoren und Videosensoren. Die Art der verwendeten Sensoren hängt dabei im wesentlichen vom Einsatzzweck ab. So sind vor allem im Bereich des autonomen Fahrens Videosensoren fast in allen Fahrzeugen anzutreffen, während die anderen Sensoren im Bereich serienreifer Fahrerassistenzsysteme sehr verbreitet sind. Dies steht in direktem Zusammenhang mit der Komplexität der Verarbeitung aber auch mit dem möglichen Informationsgewinn aus den Daten der Sensoren. Im folgenden werden die Charakteristiken der wesentlichen Sensoren näher beschrieben.

### 2.1.1 Radarsensorik

Beim Radar (Radio Detection and Ranging) besteht das Meßprinzip allgemein in der Laufzeitmessung einer reflektierten elektromagnetischen Welle, wobei zwischen verschiedenen Prinzipien, z. B. dem Pulsradar, dem frequenzmodulierten Dauerstrichradar (FMCW) und weiteren Verfahren unterschieden wird. Über die Laufzeit lassen sich der Abstand zu Objekten und zusätzlich über den Doppler-Effekt die Geschwindigkeit von Objekten direkt messen, wobei die Genauigkeit der Messung nicht entfernungsabhängig ist. Das Radar bietet damit eine robuste Erfassung beliebiger Objekte, die weitgehend unabhängig von Witterungs- und sonstigen Umweltbedingungen ist und sich allenfalls durch starke Gischt oder Schneematsch vor der Antennenlinse beeinträchtigen läßt.

Aufgrund des erforderlichen Hochfrequenzteils ist das Radar ein relativ aufwendiger und nicht besonders flexibler Sensor. Der heutzutage wesentlichste Einsatzzweck für Radarsensoren im Automobilbereich stellt die Verwendung im Rahmen des Adaptive Cruise Control (ACC) dar. Für diese Funktion wurden von mehreren Automobilzulieferern (z. B. Bosch, TRW, A.D.C. oder Delphi) Sensoren mit unterschiedlichen Meßprinzipien entwickelt, die typischerweise über einen recht eingeschränkten Sichtbereich - als Konsequenz aus den Funktionsanforderungen - mit 3-4 Radarkeulen, einem Öffnungswinkel von ca. 8-16° und einer Reichweite von ca. 150-200 m verfügen.

Ausserhalb der serienreifen Radarsysteme wird auch im Bereich der Forschung an einer Weiterentwicklung von Radarsystemen für den Automobilbereich gearbeitet. Einen guten Überblick über neuere Entwicklungen in diesem Bereich geben die Arbeiten von [Mende 99] und [Meinecke 01]. Neben den besprochenen Fernbereichs-Radarsensoren für die ACC-Anwendung gibt es auch vielfältige Entwicklungsaktivitäten von Radarsensoren für den Nahbereich (siehe z. B. <http://www.smartmicrowave.de>). Beispiele für andere Radarprinzipien als die hier genannten stellen z. B. bildgebende Radare dar [Giubbolini 97], [Cugiani, Giubbolini 98].

### 2.1.2 Lasersensorik

Obwohl Lasersensoren bereits zu Zeiten des europäischen Forschungsprogramms PROMETHEUS (Programme for a European Traffic with Highest Efficiency and Unprecedented Safety) zu Beginn der 90er Jahre im Rahmen der Entwicklung des AICC (Adaptive Intelligent Cruise Control) - dem Vorläufer der heutigen ACC-Seriensysteme - zum Einsatz kamen und gegenüber der aufwendigeren Radarsensorik Kostenvorteile besitzen, haben sie sich bislang nicht im gleichen Maße wie die Radarsysteme im Automobilbereich etablieren können. Dies hängt wohl vor allem mit den etwas ungünstigeren Eigenschaften im Hinblick auf die Umweltbedingungen, der nicht möglichen direkten Geschwindigkeitsmessung und dem Fehlen spezifischer Funktionen, welche die Vorteile von Lasersensoren ausnutzen, zusammen. Gleichwohl sind auch laserbasierte ACC-Systeme am Markt bzw. als Prototypen erhältlich (z. B. Denso, A.D.C., Spies, Hella). Ihren Ursprung haben die Lasersensoren vornehmlich im Bereich der Fabrikautomation, bei denen die Umgebung definiert, die Umweltbedingungen relativ konstant sowie Bauraum und Kosten nicht so bestimmend sind wie

im Automobilbereich. Spezielle Lieferanten von Lasersensoren vor allem für diesen Bereich sind z. B. die Firmen Riegl und Sick.

Das Meßprinzip von Lasersensoren ist das Pulslaufzeitverfahren, bei dem die Zeit zwischen dem Aussenden eines Pulses und dem Empfang des reflektierten Pulses gemessen wird. Die maximale Reichweite hängt dabei zum einen von der abgestrahlten Energie pro Puls, sowie der Empfindlichkeit des Empfängers und der (unbekannten) Reflektivität des Ziels ab. Begrenzender Faktor bezüglich der abgestrahlten Energie ist die geforderte Augensicherheit, für die vorgegebene gesetzliche Grenzwerte existieren. Als Wellenlänge kommt in den heute verfügbaren Systemen überwiegend nahes Infrarot im Bereich von  $0.9\mu\text{m}$  zum Einsatz. Einen weiteren Unterscheidungspunkt stellen Ein- oder Mehrstrahlsensoren sowie scannende oder feste Strahlanordnungen dar, die sich sowohl im Hinblick auf den Bauraum und die Kosten, als auch hinsichtlich des möglichen Erfassungsbereichs unterscheiden.

An der Universität der Bundeswehr in Hamburg (UniBwH) erfolgen seit längerem Entwicklungen auf der Basis von IBEO-Laserscannern. Diese aus der Weiterentwicklung von Sensoren zur Industrieautomation entstandenen Sensoren wurden von [Kirchner 00] eingesetzt, um in Straßenverkehrsszenen sowohl Objekte zu erkennen wie auch den Fahrbahnverlauf anhand ortsfester Objekte der Fahrbahnbegrenzung zu schätzen. Der Sensor ist dazu an der vorderen Stoßstange eines Versuchsträgers montiert und liefert im gesamten nicht verdeckten Bereich ein Meßprofil mit Winkelwerten sowie zugehörigen Entfernungswerten. Die Winkelschrittweite beträgt hierbei ca.  $0.25\text{-}0.35^\circ$  und die Meßunsicherheit ca. 50 cm bei einer Reichweite von über 100 m. Allein aufgrund der Baugröße ist dieser Sensor aber nicht für den automobilen Serieneinsatz geeignet.

An der Universität Ulm setzt man ebenfalls IBEO-Laserscanner ein, wobei diese über mehrere parallele Strahlebenen verfügen. Ein mögliches Einsatzziel stellt die Erkennung und Klassifikation von Fußgängern dar, wobei insbesondere die Beine und deren Bewegungsmuster als Erkennungsmerkmal herangezogen werden [Fuerstenberg et. al. 02].

Über diese Scannerprinzipien hinaus geht Dornier mit seiner Laser-Entfernungskamera, bei der es sich um einen faseroptischen Scanner handelt. Hierbei wird eine Scanebene über Spiegel realisiert, während die zweite Scanebene über ein Faserarray bereitgestellt wird. Der Sensor liefert in einem Detektionsbereich von  $60^\circ$  horizontal,  $31^\circ$  vertikal und einer Bildfolge von 4 Hz ein  $128 \times 63$  Pixel großes Bild. Die Reichweite beträgt maximal 60 m [Eibert et.al. 98]. Für einen Einsatz im Automobilbereich ist zum einen die Scanrate nicht hoch genug, zum anderen ist der Sensor zu groß, zu schwer und zu teuer.

Ein weiteres Meßprinzip realisiert die DASA (EADS) mit ihrer Laserkamera, welche ein kombiniertes Laufzeit/Intensitätsmeßverfahren einsetzt. Die Szene wird pulswise mit einem Laserscheinwerfer beleuchtet. Durch Aufnahme einer Bildfolge mit bzw. ohne Shutterunterstützung durch eine CCD-Kamera kann der reflektierte Laserpuls in einem bestimmten Meßvolumen bestimmt werden. Der horizontale Öffnungswinkel beträgt  $42^\circ$ , der vertikale Öffnungswinkel  $32^\circ$ . Das Meßergebnis besteht in einem Entfernungs- sowie einem Intensitätsbild mit halber Videoauflösung (VGA) mit 10 Hz. Dieses nicht-scannende System hat allerdings nur eine maximale Reichweite von ca. 30 m.

Im Rahmen von Fahrerassistenzsystemen werden Lasersensoren sowohl mit scannender

wie mit fester Anordnung im Rahmen von ACC-Systemen eingesetzt. Bedingt durch die funktionalen Anforderungen ergeben sich ähnliche Leistungsdaten wie bei den o. g. Radarsensoren.

Bei dem in dieser Arbeit eingesetzten Lasersensor handelt es sich um einen nichtscannenden Mehrstrahllaser, die Einzelheiten dazu werden in Kapitel 4 dargestellt.

### 2.1.3 Videosensorik

Gegenüber den vorgenannten Sensorprinzipien des Radars und des Lasers hebt sich die Videosensorik sowohl bzgl. der Datenmenge als des Verarbeitungsaufwands ab. Während ein Meßwert eines Radar- oder Lasersensors (nach einer Vorverarbeitung) im Hinblick auf die Umfelderkennung bereits eine gewisse Aussage darüber darstellt, daß sich an der gemessenen Position (und mit der gemessenen Geschwindigkeit - beim Radar) ein Objekt befindet, sagt ein Pixel eines Kamerachips zunächst nur etwas über die Helligkeit (bei Monochrom-Kameras) des einfallenden Lichts in einem bestimmten Raumwinkel aus. Ohne Hinzunahme weiterer Kameras (Stereo) oder zusätzlichen Wissens ist damit die Aussagefähigkeit der Sensorinformation sehr begrenzt. Umgekehrt ist aber auch das Potential der Szenenerkennung durch nachgelagerte Verarbeitung und Interpretation - anders als bei Radar- und Laserdaten - nahezu unbegrenzt - gleichwohl noch lange nicht gelöst.

Da Kameras prinzipiell die gleichen Informationen erfassen, wie das visuelle System des Menschen, sind sie nahezu ideal für Fahrzeugführungsaufgaben geeignet, da die vorhandene Verkehrsinfrastruktur gerade hierfür geschaffen wurde. Sie ist damit neben der aufwendigen Rohdatenverarbeitung, die viel Rechenleistung erfordert, den gleichen Einschränkungen wie der Mensch unterworfen, z. B. hinsichtlich der Wetterabhängigkeit.

Hinsichtlich der Bildsensoren kommen verschiedene Prinzipien zum Einsatz. Gegenüber den früher üblichen CCD-Sensoren gewinnen im Hinblick auf den Dynamikbereich und die Kosten CMOS-Kamerachips zunehmend an Bedeutung, was den Einsatz im automobilen Umfeld betrifft. Freiheitsgrade bestehen in der Wahl der Chipgröße sowie der verwendeten Optik, die in Kombination mit der Größe der wahrzunehmenden Objekte die Reichweite und den Öffnungswinkel bestimmen. Um die Limitierung des Gesichtsfeldes zu umgehen gibt es Lösungen zur aktiven Blickrichtungssteuerung bzw. bi- oder multifokale Kameraanordnungen. Eine spezielle Ausgestaltung der Bildchips gestattet auch eine Empfindlichkeit im Bereich des Nah- oder Ferninfrarotbereichs. Diese kann dazu genutzt werden, um auch bei Dunkelheit Objekte zu erkennen, indem man sie im Fall des Nahinfrarot aktiv beleuchtet oder im Fall des Ferninfrarots die Wärmestrahlung von Objekten direkt erfassen kann.

Die nachgelagerte Bildauswertung eines einzelnen Videobildes erfolgt fast immer modellbasiert, im Gegensatz dazu stellt eine Zwei-Kamera-Anordnung in Form einer sog. Stereo-Kamera im Prinzip einen eigenen Sensor dar, der durch Auswertung von Disparitäten zwischen den Bildern beider Kameras die Entfernung signifikanter Punkte im Bild bestimmt und damit ein den Radar- und Lidarsensoren vergleichbares Ergebnis liefert.

Die Videosensorik ist auch der einzige Sensor, der es derzeit gestattet, die Fahrbahn und die auf ihr befindlichen Fahrspuren zu erfassen. Über diese Anwendung sind auch bereits erste

Systeme als sog. Spurverlassenswarner - vornehmlich im LKW-Bereich - im Einsatz. Die in diesem Bereich in verschiedenen Versuchsfahrzeugen und Systemen realisierten Fähigkeiten ebenso wie die Fähigkeiten im Bereich der Objekterkennung werden nicht an dieser Stelle, sondern im folgenden Abschnitt 2.2 im Rahmen der Umfelderkennung in (semi-)autonomen Fahrzeugen behandelt.

## 2.2 Umfelderkennung in (semi-)autonomen Fahrzeugen

Im folgenden werden einige prominente Vertreter aus dem Bereich des (semi-)autonomen Fahrens exemplarisch vorgestellt, gefolgt von einem groben Überblick einiger weiterer weltweiter Aktivitäten. Auf eine erneute Darstellung der in der Einleitung bereits ausführlich erläuterten Aktivitäten an der Universität der Bundeswehr in München wird an dieser Stelle verzichtet.

### Daimler-Chrysler (vormals Daimler-Benz)

Parallel zu den Arbeiten an der Universität der Bundeswehr München wurden bei der damaligen Daimler-Benz Forschung zwei korrespondierende Versuchsträger aufgebaut. Beim 1989 aufgebauten VITA I handelte es sich um einen 7.5t Kastenwagen, der automatisch quer- und längsgerichtet auf der Straße fahren konnte. Als Sensoren dienten mehrere Kameras in einer für Autobahnfahrten ausgelegten Anordnung. Ein erweitertes System wurde 1994 im Rahmen der PROMETHEUS-Abschlußdemonstration mit dem VITA II, einem S-Klasse PKW, gezeigt [Ulmer 94a], [Ulmer 94b].

Neben den Arbeiten im Rahmen von PROMETHEUS erfolgten bei Daimler-Benz auch vielfältige andere Aktivitäten. Der PKW OSCAR wies 1994 die robuste Spurführung auch bei hohen Geschwindigkeiten nach. Mit dem LKW OTTO wurde 1995 im Rahmen des europäischen Chauffeur-Projekts die autonome Kolonnenfahrt mittels Kamera und Markierungen auf dem vorausfahrenden LKW nachgewiesen [Schulze 97].

Die Ausdehnung der bis dato zumeist auf den Autobahnbereich beschränkten Arbeiten begann 1998 mit dem PKW UTA (Urban Traffic Assistant) und in der Folge UTA II. Beide Versuchsträger sind für Innenstadtszenarien gedacht, welche sich typischerweise durch geringe Geschwindigkeiten und hohe Szenenkomplexität mit mehreren Fahrzeugen, Fußgängern, Verkehrszeichen und Ampeln auszeichnen. UTA war hierzu mit einem Stereo-Kamerasystem für die Frontsicht ausgerüstet, während bei UTA II neben mehreren Kameras auch ein Radarsensor hinzugenommen wurde und mit den Videosensoren fusioniert wurde [Franke et. al. 96], [Franke et. al. 97], [Franke et. al. 98a], [Franke et. al. 98b].

### CMU

An der Carnegie Mellon University in Pittsburgh wird ebenfalls seit langem auf dem Gebiet des autonomen Fahrens gearbeitet. Unter der Leitung von Charles Thorpe wurden



im Rahmen des zivilen AHS (Automated Highway System)- und des militärischen UGV (Unmanned Ground Vehicle)-Programms verschiedene Versuchsträger aufgebaut. Die Navlab genannten Versuchsträger waren mit verschiedenen Sensoren ausgerüstet; für die neueren Versionen Navlab 6...10 waren dies mehrere Kameras, Radarsensoren für den Front- und Seitenraum sowie ein Lasersystem für den rückwärtigen Raum. Die Fähigkeiten reichen von der automatischen Spurhaltung über die Hinderniserkennung und -vermeidung bis zum automatisch durchgeführten Spurwechsel. Details hierzu können den zahlreichen Veröffentlichungen, z. B. [Pommerleau 95], [Pommerleau 97], [Thorpe et. al. 97], [Batavia et. al. 97], [Batavia et. al. 98], [Jochem et. al. 95], [Williamson et. al. 98] entnommen werden.

## IITB

Am Fraunhofer Institut für Informations- und Datenverarbeitung wurde unter der Leitung von H.-H. Nagel und W. Enkelmann ein BMW 735 als Versuchsträger aufgebaut, der mit Kameras und einem Radarsensor ausgerüstet war, der über die Fähigkeiten der Spurhaltung auf Autobahnen, der Hinderniserkennung sowie der Detektion von Kreuzungen und Einmündungen verfügt [Krüger et. al. 95], [Enkelmann 97], [Willersinn et. al. 97], [Risack et. al. 98]. Ein weiterer Versuchsträger - DARVIN (Driver Assis using Realtime Vision for INnercity areas) - war für die Beherrschung innerstädtischer Verkehrsszenarien mit einem Stereokamerapaar ausgestattet [Nagel et. al. 98a], [Nagel et. al. 98b], [Heimes et. al. 98].

## Universität Parma

Die Gruppe um Alberto Broggi an der Universität Parma hat sich das autonome Fahren mit einfachen Hilfsmitteln zum Ziel gesetzt. Dies gilt zum einen für die Aktorik, die aus einem einfachen Elektromotor am Lenkrad besteht ebenso wie für die Sensoren, bei denen es sich um zwei Low-Cost Kameras handelt, deren Bilder auf einem Standard-PC ausgewertet werden. Als Versuchsträger dient ein Lancia Thema 2000, mit dem die optische Spurhaltung auf Autobahnen und die Hinderniserkennung demonstriert werden konnten [Bertozzi et. al. 97], [Bertozzi et. al. 98a], [Bertozzi et. al. 98b], [Broggi et. al. 99].

## Volkswagen

Ein ganz anderes Ziel als die Beherrschung des Autobahnszenarios verfolgte Volkswagen bei seinem Projekt Autonomes Fahren. Hier ging es darum, den für menschliche Testfahrer sehr belastenden Betrieb auf einem Teil des firmeneigenen Prüfgeländes zu Testzwecken zu automatisieren. Da es sich bei den Fahrzeugen um normale Serienfahrzeuge ohne spezielle Ausrüstung handelte, mußte sowohl Sensorik wie Aktorik entsprechend ausgelegt sein. Als Akteur diente dabei ein Fahrroboter, der anstelle des menschlichen Fahrers auf den Fahrersitz geschnallt werden konnte und sowohl Lenkung wie Gas, Bremse, Kupplung und Schaltung bedienen konnte. Der Hauptsensor zur Fahrzeugführung war auf dem definierten, vermessenen Prüfgelände ein hochgenaues DGPS-System, während die umfelderfassenden

Sensoren der Absicherung der Fahrt und dem Erkennen von Hindernissen dienen. Vor dem Hintergrund des beabsichtigten Einsatzes wurde insbesondere Wert auf überlappende Erfassungsbereiche mehrerer redundanter Sensoren unterschiedlicher Sensorprinzipien gelegt: (Stereo-)Video, Radar, zwei Laserscanner [Weisser et. al. 98].

## Sonstige weltweite Aktivitäten

Neben den soeben genannten Aktivitäten erfolgen zur Zeit bzw. erfolgten in den vergangenen Jahren eine Reihe weiterer Aktivitäten, die im folgenden nur knapp angerissen werden.

### Japan und Asien

In Japan sind vor allem die Automobilhersteller - teilweise in Kooperation mit dem MITI - sehr aktiv. Die Informationen über die laufenden Arbeiten sind allerdings eher verhalten. Bezüglich des autonomen Fahrens gab es bei Toyota Ansätze mittels Kameras und eines Radarsensors automatisch quergeregelt zu fahren und einem vorausfahrenden Führungsfahrzeug zu folgen [Tsugawa 93]. Nissan demonstrierte basierend auf Videosensorik eine automatische Querführung sowie ein automatisches Ausweichmanöver bei niedrigen Geschwindigkeiten [Tsugawa et. al. 93]. Einen Abstandsregeltempomaten in Verbindung mit einer automatischen Kollisionsvermeidung zeigte Mitsubishi [Mimuro et. al. 97]. Mazda [Yamamoto et. al. 99], Hyundai [Choi et. al. 95] [Kim et. al. 95] [Lee et. al. 95] und Subaru zeigten in Verbindung mit einer Kamera Systeme zur automatischen Querführung.

Darüberhinaus arbeiten die japanischen Hersteller im Rahmen des ASV (Advanced Safety Vehicle)-Programms an unterschiedlichen Konzepten von Fahrerassistenzfunktionen zur Erhöhung der Sicherheit, die auf dem Einsatz von Laserscannern (Honda [Kamiya et. al. 97], Mazda [Doi et. al. 97]), einer Kombination von Kameras mit Radar (Nissan [Obara et. al. 97]) oder eines Laserscanners mit mehreren Kameras (Mitsubishi [Mimuro et. al. 97]) basieren.

### USA

In den USA arbeitete neben der CMU noch die Ohio State University in Kooperation mit Honda an einem Radar- und Kamera-basierten System zur automatischen Abstandsregelung, der Spurhaltung und dem Spurwechsel [Özgüner et. al. 97] [Schneider et. al. 98]. Die University of California in Berkeley arbeitete ebenfalls zusammen mit Honda an der automatischen Querführung mittels Kameras [Kosecka et. al. 97] [Kosecka et. al. 98].

### Europa

In Europa ist als weiterer Vertreter das Institut für Neuroinformatik an der Ruhruniversität in Bochum zu nennen. Hier dient eine Kamera als Hauptsensor, die um einen Radarsensor ergänzt wird. Die Radarinformation dient primär zur Unterstützung bei der Segmentierung

des Videobildes im Rahmen der Fahrzeugerkennung [Handmann et. al. 98a] [Handmann et. al. 98b].

Als ein Automobilzulieferer ist die Firma Bosch auf dem Forschungsgebiet sehr aktiv. Es wurden Systeme zur Verkehrszeichenerkennung, zur automatischen Querführung oder - in Kombination mit einer videobasierten Objekterkennung - auch zur Stop & Go - Fahrt auf der Basis einer Stereo-Bildverarbeitung gezeigt [Stiller et. al. 97], [Stiller et. al. 98], [Goldbeck et. al. 98], [Lorei et. al. 99].

Bei BMW wiederum werden weniger die technischen Möglichkeiten in Richtung des autonomen Fahrens betrachtet, als vielmehr im Rahmen des Connected Drive Projekts Wege einer möglichst guten, komplementären Ergänzung der Fähigkeiten des menschlichen Fahrers mit einem maschinellen Assistenzsystem betrachtet [Wisselmann et. al. 04].

## 2.3 Informationsfusion

Wie dem vorhergehenden Abschnitt zu entnehmen ist, kommen in vielen der vorgenannten Projekte zum (semi-)autonomen Fahren Mehrsensoranordnungen vor. Eine der Hauptaufgaben besteht dann in der Datenfusion zwischen den verschiedenen Sensoren.

Das im Projekt Autonomes Fahren von Volkswagen zur Fusion der Sensoren eingesetzte Verfahren beschreibt [Becker 01]. Vor der eigentlichen Fusion auf Objektebene werden die Sensordaten zeitlich und räumlich ausgerichtet, ehe sie paarweise assoziiert werden. Die heterogene Fusion von Sensordaten hat auch die Arbeit von [Stüker 04] zum Inhalt. Hier werden die Daten mehrerer asynchron arbeitender Sensoren wie Fern- und Nahbereichsradar, mehrere Laserscanner und eine Videokamera auf Objektebene miteinander fusioniert. Der Schwerpunkt in dieser Arbeit liegt auf der Entwicklung von erweiterten Meßmodellen - insbesondere für Laserscanner - und modifizierten Filterverfahren bei wechselnden Aspektbedingungen der beobachteten Objekte.

[Hofmann et. al. 01] beschreiben ein auf dem 4D-Ansatz basierendes System mit einer Kombination aus einem Fernbereichsradar sowie einer Mehrkameraanordnung. Das Radar dient zur Detektion und zum Tracking der Objekte, während das Video zur Fahrspurerkennung und Objektverifikation herangezogen wird.

Ein Verfahren, das im Gegensatz zu den vorgenannten nicht auf dem Einsatz eines Kalman-Filters zur Informations-Fusion beruht, beschreibt [Coué et. al. 02], der einen Fusionsansatz auf der Basis Bayes'scher Programmierung beschreibt, der im europäischen Projekt Car-sense eingesetzt wird.

Enge Verwandtschaft mit dem in dieser Arbeit verfolgten Ansatz zur Datenfusion und Szeneninterpretation weist der in [Kaempchen et. al. 04] beschriebene Ansatz auf. Hier werden die Daten eines in 3 Ebenen scannenden Lasers mit den Daten einer Monovideokamera im Rahmen einer Low-Level-Fusion zusammengeführt. Hauptziel der Fusion ist die Verbesserung der Schätzung der lateralen Ablage im Fernbereich. Wesentlich ist die Assoziation von Bildbereichen mit den vom Laserscanner gelieferten Entfernungswerten in den 3 Erfassungsebenen. Je nach Entfernungsbereich werden verschiedene Verfahren einge-

setzt, um die Objekte im Videobild zu verifizieren: Im Fernbereich wird in einem Binärbild nach der Fahrzeugunterkante und den Reifen gesucht; im Nahbereich wird versucht, die seitlichen Objektkanten und bestimmte Symmetriemerkmale zu detektieren. Zur zeitlichen Verfolgung der Objekte im Videobild werden signifikante Merkmale betrachtet, wobei die Entfernungsinformation des Lasers als Stütze zur Skalierung der verfolgten Merkmale dient.

Im Hinblick auf den Einsatz von Mehrsensoranordnungen im Rahmen von Fahrerassistenzsystemen werden von verschiedener Seite auch Überlegungen zur Architektur solcher Anordnungen, zur Abstraktion von konkreten Sensoren und zur strukturellen Gestaltung einer Aufmerksamkeitssteuerung angestellt [Naab 04], [Darms, Winner 04].



# Kapitel 3

## Maschinelle und Biologische Wahrnehmung

Die im vorhergehenden Kapitel genannten Verfahren zur Szenenerkennung haben die Vielfalt möglicher Ansätze zur Umfelderkennung und Szeneninterpretation bereits angedeutet. Unabhängig von der Herangehensweise ist eine der Kernfragen in den Diskussionen stets, inwieweit sich technische Ansätze zur Objekterkennung am menschlichen Wahrnehmungsprozeß orientieren sollten. Dieser Frage wird allgemein noch zugestimmt, da zum einen der Mensch in diesem Bereich offensichtlich außerordentliches leistet und andererseits bei der technischen Entwicklung jeder Mensch aus der eigenen Erfahrung schöpfen kann. Allerdings ist gerade die visuelle Wahrnehmung des Menschen aufgrund ihrer Komplexität noch immer nur zu einem Bruchteil erforscht. Dieser Umstand bildet den ersten Ansatz für unterschiedliche Systeme. Die strukturellen Ansätze versuchen, menschliche Gehirnstrukturen in Form neuronaler Netze - erheblich vereinfacht - technisch nachzubilden. Im Gegensatz dazu sind viele analytische Ansätze auf wesentlich abstraktere Weise an den menschlichen Wahrnehmungsprozeß angelehnt, indem sie zwar Verarbeitungsstufen nachempfinden, hierzu jedoch analytische, ingenieurmäßige Methoden heranziehen.

In diesem Abschnitt soll aus der Sicht des Autors heraus der (visuelle) Wahrnehmungsprozeß in seinen für die Objekterkennung relevanten Aspekten exemplarisch dargestellt werden. Dabei soll weniger die Diskussion psychologischer / physiologischer Forschungsergebnisse im Mittelpunkt stehen, denen man in der hier gebotenen Kürze nicht gerecht werden kann - als vielmehr die Darstellung eines plausiblen Gesamtansatzes, der die technische Umsetzung motiviert. Da der Mensch - im Gegensatz zu anderen Lebewesen (z. B. Ultraschall bei Fledermäusen) - zur geometrisch-räumlichen Szenenerfassung lediglich das visuelle System besitzt, beschränkt sich die Diskussion im folgenden auch auf den Aspekt der visuellen Wahrnehmung. Im Zusammenhang mit dem Übergang auf das maschinelle Sehen treten dann auch andere Sensortechnologien wieder in den Fokus.

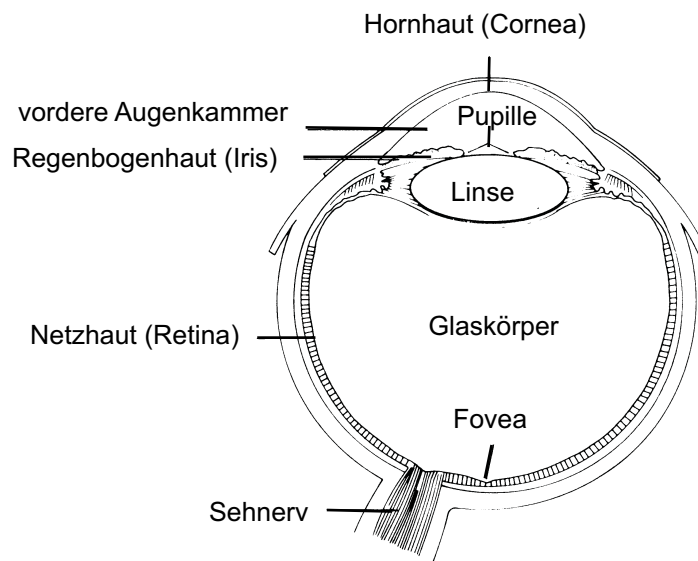


Abbildung 3.1: Schnittdarstellung des menschlichen Auges

## 3.1 Menschliche Wahrnehmung

### 3.1.1 Das Wahrnehmungssystem

Den Ausgangspunkt der Diskussion soll der visuelle Wahrnehmungsapparat des Menschen bilden. Dieser sei durch die folgenden Eigenschaften grob - im Rahmen dieser Diskussion hier jedoch ausreichend genau - charakterisiert:

- Die Netzhaut des menschlichen Auges (Abb. 3.1) besteht aus  $120 \cdot 10^6$  die Helligkeit wahrnehmenden Stäbchen und  $6 \cdot 10^6$  Zäpfchen, welche für die Farbwahrnehmung verantwortlich sind. Aufgrund der unterschiedlichen Dichte ist die Ortsauflösung im Farbbereich wesentlich schlechter als im Grauwertbereich.
- Die Abbildung durch das menschliche Auge ist erheblich verzerrt. Erst durch nachgeschaltete Verarbeitungsebenen entsteht ein korrigiertes "unverzerrtes" Bild.
- Die durch die Augen gebildete Stereobasis beträgt lediglich ca. 7 cm und ist damit nur bis zu einer geringen Entfernung von unter 10 m nutzbar. (Das entspricht etwa dem unmittelbaren Aktionsradius des Menschen.)
- Das menschliche Auge kann nur einen sehr begrenzten Ausschnitt im Sehzentrum fokussieren, der restliche Sehbereich wird nur undeutlich wahrgenommen. Auf den Bereich des Sehzentrums entfällt aber nur ein geringer Teil der visuellen Verarbeitungsleistung.

Ergänzt werden diese Eigenschaften des visuellen Wahrnehmungsapparates um die allgemeinen Gedächtnis- und Abstraktionsfähigkeiten des Menschen. Diese sind es auch, die

in Kombination mit obigen Eigenschaften den Gedanken nahelegen, daß der Mensch die ihn umgebende Welt nur mittelbar in Form einer inneren Vorstellungswelt erlebt: Da der Mensch die ihn umgebende Welt nur mittelbar durch Vermittlung seiner Sinnesorgane wahrnimmt, kann er nach [Kant] die Dinge und die Welt “an sich“ nicht erfassen, sondern jeder Mensch hat seine eigene Vorstellungswelt. Der Austausch mit anderen Menschen über diese Vorstellungswelt erfolgt über zu definierende, allgemeinverbindliche Begriffe. Die Vorstellungswelt entsteht zum einen aus der Summe der von den verschiedenen Sinnesorganen zu einem Zeitpunkt aufgenommenen Eindrücken und zum anderen aus den Erfahrungen sämtlicher vergangener Sinneseindrücke. So lernt der Mensch im Laufe seines Lebens etwa, daß Glas durchsichtig ist, oder wie der Schattenwurf bei gegebenem Sonnenstand ausfällt. Auch, daß die verzerrt durch das Auge aufgenommenen Kurven in Wirklichkeit Geraden sind und von uns auch als solche erkannt werden, gehört in diese Kategorie.

Aus diesem Modell resultieren darum die Fragen:

- “Wie werden die Dinge unserer Umwelt von uns intern repräsentiert ?“ und
- “Anhand welcher Merkmale erkennen wir diese Dinge ?“.

In bezug auf die Repräsentation wird heute davon ausgegangen, daß in Abhängigkeit der Anforderungen sowohl räumliche (regelmäßige Körper) als auch piktoriale (Gesichter) Repräsentationen vom Menschen verwendet werden. Die bzgl. der Hell-Dunkel-Verarbeitung wesentlich ausgeprägtere Empfindlichkeit legt die Vermutung nahe, diese liefere die für die Erkennung signifikanteren Merkmale. Die - im Laufe der Evolution erst spät entwickelte - Farberkennung ergänzt diese lediglich und dominiert die Wahrnehmung nur in Einzelfällen.

### 3.1.2 Wahrnehmung statischer Szenen

Die Interpretation eines statischen (Einzel-)Bildes kann etwa verglichen werden mit einem Schnappschuß der realen Welt, beispielsweise in Form eines Fotos. Auch beim Betrachten eines Fotos können wir nicht die ganze auf dem Foto festgehaltene Szene in einem “Augenblick“ erfassen, sondern werden von den am meisten “ins Auge springenden“ Szeneninhalten ausgehend nach weiteren zu unserer bisher aufgebauten Vorstellung passenden Merkmalen suchen. Diese Interpretation kann von einer Initialerwartung ausgehen, wenn Vorabinformationen über den Bildinhalt vorliegen (“Foto am Meer“), bei denen bereits zu Beginn nach bestimmten Merkmalen gesucht wird (“Wasser“). Bei dieser Vorgehensweise wird mit jedem neu entdeckten Szenenmerkmal versucht, dieses möglichst widerspruchsfrei in die bisherige Szenenvorstellung zu integrieren. Dabei werden i. a. sowohl die Bedeutungen bereits entdeckter als auch des gerade entdeckten Merkmals adaptiert werden müssen. Analog zur Betrachtung eines Fotos kann man sich viele Situationen mit ähnlichem Interpretationsschema vorstellen, wobei es sich dabei meist um sog. Aufwachszenarien handelt, also Situationen in denen man über einen längeren Zeitraum keine visuelle Informationen über seine Augen aufnimmt. Das Vorwissen bei der Szeneninterpretation nach dem Öffnen der Augen wird u. a. davon bestimmt, ob und inwieweit wir uns in dieser Zeit gegenüber unserer Umwelt bewegt und damit den Szeneninhalt geändert haben. Das Vorwissen wird desto geringer sein, je unbewußter wir diese Bewegung erlebt haben (z. B. im Schlaf).



Anhand dieses Beispiels sollte verdeutlicht werden, wie sich aus den konkurrierenden Deutungsmöglichkeiten resultierend aus:

- dem Vorwissen über die Szene und die Objekte in der Szene,
- den neu entdeckten Merkmalen/Merkmalstruppierungen und,
- den bereits entdeckten und in die Interpretation integrierten Merkmalen

eine konsistente Interpretation der Szene entwickelt. Die Gewichtung der einzelnen Möglichkeiten hängt vom jeweiligen Vertrauen in diese ab. Je signifikanter eine neu entdeckte Merkmalsgruppierung ist, desto größer wird der Einfluß auf die Interpretation sein. Andererseits wird das Vertrauen in eine Merkmalsgruppierung desto ausgeprägter sein, je besser sich diese in die Interpretation einfügt.

Das Beispiel der Betrachtung und Interpretation eines Fotos umfaßt noch einen weiteren Aspekt statischer Szenen. Da ein Foto nur ein zweidimensionales Abbild einer dreidimensionalen Szene ist, kann die räumliche Tiefe der abgebildeten Szene nicht dem Foto als solchem entnommen werden. Dennoch haben wir beim Betrachten eines Fotos eine Vorstellung von der räumlichen Tiefe der Szene und den räumlichen Dimensionen der darin enthaltenen Objekte. Dies ist ein Indiz dafür, daß wir Menschen eine räumliche Vorstellung einer Szene aufbauen, und daß wir dazu auch dann in der Lage sind, wenn die Szene von uns gar nicht räumlich wahrgenommen werden kann, wie im Falle des Fotos. Aber auch, wenn es sich nicht um ein Foto, sondern um eine reale statische Szene (ohne Bewegung) handelt, haben wir vom größten Teil der Szene nur ein zweidimensionales Abbild. Die Raumtiefe, bis zu welcher die durch das menschliche Augenpaar gebildete Stereobasis Tiefeninformationen liefert beträgt wie o. a. lediglich ca. 10 m. Es muß also andere Interpretationsmechanismen geben, anhand derer wir auf die räumliche Gestalt von Objekten in einer Szene schließen.

Aus der Gestalttheorie [Wertheimer 23] ist bekannt, daß der Mensch bei der Erkennung von Objekten zunächst einfache geometrische Formen (Kreise, Rechtecke, etc.) bzw. Verhältnisse (rund, rechtwinklig, etc.) bevorzugt. Darüberhinaus gibt es eine Reihe von “Gesetzen“, nach denen Menschen Merkmale extrahieren oder zu höheren Merkmalen gruppieren, wie insbesondere die Arbeit von [Lowe 85] gezeigt hat oder in neuerer Zeit auch die Arbeit von [Hoffman 98] mit dem provozierenden Untertitel “How we create what we see“.

Hat sich auf der Basis dieser Mechanismen letztlich eine Deutung eines statischen Bildes ergeben, so ist damit auch ein bestimmter Interpretationsfreiraum ausgeschöpft, der in Abhängigkeit von Vorwissen und Merkmalssignifikanz begrenzt wird. In vielen Fällen reichen die in einem Bild enthaltenen Informationen jedoch nicht aus. Diese Beschränkung kann bewußt oder auch unbewußt wahrgenommen werden, wenn sich bestimmte Widersprüche anhand eines Bildes nicht beheben lassen. Auf diesen unbewußten Interpretationsmechanismen beruhen auch viele optische Täuschungen, wie sie beispielsweise in [Zimbardo 83] zu finden sind.

### 3.1.3 Wahrnehmung dynamischer Szenen

Die zuvor genannten Widersprüche lassen sich erst durch Einbeziehen zusätzlicher Perspektiven einer Szene auflösen. Solche zusätzlichen Perspektiven lassen sich jedoch nur durch eine Stereobasis oder durch eine Relativbewegung zwischen Szene und Beobachter erhalten. Ersterer Fall soll aufgrund der geringen Stereobasis des Menschen hier nicht relevant sein, letzterer Fall bedeutet den Übergang von der statischen zur dynamischen Szeneninterpretation. Damit die Hinzunahme weiterer Perspektiven jedoch sinnvoll genutzt und also in die Szeneninterpretation einfließen kann, müssen die jeweils sichtbaren Objekte und deren Merkmale in den einzelnen Perspektiven einander zugeordnet werden, d. h. es müssen Korrespondenzen zwischen den einzelnen Abbildungen hergestellt werden. Die Bildung dieser Korrespondenzen zählt zu den Hauptproblemen der Umfeldwahrnehmung. Sie wird umso leichter sein,

- je geringer die relativen Änderungen der Aspektbedingungen sind und
- je weniger abhängig die Merkmale von den Aspektbedingungen (Invarianten) sind.

Um auf das Beispiel mit den Fotos zurückzukommen, wäre eine Möglichkeit, ein zweites Foto eines Szeneninhalts, welches unter einer gänzlich anderen Perspektive entstanden ist, hinzuzunehmen. Sofern das Wissen vorhanden ist, daß es sich um den gleichen Szeneninhalt handelt, wird man versuchen, im zweiten (später erhaltenen) Foto bestimmte signifikante Merkmale des ersten Fotos "wiederzuentdecken". Je mehr Informationen man über die Szenenänderung besitzt, desto ausgeprägter wird die Erwartungshaltung aufgrund der vom Menschen anhand des ersten Fotos aufgebauten Szenenvorstellung sein. Wesentlich hierbei ist der inkrementelle Charakter des Erkennungsvorgangs, demzufolge die Auswertung des zweiten Fotos unter Bezugnahme auf die bereits erhaltene Interpretation des ersten Fotos erfolgt.

Wenn es sich statt der Fotos wieder um reale Szenen handelt, kann die Änderung von Perspektivbedingungen auf der Basis von Eigenbewegungen des Beobachters oder von Fremdbewegungen der Objekte in der Szene erfolgen. Durch diese Einbeziehung von Bewegungen in die Szeneninterpretation tritt die Zeit als zusätzliche in der Repräsentation zu berücksichtigende Dimension neben die räumliche Repräsentation. Um die Szene auf der Basis der durch die Eigenbewegung des Beobachters entstehenden Aspektänderungen deuten zu können, muß der Beobachter seine Eigenbewegung kennen. Diese Tatsache wird zwar jedem als trivial erscheinen, wenn man sich jedoch überlegt, aufgrund welcher Informationen wir unseren eigenen Bewegungszustand wahrnehmen, so ergeben sich drei Quellen:

- das Wissen über den Zustand unserer Fortbewegungsorgane
- die sensorischen Informationen unseres Gleichgewichtssinnes
- die visuellen Informationen des sich durch die Bewegung ändernden Blickfeldes.

D. h. aber nichts anderes, als daß dieselben visuellen Informationen, anhand derer wir unsere Umwelt interpretieren, auch mit zur Schätzung des Eigenzustands des Beobachters herangezogen werden.

Ebenso wie die visuellen Informationen zur Schätzung des Eigenzustands dienen, können aus ihnen auch die Zustände anderer bewegter Objekte in der Szene ermittelt werden, von denen wir nach Newton in jedem Fall wissen, daß die Objekte als massebehaftete Körper bestimmten physikalischen Bewegungsgesetzen gehorchen. In Verbindung mit entsprechendem Hintergrundwissen über die für bestimmte Objekte geltenden Bewegungsgesetze (z. B. können Autos sich nicht seitwärts bewegen und Menschen ohne Hilfe nicht fliegen), können dann auch diese Objektbewegungen in die Interpretation einbezogen werden.

Wenn wir also davon ausgehen, daß wir eine Vorstellung über den eigenen Bewegungszustand und den aller Objekte haben, so können wir von einem Moment zum nächsten Annahmen über die durch die Bewegungen in der Szene auftretenden Veränderungen treffen, und diese entsprechend bei der Interpretation berücksichtigen. In diesem Zusammenhang tritt nun neben die bereits o. g. inkrementelle Interpretation der räumlichen Dimension jene der zeitlichen Dimension, denn auch unsere Vorstellungen über die Bewegungszustände der Objekte ist nur bedingt genau. In bezug auf das Eigenobjekt kann sie gestützt werden durch die zusätzlichen Informationen der anderen Sinnesorgane, in bezug auf Objekte kann deren Bewegung nur wesentlich ungenauer geschätzt werden, da sie:

- uns nicht bekannten äußeren Störeinflüssen (passive und aktive Objekte)
- uns nicht bekannten inneren Steuereinflüssen (aktive Objekte)

unterliegen. Mit “aktiven“ Objekten sind im Gegensatz zu “passiven“ Objekten in diesem Zusammenhang jene Objekte gemeint, die ihre Lage aus eigenem Antrieb heraus verändern können.

### 3.1.4 Aufmerksamkeits-/Blickrichtungs-Steuerung

Sowohl bei statischen als auch bei dynamischen Szenen kann die volle Aufmerksamkeit wie o. a. nur auf einen kleinen Bildbereich gerichtet werden. Die Kriterien, nach denen die Aufmerksamkeit gesteuert wird, bilden einen umfassenden eigenen Themenkomplex. Die Verfolgung der Aufmerksamkeitssteuerung eines Menschen beispielsweise gestattet wesentliche Rückschlüsse auf die Interpretation einer Szene. Dabei konkurrieren signifikante Merkmale im Bild mit der modellgesteuerten Merkmalssuche um die Aufmerksamkeit des Beobachters. Wohin diese gelenkt wird, hängt von der angenommenen Relevanz in bezug auf das Ziel des Beobachters ab. I. a. wird sich die Aufmerksamkeit insbesondere auf folgende Szenenbestandteile konzentrieren:

- Objekte, in bezug auf die sich die Aspektbedingungen geändert haben zur Erzielung einer besseren räumlichen Vorstellung.
- Objekte, denen gegenüber eine erhebliche Relativbewegung oder Änderung derselben besteht.
- Objekte, die aufgrund von Relativbewegungen an der Peripherie in das Blickfeld eintreten.

- Objekte, die aufgrund von Relativbewegungen aus dem Bildzentrum heraus expandieren.
- Objekte, deren Bedeutung sich in bezug auf das Ziel des Beobachters geändert hat.

Eine Vorstellung über Form und Bewegung eines Objektes kann nur auf der Basis der Blickzuwendung und der Relativbewegung zwischen Objekt und Beobachter entstehen. Damit können alle im Blickfeld eines Beobachters befindlichen für diesen relevanten Objekte als in der Szenenvorstellung des Beobachters repräsentiert angesehen werden. Neu in das Blickfeld eintretende Objekte werden - entsprechend ausgeprägte (signifikante) Merkmale und Bildverschiebungen vorausgesetzt - durch Blickzuwendung fokussiert und in das zeitlich-räumliche Szenenmodell integriert werden.

### 3.1.5 Bewegungssteuerung

Die im vorhergehenden Abschnitt diskutierten Möglichkeiten zur Aufmerksamkeits- und Blickrichtungssteuerung können nur in dem Maß zur Szenenerkennung beitragen, wie dies unter den aktuellen Aspektbedingungen möglich ist. Viele bedeutungsrelevante Szeneninhalte werden aber erst dann für die Wahrnehmungsorgane sichtbar, wenn sich zuvor die Aspektbedingungen geändert haben. Um also eine Szene vollständig zu explorieren, kann es daher oftmals sinnvoll sein, gezielt eine Eigenbewegung durchzuführen, welche zu geänderten Aspektbedingungen in bezug auf den relevanten Szenenausschnitt führt und damit bisher nicht beobachtbare Größen zugänglich macht.

Im Gegensatz zu der "normalen" hierarchischen Abfolge, bei welcher die Ziele der Eigenbewegung maßgeblich sind für die Wahrnehmung und deren Steuerung, verkehrt sich im Falle dieser sog. Szenenexploration die Rangfolge in ihr Gegenteil: Die Wahrnehmung gibt der Bewegungssteuerung die Ziele vor, um die Szene optimal erfassen zu können [Pellkofer 03].

## 3.2 Maschinelle Wahrnehmung

Nach dieser allgemeinen Beschreibung des visuellen Erkennungsprozesses und seines Umfeldes aus der Sicht des Autors sollen nun Ansätze aus dem Bereich der Technik diskutiert werden, welche insbesondere visuelle Erkennungsaufgaben zum Ziel haben. Im Mittelpunkt soll dabei der an der Universität der Bundeswehr München verwendete Ansatz zum dynamischen maschinellen Sehen stehen.

### 3.2.1 Klassischer 4D-Ansatz zur visuellen Wahrnehmung

Den Kern der Bildinterpretation bildet bei diesem Ansatz ein Weltmodell, welches die Welt als aus Objekten dreidimensionaler Gestalt bestehend modelliert, die aufgrund ihrer Masse den physikalischen Bewegungsgesetzen über der Zeit unterworfen sind. Das durch Körper (3D) und Zeit (1D) gegebene Modell wird deshalb auch als 4D-Weltmodell bezeichnet. Über

diese Annahme eines Weltmodells hinaus beschreibt der sog. 4D-Ansatz [Dickmanns 87] die Art und Weise, wie sich dieses Weltmodell auf der Basis 2-dimensionaler visueller Informationen ausprägt. Der Erkennungsvorgang hat dabei nach Abb. 3.2 rekursiven Charakter. Ausgehend von der 4D-Weltvorstellung, die der Beobachter zu einem Zeitpunkt besitzt,

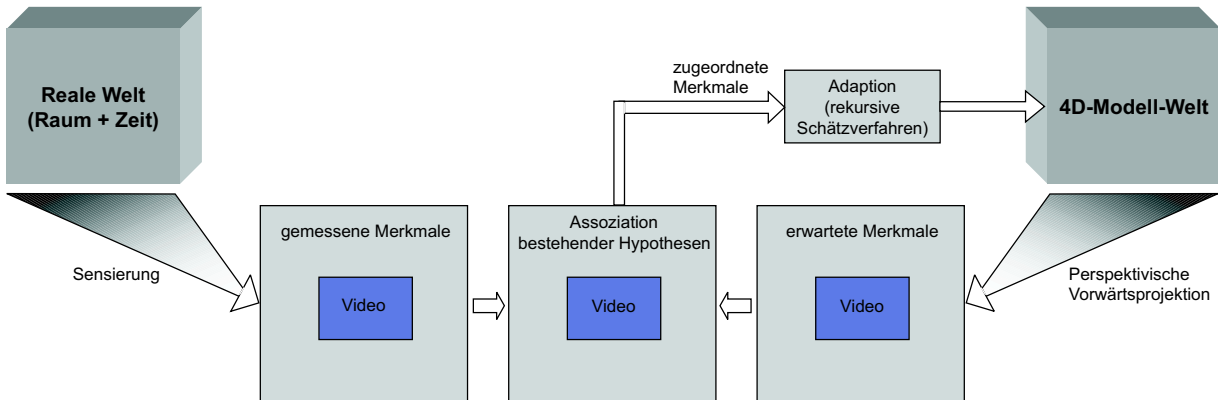


Abbildung 3.2: 4D-Ansatz zum dynamischen maschinellen Sehen (nach Dickmanns)

kann er bei Kenntnis der geometrischen Abbildungseigenschaften anhand seiner Modellvorstellung durch Vorwärtsprojektion der Szene ein imaginäres 2D-Modellbild konstruieren, welches die Erwartung des Beobachters über das von seinen Sinnesorganen im nächsten Augenblick aufgenommene Bild wiedergibt. Aus dem Vergleich zwischen den Bildmerkmalen im konstruierten Modellbild (Modellmerkmale) und dem im aufgenommenen Bild tatsächlich gemessenen Merkmalen (Datenmerkmalen) ergeben sich Diskrepanzen, die von Fehlannahmen über die Form oder Bewegung von Objekten in der Szene oder Fehler in bildaufnehmenden Sensoren herrühren können. Nach einer Zuordnung der erwarteten zu den gemessenen Merkmalen werden nun lediglich diese Abweichungen zwischen erwarteter und gemessener Merkmalslage über rekursive Schätzverfahren auf die Modellvorstellung zurückgekoppelt und somit zu einer Korrektur derselben benutzt. Das herausragende Kennzeichen, welches die Eleganz dieses Ansatzes ausmacht, ist diese direkte Rückkopplung der Meßwertabweichungen. D. h. es findet keine Interpretation isolierter 2D-Bildinformationen und damit keine Inversion der Abbildungsgleichungen statt, die erst anschließend zueinander in Beziehung gesetzt werden, sondern die Interpretation von Diskrepanzen zwischen Erwartung und Messung findet unter direkter Bezugnahme auf das Weltmodell statt und steht damit in Übereinstimmung mit der im vorhergehenden Abschnitt genannten rekursiven Szeneninterpretation der menschlichen Wahrnehmung.

Der 4D-Ansatz hat sich - wie in Kapitel 1 beschrieben - im Einsatz zur autonomen Führung und Eigenzustandsschätzung von Straßenfahrzeugen, bei autonomen Landeanflügen von Luftfahrzeugen und zur Verfolgung einfacher Objekte in Straßenverkehrsszenen am ISF seit Jahren bewährt. Ein wesentlicher Vorteil des 4D-Ansatzes ist in dem - im Vergleich zu anderen Verfahren - geringen Rechenaufwand zu sehen. Insbesondere in den Anfangsjahren der autonomen Fahrzeugführung konnten mit - im Vergleich zu heute - bescheidener Rechenleistung, autonome Fahrten mit bis zu 96 km/h bei Auswertung von 25 Bildern/s durchgeführt werden. Wesentlichen Anteil daran hatte auch die intelligent gesteuerte Merkmalsextraktion, die lediglich in den aus der Prädiktion der Modellvorstellung sich ergebenden

den relevanten Bildbereichen mit adaptierten Meßverfahren vorgenommen wurde. Damit diese Bereiche möglichst klein, und die Korrespondenzbildung möglichst eindeutig waren, war eine hohe Bildfrequenz erforderlich, da nur dann die Änderungen zwischen den einzelnen Bildern der Bildfolge relativ gering sind.

Aus dieser Beschreibung des 4D-Ansatzes folgt, daß er zunächst “nur“ zur zeitlichen Verfolgung bereits bekannter Objekte, d. h. rein modellgetrieben, sinnvoll einzusetzen ist. Sowohl die Auswahl und Korrespondenzfindung der Merkmale, die Bildung von Modellen und die Initiierung und Verifizierung von Objekthypothesen sind eigene Komponenten des Erkennungsprozesses, die sich zur räumlich-zeitlichen Modellierung der Methoden des 4D-Ansatzes bedienen.

### 3.2.2 Verallgemeinerter 4D-Ansatz zur multisensoriellen Wahrnehmung

Wie im vorangegangenen Abschnitt dargestellt wurde der 4D-Ansatz ursprünglich für den Einsatz im Rahmen der Bildfolgenverarbeitung konzipiert. Dies wurde unter anderem durch die zum Zeitpunkt seiner erstmaligen Vorstellung allgemein verbreiteten Vorgehensweisen zur Bildfolgenauswertung unter Berücksichtigung der dafür erforderlichen Rechenleistung und im Hinblick auf die damals (1987) verfügbare Rechenleistung befördert. In der Zwischenzeit hat sich der rekursive räumlich-zeitliche Ansatz zur Szeneninterpretation nicht allein im Bereich der Bildfolgenauswertung auf breiter Front durchgesetzt, sondern ist - getrieben durch die gerade im automobilen Bereich vielfach eingesetzten andersgearteten Sensorprinzipien - auch auf beliebige andere Sensoren zur Umfelderkennung ausgedehnt worden. In einem Zug mit der Erweiterung läßt sich auch die anfangs offen gebliebene Frage nach der Instanziierung der initialen Hypothese vor dem Beginn der rekursiven Schätzung integrieren. Eine mögliche Darstellung des verallgemeinerten 4D-Ansatzes ist in Abb. 3.3 gegeben. Wie man erkennt, sind alle Module, in welchen Merkmale auftreten, jetzt neben dem Block für den Videosensor um Verarbeitungsblöcke für weitere Sensoren - in diesem Beispiel für einen Lidar- und einen Radarsensor - ergänzt. Diese Ergänzung ist nicht nur in dem Sinne zu verstehen, daß jeder Sensor seine jeweils eigenen Merkmale vermisst, sondern in Verbindung damit sich auch die jeweiligen Verarbeitungsblöcke in den Modulen voneinander unterscheiden, d. h. auf jeden Sensor angepaßt sind. Die im Bild 3.2 noch als “Perspektivische Abbildung“ bezeichnete Transformation vom Zustandsraum der 4D-Modellwelt in den Merkmalsraum des Videosensors wird verallgemeinert als Meßmodell bezeichnet, welches individuell für jeden Sensor die Transformation in den ihm eigenen Merkmalsraum durchführt. Der Vergleich der erwarteten Modell-Merkmale mit den tatsächlich gemessenen Daten-Merkmalen erfolgt im Modul “Assoziation bestehender Hypothesen“ ebenfalls für jeden Sensor separat. Der Filterschritt, in dessen Rahmen die Adaption der bestehenden Hypothesen an die mit erwarteten Merkmalen assoziierten gemessenen Merkmale erfolgt, ist hingegen für alle Sensoren identisch, da hier keine sensorspezifischen Informationen mehr benötigt werden. Im Gegensatz dazu ist das Modul “Instanziierung neuer Objekthypothesen“ wiederum sensorabhängig. Es hat unter Zuhilfenahme geeigneten Hintergrundwissens die Aufgabe, aus nicht mit bereits bestehenden Objekthypothesen assoziierten Merkmalen bzw. im Falle des Videosensors speziell daraufhin gemessenen Merkmalen neue Objekthy-

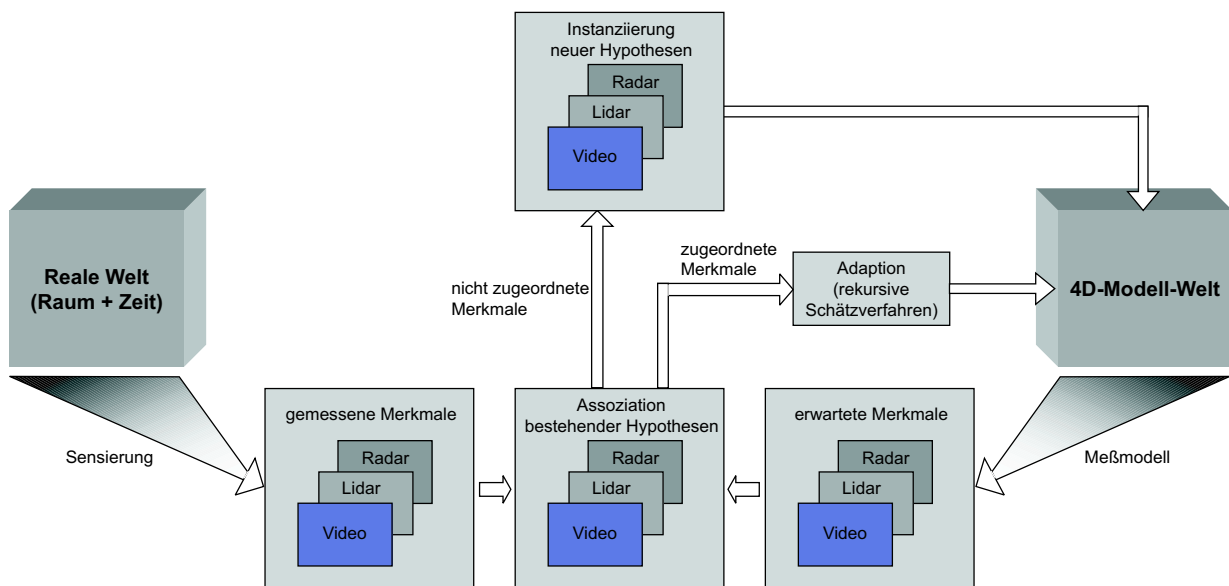


Abbildung 3.3: Verallgemeinerter 4D-Ansatz zum multisensoriellen maschinellen Sehen

pothesen zu generieren. Wie im weiteren Verlauf der Arbeit noch deutlich wird, führt diese Verallgemeinerung des klassischen 4D-Ansatzes um andere Sensortypen sowie die explizite Hereinnahme der Instanziierung als zusätzliches Modul zu einem insgesamt konsistenten Gesamtrahmen für die Szenenerkennung mittels beliebiger umfelderfassender Sensorik.

## Teil II

# Multisensorielle maschinelle Wahrnehmung





# Kapitel 4

## Sensoren

In diesem Kapitel werden die während der Versuche eingesetzten Sensoren beschrieben. Wie in der Einleitung bereits erwähnt, handelt es sich bei den umfelderfassenden Sensoren um eine monochrome Videokamera in Kombination mit einem mehrstrahligen Lidarsensor. Diese Kombination stellt einen Kompromiß dar aus technisch-wissenschaftlich wünschenswertem und technisch mittelfristig realisierbaren Wegen. Technisch-wissenschaftlich gesehen wäre der Einsatz sowohl mehrerer Kameras unterschiedlicher Brennweite ebenso wie der Einsatz eines scannenden Lasersystems mit höherer Winkelauflösung und mehreren Scanebenen wünschenswert gewesen. Aus technisch-realiserbarer Sicht - im Sinne eines möglichen Serieneinsatzes - wäre die Tendenz eher in die Richtung eines Radarsensors anstelle des Mehrstrahllasers gegangen.

### 4.1 Video

In den Versuchsfahrzeugen VaMP und insbesondere VaMoRs gelangen zum Zwecke des autonomen Fahrens komplexe, multifokale Kamerasysteme zum Einsatz [Rieder 00], [Pellkofer 03]. Um mit den eingesetzten großen Brennweiten trotz des damit verbundenen geringen Öffnungswinkels weite Raumbereiche abdecken zu können, sind die Kameras auf einer Zweiachs-(VaMoRs) bzw. Einachs-(VaMP) Plattform montiert. Diese sind u. a. in der Lage, schnelle Sakkaden und eine Stabilisierung in der Nicklage durchzuführen. In der o. g. Arbeit von Rieder wird u. a. eine auf der Plattform befindliche divergente Stereokameraanordnung betrachtet. Die Steuerung der Blickrichtung erfolgt priorisiert durch die in [Pellkofer 03] beschriebene Komponente zur Aufmerksamkeitssteuerung.

In dieser Arbeit wird in Verbindung mit der Lidarsensorik und der industriellen Zielrichtung eine einfache Anordnung mit nur einer festen Kamera mittlerer Brennweite betrachtet. Die Notwendigkeit der Aufmerksamkeitssteuerung im Sinne von Verteilung der Rechenressourcen bleibt aber auch hierbei bestehen.

Bei dem Videosensor wird eine Standard Schwarzweiß-CCD-Kamera der Marke TELI eingesetzt. Hierbei handelt es sich um eine Kamera mit einem 1/3" CCD-Chip, der über eine

Größe von  $752 \times 582$  aktiven Pixeln verfügt, wobei die Pixel mit  $6,5 \times 6,25 \mu\text{m}$  nahezu quadratisch sind. Unter Verwendung eines 16 mm Objektivs ergibt sich ein horizontaler Öffnungswinkel von

$$\alpha = 2 \arctan \frac{b_{\text{Chip}}}{2f} = 2 \arctan \frac{5,2 \text{ mm}}{2 \cdot 16 \text{ mm}} \approx 18,5^\circ \quad (4.1)$$

und damit eine Auflösung von  $0,02^\circ/\text{Pixel} = 0,4 \text{ mrad}/\text{Pixel}$ .

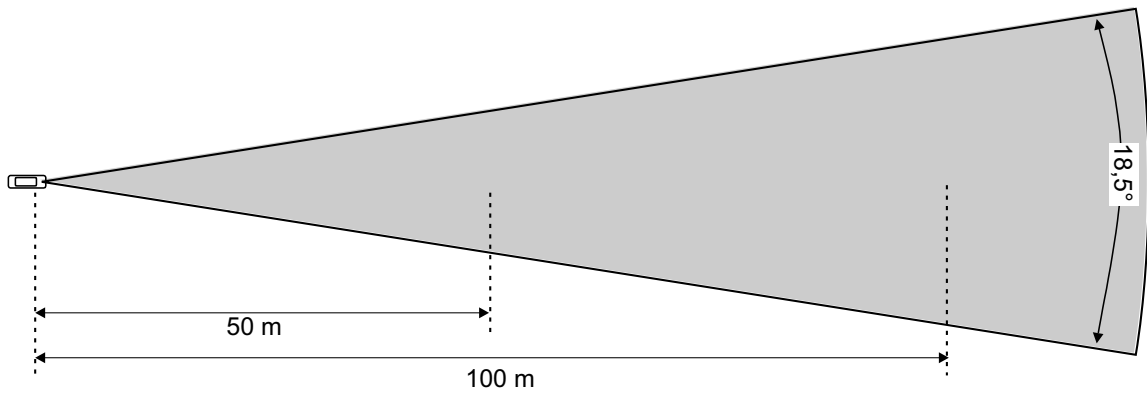


Abbildung 4.1: Erfassungsbereich der Videokamera

Die Abbildungseigenschaften der Kamera werden durch ein physikalisches Modell beschrieben, welches einen Punkt  $P_K = [x_K, y_K, z_K]^T$  aus dem dreidimensionalen Kamerakoordinatensystem in einen Punkt  $P_B = [x_B, y_B, z_B]^T$  in der zweidimensionalen Bildebene abbildet. Das Kamerakoordinatensystem hat seinen Ursprung im Projektionszentrum der Kamera, mit der x-Achse entlang der optischen Achse ausgerichtet (siehe Bild 4.2). Das Bildkoordinatensystem befindet sich mit der yz-Ebene in der Bildebene, die x-Achse ist parallel zur optischen Achse ausgerichtet. Die Abbildung besteht aus einer Zentralprojektion mit der Brennweite  $f$  gefolgt von einer Verschiebung um  $t_B = [0, t_y, t_z]^T$  und einer Skalierung mit  $s_B = [0, s_y, s_z]^T$ . Die Transformationen lassen sich als homogene Matrizen (Anhang A) darstellen:

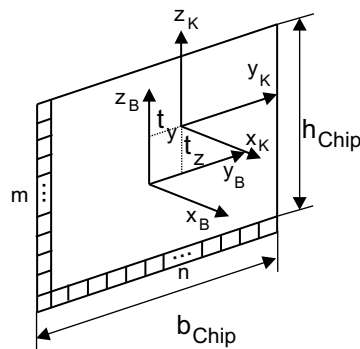


Abbildung 4.2: Kamera- und Bild-Koordinatensystem

(Zentral-)Projektion:

$$T_P(f) = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & f & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix} \quad (4.2)$$

Verschiebung:

$$T_T(t) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & t_y \\ 0 & 0 & 1 & t_z \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (4.3)$$

Skalierung:

$$T_S(s) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & s_y & 0 & 0 \\ 0 & 0 & s_z & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (4.4)$$

Da angesichts der verwendeten Brennweite von  $f = 16$  mm keine Linsenverzerrungen modelliert werden, können die Transformationen zusammengezogen werden zu:

Gesamtabbildung:

$$T_A(f, t, s) = T_S T_T T_P = \begin{bmatrix} 0 & 0 & 0 & 0 \\ t_y & f s_y & 0 & 0 \\ t_z & 0 & f s_z & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix} \quad (4.5)$$

## 4.2 Lidar

Bei dem in den Versuchen eingesetzten Lidarsensor handelt es sich um den Prototypen eines 28-kanaligen Mehrstrahllasers, der nach dem Pulslaufzeitprinzip arbeitet. Jeder der 28 mehrzielfähigen Kanäle verfügt über eine eigene Sende- und Empfangsdiode. Aus den Rückstreusignalen werden durch den im Sensor integrierten Mikrocontroller die Zielentfernungen extrahiert und mit einer Zykluszeit von 60 ms auf dem CAN(Controller Area Network)-Bus ausgegeben. Abbildung 4.3 zeigt die Strahlgeometrie des Lidarsystems. Wie man sieht, besteht das System aus 16 zentralen Kanälen mit einer horizontalen Strahlaufweitung von  $0.9^\circ$  ( $\approx 15$  mrad/Strahl) sowie jeweils 6 peripheren Kanälen zu beiden Seiten mit einer horizontalen Strahlaufweitung von  $1.7^\circ$  ( $\approx 30$  mrad/Strahl) sowie Lücken zu ebenfalls jeweils  $1.7^\circ$  zwischen den peripheren Strahlen. Der solchermaßen gestaltete Erfassungsbereich des Lidars folgt damit direkt dem auch durch das menschliche visuelle System realisierten Ansatz der Fokussierung auf das Zentrum und der größeren Auflösung in der Peripherie. (Die beim Menschen aus der Detektion in den peripheren Bereichen resultierende Blickzuwendung muß allerdings durch den festen Einbau des Sensors unterbleiben.) Wie aus Bild 4.3 zu erkennen ist, haben die Strahlen im Innenbereich eine Reichweite von 150 m, während die äußeren Strahlen in der Reichweite auf ca. 50 m begrenzt sind.

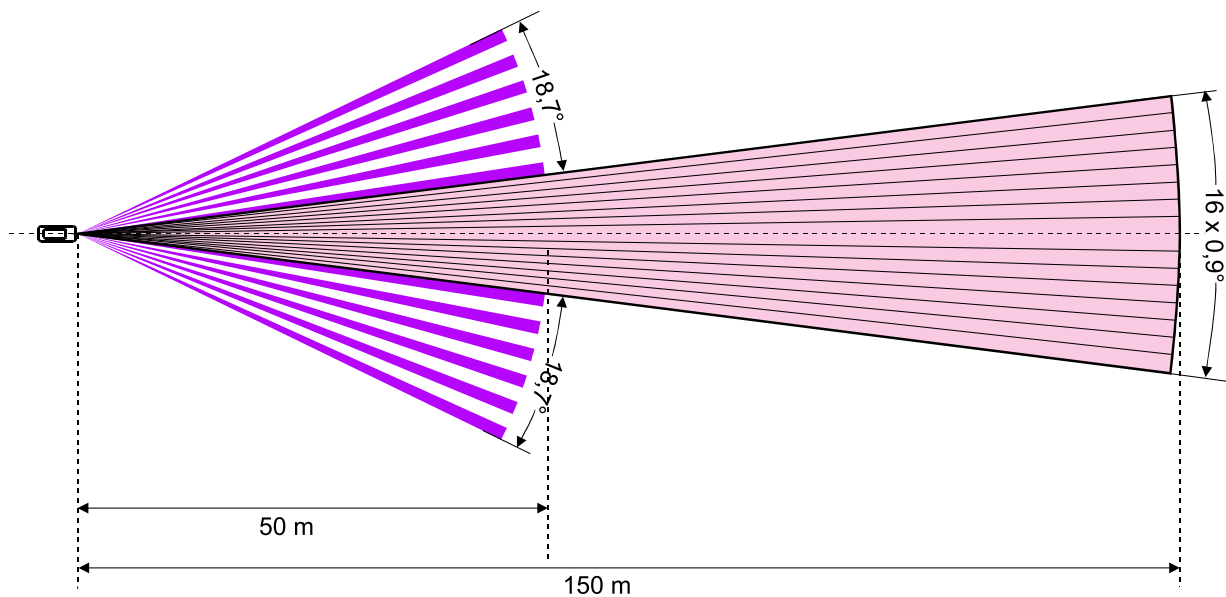


Abbildung 4.3: Strahlgeometrie des Mehrstrahl-Lidars

### 4.3 Fahrzeugsensorik

Nur der Vollständigkeit halber sei hier auch kurz die verwendete Fahrzeugsensorik genannt. Als solche gelangen in dieser Arbeit lediglich die serienmäßig verfügbaren Sensoren zum Einsatz. Im einzelnen handelt es sich um folgende Informationen:

- Raddrehzahlsensoren (ABS) zur Bestimmung von Geschwindigkeit und Drehrate
- Gierratensensor (ESP) zur Drehratenbestimmung um die Hochachse
- Lenkradwinkel

Weitere auf den CAN-Bussen im Fahrzeug verfügbaren Signale werden zwar im Rahmen von - auf der Umfelderkennung aufsetzenden - Funktionen benötigt, sind aber für die hier betrachtete Umfeldwahrnehmung selbst nicht direkt erforderlich, sondern allenfalls von mittelbarem Interesse. Ein Beispiel wäre etwa die Aufmerksamkeitssteuerung auf eine benachbarte Fahrspur bei erkannter Spurwechselabsicht anhand des Blinkersignals.

In den Versuchsfahrzeugen VaMP und VaMoRs war dagegen - u. a. zum Querfeldeinfahren mit dem VaMoRs - weitere Inertialsensorik mit zum Teil deutlich höherer Signalgüte gegenüber in der Automobilindustrie üblichen Seriensensoren verfügbar [Siedersberger 03].

# Kapitel 5

## Rekursive Schätzverfahren in der Umfelderfassung

Die Repräsentation der Objektmodelle auf der 4D-Objektebene erfolgt in der Zustandsraumbeschreibung, wie sie in der Regelungstechnik eingeführt ist, siehe z. B. [Föllinger]. Um den Zustand der Objekte zu schätzen, gelangen Verfahren zur rekursiven Zustandsschätzung nach [Kalman 60] zum Einsatz. Der Einsatz des Kalman-Filters im Rahmen der Interpretation dynamischer Bildfolgen zur rekursiven Form- und Bewegungsschätzung erfolgte erstmalig in der Arbeit von [Wünsche 87]. Auch die spezielle Implementation des Kalman-Filters in faktorisierter, numerisch stabiler Form ( $UDU^T$ -Zerlegung), wie sie in dieser Arbeit zum Einsatz gelangt, erfolgte erstmals durch [Wünsche 87].

An dieser Stelle soll keine Darstellung der Kalman-Filterung an sich gegeben werden, dazu sei auf die Originalarbeit von [Kalman 60] und auf die breite Literatur zu diesem Themenkreis verwiesen [Maybeck 79], [Loffeld 90], [Gelb 84], [Brammer, Siffling 75]. Eine kurze Darstellung zur Klärung der hier verwendeten Nomenklatur findet sich in Anhang B. Vielmehr sollen hier zwei Aspekte, welche wesentlich sind zum Verständnis der Verwendung des Kalman-Filters im Rahmen dieser Arbeit noch einmal kurz betont werden, die insbesondere zum Verständnis der Problematik bei der Assoziation von Meßwerten von essentieller Bedeutung sind. Es handelt sich dabei zum einen um das Thema des Filter-Tunings, sowie darauf aufbauend, spezieller Aspekte der Meßwertassoziation in Verbindung mit den umfelderfassenden Sensoren.

### 5.1 Kalman-Filter-Tuning

Die Kalman-Filterung von Daten geht von einem Optimalitätsprinzip aus, welches nur dann gültig ist, wenn die Kombination aus dynamischem Modell sowie den Störprozessen das System und die auf dieses einwirkenden Einflüsse auch tatsächlich hinreichend genau charakterisieren. Jenseits dieses qualitativen Aspekts muß auch die quantitative Modellierung wohlüberlegt gewählt werden, insbesondere die Systemfehlerkovarianz  $\underline{Q}$ , die Meßfehlerkovarianz  $\underline{R}$  und die initiale Schätzfehlerkovarianz  $\underline{P}_0$  zum Zeitpunkt  $\bar{k} = 0$  betreffend.

Die im realen System auftretenden Störungen müssen in jedem Fall den angenommenen Verteilungen gehorchen, da sonst

- Im Falle zu groß angenommener Störungen die Genauigkeit und damit die mögliche Dynamik des Filters unnötig verschlechtert wird (zu ausgeprägtes Tiefpaßverhalten).
- Im Falle zu klein angenommener Störungen das Filter nicht konvergiert, da die Residuen nicht der angenommenen Innovationskovarianz gehorchen. Dadurch werden eigentlich durch Störungen bedingte Residuen nicht als solche angesehen, sondern das Filter interpretiert sie als reale Änderungen des Systemzustands.

Das Vorgehen, ein Kalman-Filter auf ein bestimmtes System und einen bestimmten Anwendungsfall einzustellen, wird als Filter-Tuning bezeichnet. Eine zentrale Rolle spielt dabei die Innovationskovarianz nach Gl. B.20. Anhand dieser kann überprüft werden, ob die Meßwerte der angenommenen Residuenverteilung genügen. Um das Filter möglichst empfindlich auszulegen ohne die Stabilität/Konvergenz zu gefährden kann eine Meßwertselektion in Form einer Ausreißerdetektion zur Laufzeit durchgeführt werden. Meßwerte, die außerhalb eines gewählten Vertrauensintervalls liegen, das sich aus der  $\chi^2$ -Verteilung ergibt, werden als Ausreißer verworfen. Wählt man ein Vertrauensintervall von  $2\sigma$  ( $\sigma$  = Standardabweichung) so liegen 95 % der Meßwerte innerhalb dieses Intervalls, bei  $3\sigma$  sind es 99 %. Allerdings ergibt eine Ausreißerdetektion natürlich nur so lange einen Sinn, wie der Anteil der Ausreißer nur gering ist. Im Grenzfall werden so viele Meßwerte als Ausreißer eingeordnet, daß die in den verbleibenden Meßwerten enthaltenen Informationen nicht ausreichen, um den Schätzfehler entscheidend zu reduzieren, so daß dieser über der Zeit anwächst. Durch den anwachsenden Schätzfehler weitet sich das Vertrauensintervall aus und damit auch die Anzahl akzeptierter Meßwerte. Die Ausreißerdetektion allein garantiert aber noch nicht die Konvergenz des Filters. Wenn alle Meßwerte sich am Rande des Vertrauensintervalls drängen, werden sie zwar akzeptiert, sie genügen aber dennoch nicht der angenommenen gaußförmigen Verteilung. Auch ein solches Filter ist als "mismatched" anzusehen.

## 5.2 Meßwertassoziation bei umfelderfassenden Sensoren

Bei Einsatz des Kalman-Filters in der Umfelderfassung nach dem 4D-Ansatz ergibt sich bzgl. der Bildmessung als weiteres Problem, jenes der Zuordnung von Meßwerten zu den erwarteten Referenzwerten. Dieser allgemein als Korrespondenzproblem (hier zwischen den Modell- und den Datenmerkmalen) bezeichnete Unsicherheitsfaktor stellt eines der Hauptproblemfelder bei der Umfeldwahrnehmung dar. Bezogen auf das Kalman-Filter ist das Zuordnungsproblem insofern von Bedeutung, als es direkt in Zusammenhang mit der Konvergenz bzw. Divergenz des Filters steht. Um die Problematik zu verdeutlichen, seien die folgenden Szenarien bei der Messung gegeben:

1. Es wird *kein* Meßwertkandidat für den Referenzwert gefunden.
2. Es wird *ein* zugehöriger Meßwertkandidat für den Referenzwert gefunden.

3. Es wird *ein nicht zugehöriger* Meßwertkandidat für den Referenzwert gefunden.
4. Es werden *mehrere* Meßwertkandidaten für den Referenzwert gefunden. Der *zugehörige* Meßwertkandidat *befindet sich* darunter.
5. Es werden *mehrere* Meßwertkandidaten für den Referenzwert gefunden. Der *zugehörige* Meßwertkandidat *befindet sich nicht* darunter.

Berücksichtigt man auch noch das Vertrauensintervall der Referenzwerte, entstehen weitere Szenarien, bei denen die Meßwertkandidaten und der zugehörige Meßwert innerhalb oder außerhalb dieses Intervalls liegen können. Grundsätzlich besteht zwischen allen Szenarien kein Unterschied, denn der Umstand, sich bei Vorliegen nur eines Meßwertkandidaten nicht unter mehreren entscheiden zu müssen, bedingt keineswegs die Korrektheit dieser Entscheidung. Einfache Zuordnungsmethoden wie etwa das sog. Nächste-Nachbar-Filter (engl. : Next-Neighbour-Filter (NNF)), die stets den dem Referenzwert zunächst liegenden Meßwertkandidaten selektieren, führen i. d. R. nicht zum gewünschten Ergebnis. Als aufwendigere Verfahren bieten sich an:

- Statistische Verfahren, welche die auftretenden Residuen beobachten.
- Zusätzliche Attribute von Messungen, die insbesondere bei Bildmerkmalen oft vorhanden sind und über die Modellwissen vorliegt, können die Zuordnung erleichtern.
- Eigenschaften der Messungen in Verbindung (Gruppierung) mit anderen (benachbarten) Messungen.
- Die Unsicherheit der Meßwertzuordnung kann über eine Anpassung (Parametrisierung) des Meßrauschens individuell für jeden Meßwert in die Schätzung eingebracht werden. Meßwerte mit höherer Sicherheit in der Zuordnung werden stärker und unsichere (Fehl-)Zuordnungen nur geringer gewichtet in die Schätzung eingehen.

Diese zusätzlichen Modellannahmen können dazu beitragen, die Meßwertzuordnung zu verbessern. Sie müssen aber in Kombination mit dem Kalman-Filter angewendet werden, welches seinerseits das Zustandsraummodell des betrachteten Systems wiedergibt.

Wenn also die erwarteten Referenzwerte des Kalman-Filters über einen langen Zeitraum stets weit neben den letztlich zugeordneten Meßwertkandidaten liegen, kann das Gesamtmodell nicht stimmig sein.





# Kapitel 6

## Umfeldmodellierung

In diesem Kapitel stehen zunächst die Modelle für die Objektbewegung in Abschnitt 6.1 sowie die Objektform in Abschnitt 6.2 im Fokus. Anschliessend wird dann das aus solcherart modellierten Objekten aufgebaute Umfeldmodell in Form eines Szenenbaums dargestellt (Abschnitt 6.3). Dabei geht es zum einen (Abschnitt 6.3.1) um die eigentliche Szenenrepräsentation, zum anderen dann um die auf dieser Struktur des Szenenbaums aufsetzende Zustandsschätzung aus Umfelddaten (6.3.2).

### 6.1 Bewegungsmodelle

Die Bewegungsmodelle dienen zur Beschreibung der dynamischen Aspekte von Objekten. Es wird unterschieden zwischen Modellen für die Beschreibung der Bewegung von Objekten sowie dem Dynamikmodell für die Eigenbewegung. Die Auswahl der Modelle hängt zum einen von dem beabsichtigten Einsatz des Systems in einem bestimmten Szenario ab, als auch von den grundsätzlichen Möglichkeiten, die Modellgrößen bestimmen zu können. Letzterer Punkt gilt insbesondere für (Fremd-)Objekte, die ja nur anhand der Umfeldsensorik erfaßt werden können, was aufgrund der begrenzten Beobachtbarkeit sowie der Meßgenauigkeit der Sensoren zu Einschränkungen führt.

#### 6.1.1 Objekte

Den Bewegungsmodellen für Objekte liegt die Modellierung als bewegter Massepunkt zugrunde. Längs- und Querdynamik werden dabei als getrennt voneinander modelliert und geschätzt. Es wird unterschieden zwischen einem Modell mit der Annahme konstanter Beschleunigung, einem Modell mit der Annahme einer konstanten Geschwindigkeit sowie einem stationären Modell. Durch die entsprechende Kombination eines Modells für die Längsführung sowie eines Modells für die Querführung kann man den unterschiedlichen Objektklassen sowie den Anforderungen unterschiedlicher Situationen gerecht werden. Denkbar ist auch eine Manöverdetektion in Kombination mit einer Modellumschaltung für ein

und dasselbe Objekt. Darüberhinausgehende Techniken wie beispielsweise die Methode der “Interacting Multiple Models” (IMM) werden hier nicht betrachtet

### Modell “Konstante Beschleunigung“

Beim Modell der “Konstanten Beschleunigung“ ist der Zustandsvektor mit  $x = (p, v, a)^T$  gegeben, worin  $p$  die Position,  $v$  die Geschwindigkeit und  $a$  die Beschleunigung darstellen. Die zugehörige Systemmatrix ist gegeben durch

$$\Phi_k = \begin{pmatrix} 1 & T & T^2/2 \\ 0 & 1 & T \\ 0 & 0 & 1 \end{pmatrix} \quad (6.1)$$

und die Systemstörung  $w_k$  sowie die korrespondierende Kovarianzmatrix  $Q_k$  durch

$$w_k = \begin{pmatrix} T^2/2 \\ T \\ 1 \end{pmatrix} \sigma_{a,k} \quad (6.2)$$

$$Q_k = w_k w_k^T = \begin{pmatrix} T^4/4 & T^3/3 & T^2/2 \\ T^3/2 & T^2 & T \\ T^2/2 & T & 1 \end{pmatrix} \sigma_{a,k}^2 \quad (6.3)$$

Nach [Loffeld 90] gibt  $\sigma_{a,k}$  die maximal zulässige Objektbeschleunigung an. Über Möglichkeiten, die Systemstörungen nicht nur durch einen weißen Rauschprozeß zu beschreiben, sondern auch farbiges Rauschen zu modellieren, geben [Schmid 94] und die dort enthaltenen Referenzen Auskunft.

### Modell “Konstante Geschwindigkeit“

Dieses Modell ergibt sich direkt aus dem vorangegangenen Modell der “Konstanten Beschleunigung“ unter Vernachlässigung der Beschleunigungskomponente. Der Zustandsvektor ist mit  $x = (p, v)^T$  gegeben, die zugehörige Systemmatrix durch

$$\Phi_k = \begin{pmatrix} 1 & T \\ 0 & 1 \end{pmatrix} \quad (6.4)$$

und die Systemstörung  $w_k$  sowie Kovarianzmatrix  $Q_k$  durch

$$w_k = \begin{pmatrix} T \\ 1 \end{pmatrix} \sigma_{v,k} \quad (6.5)$$

$$Q_k = w_k w_k^T = \begin{pmatrix} T^2 & T \\ T & 1 \end{pmatrix} \sigma_{v,k}^2 \quad (6.6)$$

$\sigma_{v,k}$  gibt in diesem Fall das Geschwindigkeitsrauschen an.

### Modell “Konstante Position“

Dieses Modell ist im eigentlichen Sinn kein dynamisches Modell, da es ein stationäres Objekt beschreibt, d. h. lediglich den Positionsparameter schätzt. Der Zustandsvektor ist nur mit  $x = (p)$  gegeben, die Systemmatrix entartet zu

$$\Phi_k = \begin{pmatrix} 1 \end{pmatrix} \quad (6.7)$$

und die Systemstörung  $w_k$  und Kovarianzmatrix  $Q_k$  durch

$$w_k = \begin{pmatrix} 1 \end{pmatrix} \sigma_{p,k} \quad (6.8)$$

$$Q_k = w_k w_k^T = \begin{pmatrix} 1 \end{pmatrix} \sigma_{p,k}^2 \quad (6.9)$$

mit  $\sigma_{p,k}$  als dem Positionsrauschen.

## 6.1.2 Eigenfahrzeug

Das Bewegungsmodell des Eigenfahrzeugs ist bezogen auf die Fahrspur zu beschreiben, wie sie in Abschnitt 6.2.5 dargestellt wird. Es werden Längs- und Querdynamik sowie die Nickbewegung getrennt voneinander geschätzt.

Für die Querdynamik wird gegenüber den Bewegungsmodellen wie sie von [Behringer 96] und [Mysliwetz 90] zur autonomen Fahrzeugführung eingesetzt wurden, einige vereinfachende Annahmen getroffen. Insbesondere wird der Schwimmwinkel  $\beta$  vernachlässigt, d. h. es wird angenommen, daß der Gierwinkel stets gleich dem Bahnwinkel ist, was nach [Schmid 94] auf Autobahnen auch näherungsweise angenommen werden kann. Desweiteren wird als Eingangsgröße nicht auf den Lenkwinkel  $\lambda$  sondern auf die Gierrate  $\dot{\Psi}$  zurückgegriffen. Dadurch kann auf die Kenntnis der Fahrwerksgeometrie verzichtet werden. Als diskretes Systemmodell für Ablage  $y$  und Gierwinkel  $\Psi$  ergibt sich mit der Eigengeschwindigkeit  $v$  und der mittleren horizontalen Fahrspurkrümmung  $c_{0hm}$

$$\begin{bmatrix} y \\ \Psi \end{bmatrix}_{k+1} = \begin{bmatrix} 1 & vT \\ 0 & 1 \end{bmatrix} \begin{bmatrix} y \\ \Psi \end{bmatrix}_k + \begin{bmatrix} -vT & vT^2 \\ -v^2T^2/2 & T \end{bmatrix} \begin{bmatrix} c_{0hm} \\ \dot{\Psi} \end{bmatrix} \quad (6.10)$$

Die Nickbewegung spielt insbesondere für die Bildauswertung eine entscheidende Rolle, da sie in wesentlichen Teilen die Entfernungsschätzung von Objekten aus Videobildern konditioniert. Aber auch für andere umfelderfassende Sensoren wie das Mehrstrahl-Lidar ist die Kenntnis darüber, welche “Ebene“ aufgrund der aktuellen Nickbewegung des Fahrzeugs erfasst wird, respektive wann man “in den Boden“ bzw. “in den Himmel“ schaut von entscheidender Bedeutung. Für die Nickbewegung wird hier das von [Behringer 96] eingeführte (diskrete) Modell eines gedämpften Schwingers verwendet

$$\begin{bmatrix} \Theta \\ \dot{\Theta} \\ \Theta_{\text{Offset}} \end{bmatrix}_{k+1} = \begin{bmatrix} \Phi_{11} & \Phi_{12} & 1 - \Phi_{11} \\ -\omega_0^2 \Phi_{12} & e_\sigma(c_\omega - \frac{\sigma}{\omega} s_\omega) & \omega_0^2 \Phi_{12} \\ 0 & 0 & 1 \end{bmatrix} \quad (6.11)$$

mit

$$\begin{aligned}
\Phi_{11} &= e_\sigma(c_\omega + \frac{\sigma}{\omega}s_\omega) & \Phi_{12} &= \frac{1}{\omega}e_\sigma s_\omega \\
s_\omega &= \sin(\omega T) & c_\omega &= \cos(\omega T) \\
\omega &= \sqrt{\omega_0^2 - \sigma^2} & e_\sigma &= \exp(-\sigma T)
\end{aligned}$$

## 6.2 Formmodelle

Die Repräsentation der Objektform hat zwei Aspekte zu berücksichtigen:

1. Sie muß die Form von Objekten soweit beschreiben, wie dies zur Erfüllung der Systemaufgabe erforderlich ist. Im Falle des autonomen Fahrens zählen hierzu insbesondere die für die Navigation erforderlichen Objekte. Dabei kann es sich einerseits um Sollgrößen-Objekte, d. h. die direkte Vorgabe einer Sollbahn in Form des Fahrspurverlaufs, des befahrenen Oberflächenbereichs oder auch einer Landmarke handeln, oder aber es handelt sich um "Stör"-Objekte in Form anderer Fahrzeuge oder allgemein Hindernisse im Sinne der Kollisionsvermeidung.
2. Sie muß eine Beschreibung der Objekte durch Merkmale liefern, deren Attribute durch die zur Verfügung stehenden Sensoren vermeßbar sind.

Zur Erfüllung der ersten Forderung gehört eine geometrische Beschreibung der Form, welche die zur lokalen Navigation benötigten Zustandsgrößen liefert. Die zweite Forderung nach der Beschreibung der vermeßbaren Merkmale beinhaltet darüberhinaus noch die Beschreibung von Oberflächeneigenschaften für die umfelderfassende Sensorik. Bzgl. der visuellen Wahrnehmung sind dies beispielsweise Farbe, Textur und Reflektivität, bzgl. des Lidars ist es im wesentlichen die Reflektion. Wenn im weiteren Verlauf von diesen Oberflächen-Attributen die Rede ist, soll allgemein von photometrischen Merkmalen die Rede sein, unabhängig davon, daß sich diese Bezeichnung im engeren Sinne nur auf die Beschreibung von Eigenschaften im visuellen Sichtbereich des Menschen bezieht.

Aus der Literatur ist eine ganze Fülle unterschiedlicher Methoden zur Beschreibung der Form fester, dreidimensionaler Körper bekannt. Eine Schilderung dieser Möglichkeiten und eine Erörterung der Vor- und Nachteile der verschiedenen Repräsentationsformen enthält z. B. [Ballard, Brown 82]. In dieser Arbeit wird im Hinblick auf die verwendeten Sensoren und Meßverfahren eine oberflächen- und begrenzungslinienbasierte Repräsentation gewählt.

Die Form eines Objektes wird in einem kartesischen Koordinatensystem  $O = \{x, y, z\}$  beschrieben, welches zweckmäßigerweise oft im Objektschwerpunkt gewählt wird. Die Form wird durch eine äußere Gestalt  $G_a$  und eine innere Gestalt  $G_i$  beschrieben:

$$G = \{G_a, G_i\} \quad (6.12)$$

Die Form besteht aus den Grundelementen Flächen  $F$ , Kanten  $K$  und Vertexpunkten  $V$ . Sie ist hierarchisch aus diesen Formelementen aufgebaut (s. Abbildung 6.1).

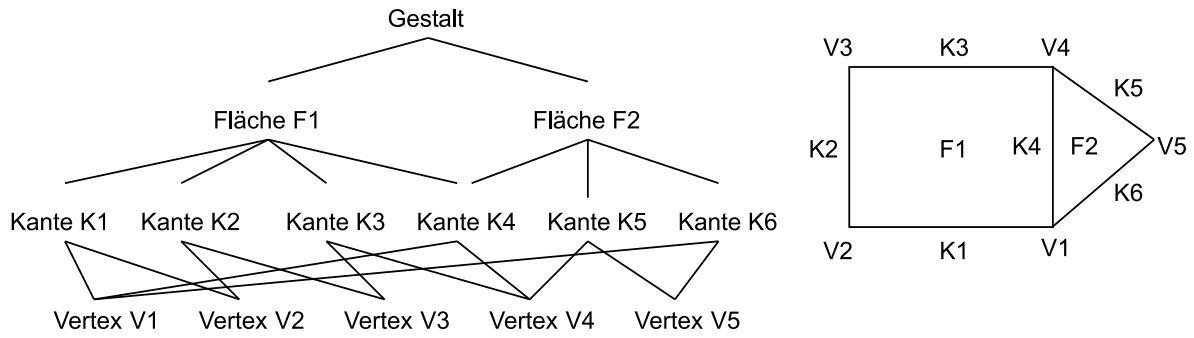


Abbildung 6.1: Hierarchische Formkonstruktion aus Formelementen an einem Beispiel

Die äußere Gestalt beinhaltet alle Merkmale, welche die äußere Hülle des Objektes bilden:

$$M_a = \{F_a, K_a, V_a\} \quad (6.13)$$

Die innere Gestalt  $G_i$  umfaßt entsprechend die nicht zur äußeren Hülle zählenden Merkmale

$$M_i = \{F_i, K_i, V_i\} \quad (6.14)$$

Die Unterscheidung zwischen den beiden Merkmalskategorien der äußeren und inneren Gestalt ist einerseits für die Sichtbarkeitsüberprüfung von Belang, andererseits ist sie eine der Möglichkeiten, eine Hierarchie in Form eines Detaillierungsgrades (engl. LOD = “Level of Detail“) zu realisieren.

### 6.2.1 Formelemente

Die Basisbausteine der Formrepräsentation bilden die sog. Formelemente. Bei diesen wird zwischen linienhaften, flächenhaften und volumenhaften Beschreibungen unterschieden, wobei letztere in dieser Arbeit nicht explizit berücksichtigt werden.

#### Linienhafte Formelemente

**Geradenabschnitte** werden durch ihren Anfangspunkt  $p_a$  und ihren Endpunkt  $p_e$  beschrieben. Ein Punkt  $p(t)$  auf dem Geradenabschnitt wird folglich durch

$$p(t) = p_a + (p_e - p_a)t \quad (6.15)$$

bestimmt.

**Fergusonkurvenabschnitte** werden durch Anfangspunkt  $p_a$ , Endpunkt  $p_e$ , sowie die beiden Tangentialvektoren  $\underline{v}_a$  im Anfangs- und  $\underline{v}_e$  im Endpunkt repräsentiert. Ein Punkt  $p(t)$  wird in Abhängigkeit der Lauflänge  $t$  auf der Fergusonkurve bestimmt durch

$$p(t) = \begin{pmatrix} t^3 & t^2 & t & 1 \end{pmatrix}^T M \quad (6.16)$$

mit

$$M = \begin{pmatrix} 2 & -2 & 1 & 1 \\ -3 & 3 & -2 & -1 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix} \quad (6.17)$$

Fergusonkurven dienen in [D. Dickmanns 97] als alleinige Bechreibungsform von Kanten. Sie werden durch Anfangs- und Endpunkt sowie durch je zu diesen Punkten gehörigen Parametervektor beschrieben.

Durch Fergusonkurven lassen sich eine Vielzahl von Kurvenformen wie etwa Geraden oder Ellipsen als “Sonderformen“ darstellen. Da die Parametrisierung sich einerseits (insbesondere in der Entwurfsphase von Formmodellen) ohne entsprechende (Hilfs-)Werkzeuge aber nur schwer handhaben läßt, andererseits eine explizite Repräsentation dieser “Sonderformen“ durch effiziente Verfahren, die die speziellen Eigenschaften dieser ausnutzen, gerechtfertigt wird, wird von dieser Möglichkeit kein Gebrauch gemacht, sondern die Anwendung von Fergusonkurven wirklich nur zur Repräsentation (schwach) gekrümmter Objekte herangezogen.

## Flächenhafte Formelemente

Flächen werden in der Formrepräsentation aus den linienhaften Formelementen als “Begrenzungs“-Kanten zusammengesetzt. Da bei mehreren zusammenhängenden Flächen (z. B. bei geschlossenen Körpern) die Trennlinien zwischen diesen als zu beiden gehörig angesehen werden, geschieht die Referenzierung der Kanten nicht direkt, sondern indirekt über sogenannte Randelemente. Die Randelemente sind bezogen auf die Fläche, der sie angehören, in der Draufsicht gesehen mathematisch positiv gerichtet. Die Randelemente verweisen auf die nur einmal definierten Kanten und geben den Umlaufsinn der Kante bzgl. der Fläche, der die Kante über das Randelement angehört, an. Eine Kante, die zwei aneinandergrenzende Flächen berandet, wird dementsprechend von zwei Randelementen mit gegenläufigem Umlaufsinn referenziert.

Neben den Randelementen, welche die Fläche definieren, können Flächen zusätzlich über weitere ihnen zugehörige Formelemente beschrieben werden. Dabei kann es sich sowohl um linienartige Strukturen, z. B. als Skelettlinien, wie auch um flächenartige Strukturen im Sinne einer Hierarchie (LOD) flächenhafter Formelemente handeln.

### 6.2.2 Topologische Beschreibung der Gesamtform

Die topologische Beschreibung der Gesamtform beinhaltet die Beschreibung der Vertexpunkte als Anfangs- bzw. Endpunkte linienhafter Formelemente. Diese referenzieren auf ihren jeweiligen Anfangs- bzw. Endvertex. Die linienhaften Formelemente werden entweder als Randelement einer oder zweier Flächen oder als inneres Element einer Fläche zugewiesen. Die Flächen wiederum werden über die Folge der sie begrenzenden Randelemente sowie ihre inneren Elemente definiert. Die Relationen sind in Form von Inzidenztabelle über Indizierungen der Formelemente repräsentiert.

### 6.2.3 Geometrische Beschreibung der Gesamtform

Die geometrische Formbeschreibung liefert die konkrete Ausprägung einer Form. Als geometrische Formparameter ergeben sich zunächst alle bei der Beschreibung der Formelemente angegebenen Parameter. Durch die Verkopplung der Formelemente durch die topologische Beschreibung reduziert sich die Anzahl wählbarer Parameter. Eine weitere Reduzierung erfolgt durch spezifische Nebenannahmen über die Gestalt (z. B. Symmetrien). Je mehr Nebenannahmen getroffen werden, desto mehr Modellwissen über die Gestalt zu repräsentierender Objekte fließt ein, und desto geringer wird i. a. die Anzahl frei zu wählender Formparameter (s. Beispiel Abb. 6.2). Wenn die Formparameter definiert sind, so kann durch Festlegung von Parameterintervallen für diese Parameter eine Einteilung dieser generischen Repräsentation in Formklassen erfolgen. Diesen kann dann in Abhängigkeit der Situation und des Abstraktionsgrades der Repräsentation eine Bedeutung zugewiesen werden, die für die Erkennung der Objektklasse genutzt werden kann. Für die Erkennung ist es ein qualitativer Unterschied, wie viele Parameter zur Repräsentation herangezogen werden müssen.

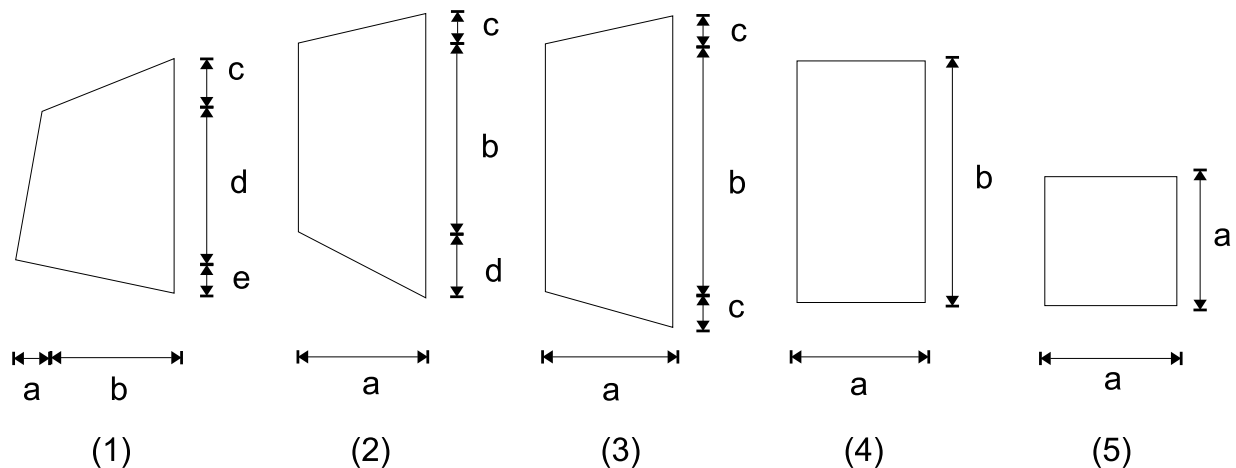


Abbildung 6.2: Festlegung von Zwangsbedingungen durch Formparametrisierung

### 6.2.4 Photometrische Beschreibung der Formelemente

Das Erscheinungs-“Bild“ von Objektmerkmalen bezogen auf einen Sensor hängt nicht allein von der Form eines Objektes sondern maßgeblich von anderen Objekteigenschaften, insbesondere von Oberflächeneigenschaften (Farbe/Reflektivität) oder allgemein Materialeigenschaften ab. Bei der passiven Videosensorik kommen die Beleuchtungsquellen als aktive Elemente hinzu. Jedem Bildmeßverfahren liegt zumindest implizit ein Beleuchtungsmodell zugrunde, häufig in Form einer diffusen Lichtquelle. Die Gründe liegen oft in der nicht gerechtfertigten bzw. nicht vorhandenen Rechenleistung zur Berücksichtigung eines expliziten Beleuchtungsmodells, oft allerdings auch schlicht in der Unkenntnis der Umgebungsbedingungen und damit der Lichtquellen.

Für den Anwendungsfall der autonomen Fahrzeugführung gibt es bei Tage i. a. nur eine



Beleuchtungsquelle, die Sonne. Je nach Bewölkungsgrad und sonstigen Umweltbedingungen (Regen) ändert sich jedoch das geltende Beleuchtungsmodell, sowie möglicherweise die Reflektionseigenschaften von Objekten (nasse Fahrbahn). Nachts gibt es zwar sehr viele verschiedene Beleuchtungsquellen etwa durch andere Fahrzeuge, Straßenbeleuchtung etc., aber es ist natürlich vollkommen sinnlos, diese explizit modellieren zu wollen. In künstlichen Umgebungen (Fabrikhalle) können demgegenüber mehrere Lichtquellen auftreten, die sich aufgrund der als konstant anzunehmenden Umgebung auch besser modellieren lassen. Darüberhinaus bieten künstliche Umgebungen manchmal den Vorteil, eine Optimierung der Beleuchtung entsprechend den Erfordernissen der Bildverarbeitung zu erlauben.

In dieser Arbeit wird vereinfachend nur ein diffuses Beleuchtungsmodell implizit angenommen. Die Objektoberflächen können als transparent oder nicht transparent beschrieben werden. Aufgrund der grundsätzlich kantenbasierten Repräsentation, bei der die Flächen als näherungsweise eben angenommen werden, machen weitergehende Annahmen und Beleuchtungsmodelle auch keinen Sinn. Denn erst, wenn die Flächenrepräsentation sich der realen Form stärker annähert und das Erscheinungsbild der Objekte mit komplexeren Methoden der Computergrafik etwa in Form von Flächenstücken beschrieben werden kann, erhalten die Ergebnisse von Beleuchtungsmodellen eine Aussagekraft.

Neben der Eigenschaft der Transparenz/Nicht-Transparenz erhalten die Flächen als weitere Eigenschaft ein Attribut zur Kennzeichnung ihrer Homogenität. Dieses Flächen-Attribut wird an die die Fläche bildenden Kanten weitergegeben, so daß diese Eigenschaft bei der Bildmessung und Primitivengenerierung ausgenutzt werden kann. In gewissen Grenzen lassen sich mithilfe des Homogenitäts-Attributs auch andere Effekte erreichen: Bei stark gekrümmten, stark reflektierenden (metallischen) Oberflächen, kann die Fläche zwar an sich einen homogenen Farbwert haben, wird aufgrund der Krümmung und Reflektion aber nie als homogen erscheinen. (Daß wir als Menschen solche Flächen als homogen wahrnehmen, hängt mit unserem Wissen über die Materialeigenschaften, die Beleuchtung und die Objekte zusammen.)

### 6.2.5 Formmodell-Datenbank

Die Formen der für die Fahrumgebungserfassung relevanten Objektklassen sind in einer Formmodell-Datenbank zusammengefaßt. Im wesentlichen wird unterschieden zwischen dem Fahrspurmodell und den Formmodellen der Objekte. Objekte werden hierbei durch ein einfaches Quadermodell zur Repräsentation allgemeiner kompakter Objekte und/oder ein Polyedermodell als Verfeinerung des Quadermodells für Fahrzeuge dargestellt.

#### Fahrbahnmodell

Eine Sonderstellung unter den Formmodellen nimmt die Modellierung der Fahrbahn oder genauer gesagt eines Fahrstreifens ein. Diese wird - den Grundsätzen der deutschen Straßenbauvorschriften für den Bau von Straßen höherer Ordnung [RAL 84] entsprechend - als Klothoidenband modelliert, wie dies in Abbildung 6.2.5 dargestellt ist. Die Modellierung erfolgt allerdings als sog. gleitende Modellierung, wie sie erstmals von Dickmanns

und Mysliwetz [Mysliwetz 90] vorgeschlagen wurde. D. h. das in Abb. 6.2.5 gezeigte Klothoidenband ist nicht ortsfest, sondern wird als “gleitender“ Fahrspurabschnitt vor dem Eigenfahrzeug hergeschoben. Dadurch ergibt sich wie in [Mysliwetz 90] eine Verkopplung mit dem Zustand des Eigenfahrzeugs.

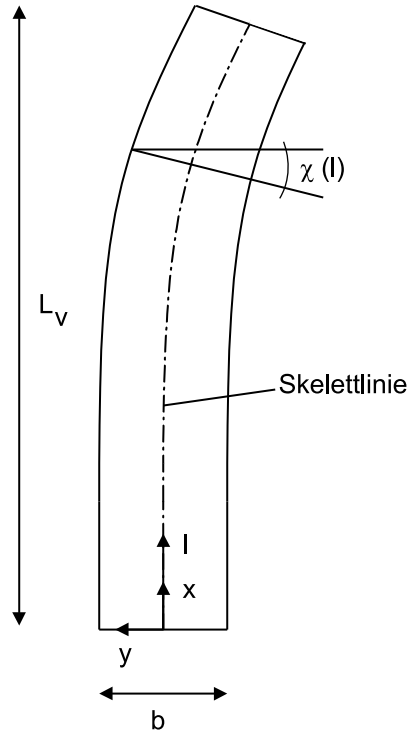


Abbildung 6.3: Formmodell Fahrspur

Die hier angenommene Modellierung folgt den Darstellungen in [Mysliwetz 90] und [Behringer 96], nimmt gegenüber diesen jedoch einige Vereinfachungen vor, die im Rahmen dieser Arbeit nicht ins Gewicht fallen. So wurde auf die Modellierung einer vertikalen Fahrspurkrümmung, die in [Mysliwetz 90] erstmals eingeführt wurde, verzichtet, ebenso auf die Modellierung einer Breitenänderung im Vorausschaubereich. Damit ergibt sich als diskretes Systemmodell der horizontalen Krümmung des Fahrbahnmodells

$$\begin{bmatrix} c_{0hm} \\ c_{1hm} \\ c_{1h} \end{bmatrix}_{k+1} = \begin{bmatrix} 1 & L_v(1-\rho)/3 & L_v(1-\rho)/3 + vT \\ 0 & \rho & 1-\rho \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} c_{0hm} \\ c_{1hm} \\ c_{1h} \end{bmatrix}_k \quad (6.18)$$

mit  $\rho = e^{-3vT/L_v}$  und

$c_{0hm}$	gemittelte horizontale Krümmung
$c_{1hm}$	gemittelte horizontale Krümmungsänderung
$c_{1h}$	horizontale Krümmungsänderung
$L_v$	Vorausschauweite
$v$	Eigengeschwindigkeit

Das diskrete (Trivial-)Systemmodell für die Fahrspurbreite lautet

$$[b]_{k+1} = [1] [b]_k \quad (6.19)$$

mit  $b$  als der Fahrspurbreite.

Die Punkte auf der Skelettlinie ergeben sich für kleine Krümmungen näherungsweise zu

$$P_s = \begin{bmatrix} x_s \\ y_s \\ z_s \end{bmatrix} \approx \begin{bmatrix} l \\ \frac{c_{ohm}}{2} l^2 + \frac{c_{lhm}}{6} l^3 \\ 0 \end{bmatrix} \quad (6.20)$$

Die Punkte auf den visuell vermeßbaren Fahrspurbegrenzungen werden darauf aufbauend wie folgt ermittelt

$$P_{r/l} = \begin{bmatrix} x_{r/l} \\ y_{r/l} \\ z_{r/l} \end{bmatrix} = P_s(l) \pm \frac{b(l)}{2} \begin{bmatrix} -\sin(\chi(l)) \\ +\cos(\chi(l)) \\ 0 \end{bmatrix} \quad (6.21)$$

### Allgemeines Quadermodell

Zur Beschreibung allgemeiner kompakter Objekte, insbesondere auch zur Beschreibung von Fahrzeugen dient in dieser Arbeit ein Quader-Formmodell, mit der in Abb. 6.4 gezeigten Gestalt und Benennungskonvention.

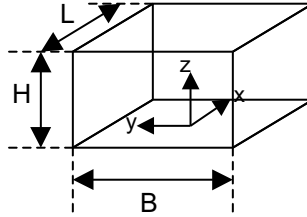


Abbildung 6.4: Formmodell Quader

Die hierarchische Zerlegung in Flächen, Kanten und Vertexknoten des Quadermodells ist in Abb. 6.5 exemplarisch dargestellt. Den einzelnen Flächen wie auch jeder einzelnen Kante können je nach zu erfassender Objektklasse photometrische Attribute mitgegeben werden. Der Ursprung des Objektkoordinatensystems liegt im Flächenschwerpunkt der Bodenfläche (Fläche 5 (Abb. 6.5)). Als Formparameter werden Länge  $L$ , Breite  $B$  und Höhe  $H$  gewählt:

$$x_f = (L, B, H) \quad (6.22)$$

### Fahrzeug-Polyedermodell

Um das Tracking von Objekten sicherer zu machen und/oder eine genauere Klassifizierung von Fahrzeugen durchzuführen, ist ein Polyedermodell wie es in [Schick 92] und [Koller 92]

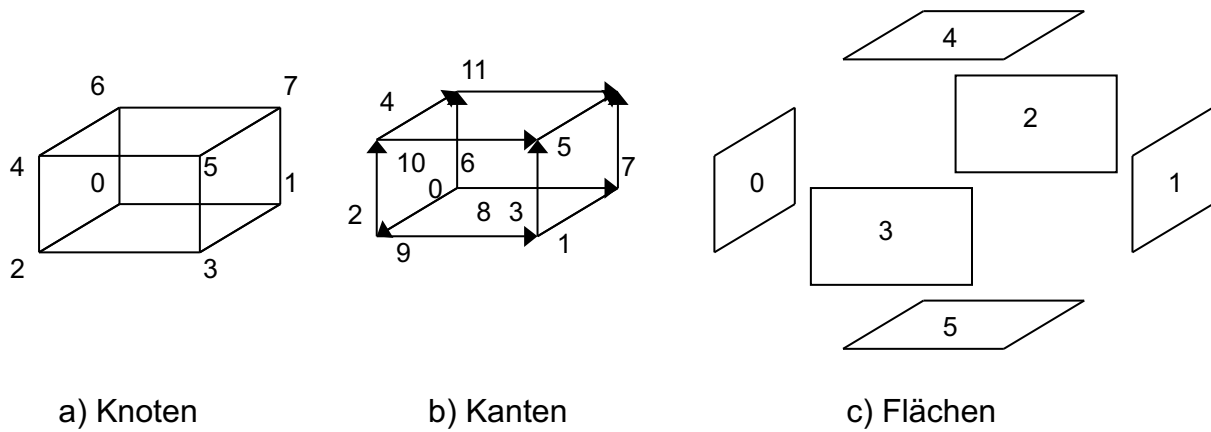


Abbildung 6.5: Quaderkonstruktion aus Formelementen

Verwendung findet, modelliert. Dieses Modell kann durch eine Variation der Formparameter und eine Festlegung von Wertebereichen für die Formparameter eine Vielzahl von Fahrzeugformen darstellen (Abb. 6.6). Zur Beschreibung dienen die in Abb. 6.7 angegebenen Formparameter:

$$x_f = (L, L_M, L_{FF}, L_D, L_{HF}, L_K, H, H_F, H_M, H_{FF}, H_H, H_K, H_{HF}, B, B_D, B_S, R_R) \quad (6.23)$$

Wie man aus Abb. 6.7 erkennt, ergeben sich geschlossene Bemaßungsketten, die bei einer Formschatzung zu einer Überbestimmung der Formparameter führen, daher wird der Formparametervektor reduziert, so daß lediglich unabhängige Parameter enthalten sind

$$x_f = (L, L_M, L_{FF}, L_D, L_K, H, H_F, H_M, H_H, H_K, B, B_D, R_R) \quad (6.24)$$

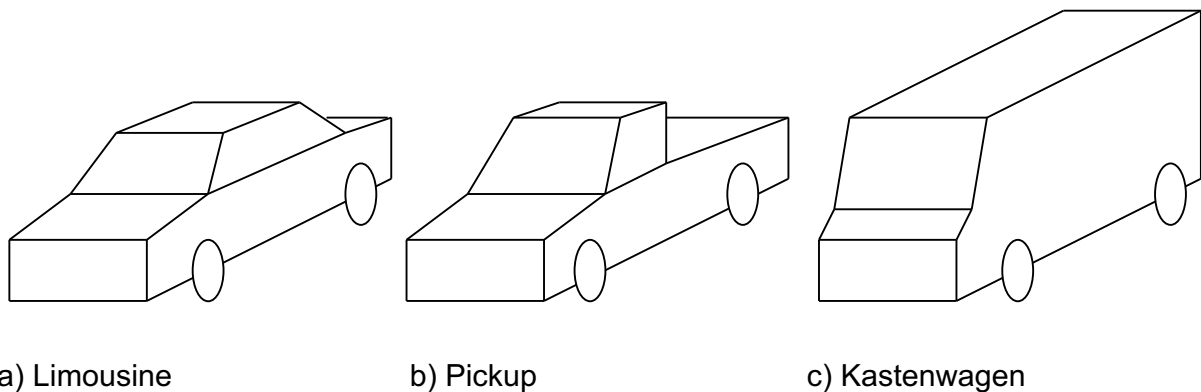


Abbildung 6.6: Beispielausprägungen des Polyedermodells

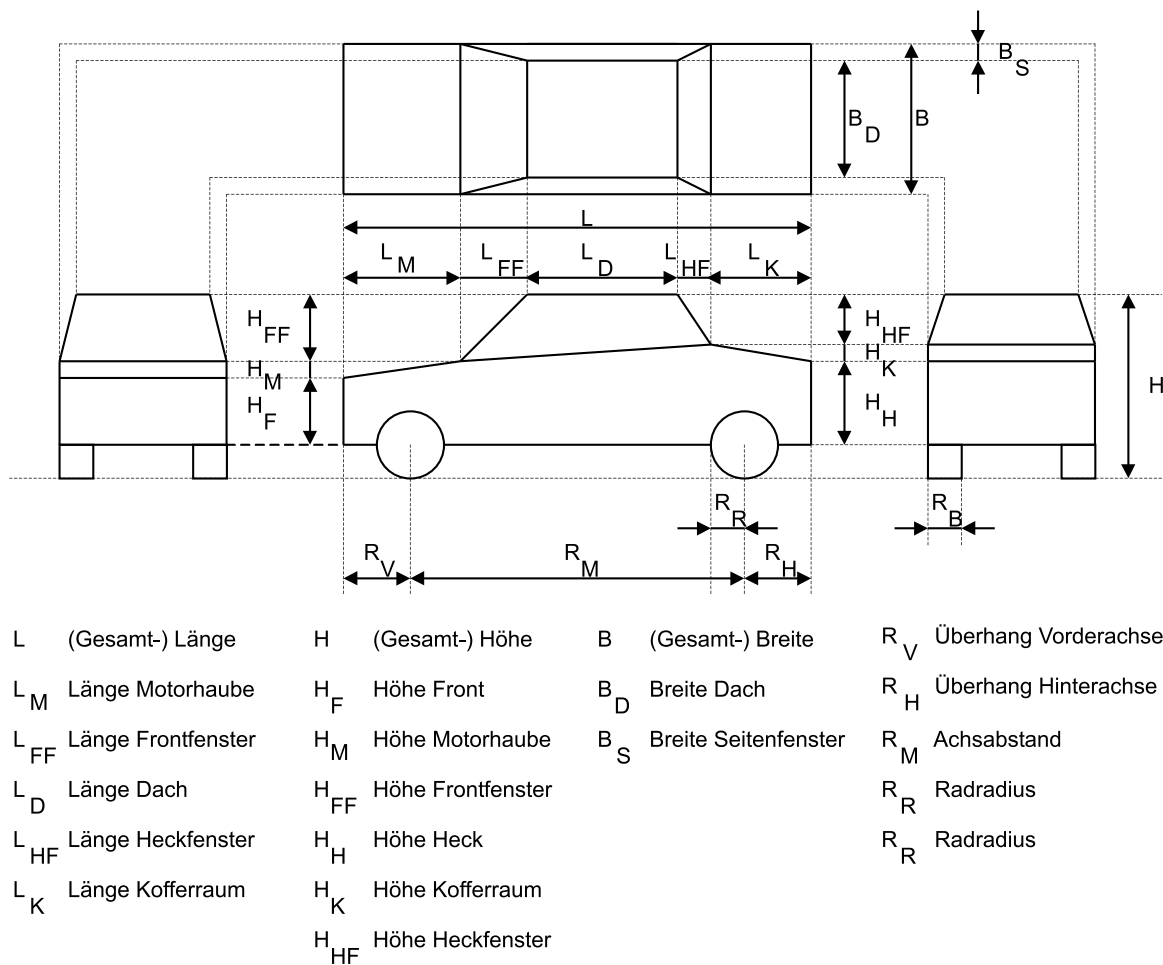


Abbildung 6.7: Polyeder Schnittdarstellung

## 6.3 Umfeldmodell und Szenenbaum

Die Gesamtheit des zu einem gegebenen Zeitpunkt vorliegenden Wissens über das Umfeld des Eigenfahrzeugs wird als Umfeldmodell oder Szene bezeichnet. Es besteht aus sämtlichen Objekten - Fahrspuren, Fahrzeugen und sonstigen Objekten - die durch die Umfeldsensorik erfasst und als Objekte instanziiert worden sind. Es folgt zunächst die Beschreibung der eigentlichen Repräsentation der Szene, gefolgt von der Darstellung der auf der Repräsentation aufsetzenden Zustandsschätzung der Objekte in der Szene.

### 6.3.1 Szenenrepräsentation

Als Organisationsform für das Umfeldmodell wird ein sog. Szenenbaum verwendet, wie er aus der Computergraphik bekannt ist und wie er erstmals in [D. Dickmanns 97] im Rahmen der visuellen Szenenerkennung zum Einsatz kam. Im Szenenbaum sind, ausgehend von einem Wurzelknoten, alle in der Szene, dem Umfeld, erfassten Objekte hierarchisch unter Angabe der jeweiligen Relativlage zum übergeordneten Knoten enthalten. Neben



Die Beziehungen zwischen den Knoten basieren auf der Angabe von homogenen Koordinaten bzw. von homogenen Transformationsmatrizen. Da nicht jede Beziehung zwischen Objekten direkt in Form dieser 6 Freiheitsgrade ( $x, y, z, \Phi, \Theta, \Psi$ ) bzw. als Polynomialmodell in diesen Freiheitsgraden gegeben werden kann, erfolgt im allgemeinen Fall eine Abbildung des Zustands zu einem gegebenen Zeitpunkt  $t$  auf die 6 Freiheitsgrade und damit auf die homogene Transformationsmatrix für diesen Zeitpunkt  $t$ :

$$\underline{x}(t) \rightarrow \underline{x}_{6\text{DOF}}(t) \quad (6.25)$$

Die Transformation zwischen beliebigen Knoten des Szenenbaums erhält man durch einfache Konkatenation der Transformationen in Form einer Matrixmultiplikation unter Beachtung der Transformationsrichtung und dementsprechender Verwendung der Vor- bzw. Rück-(inversen) Transformation. Im o. g. Szenenbaum ergibt sich die Transformation für Punkte aus dem Objektkoordinatensystem  $K_{\text{Objekt } 1}$  in das Koordinatensystem der Kamerabildebene  $K_{\text{Kamerabildebene}}$  etwa zu:

$$P_{\text{KB}} = T_{\text{KB,KA}} T_{\text{KA,S}} T_{\text{S,EGO}} T_{\text{EGO,Welt}} T_{\text{O}_1,\text{Welt}}^{-1} P_{\text{O}_1} \quad (6.26)$$

### 6.3.2 Zustandsschätzung auf Basis des Szenenbaums

Die Schätzung der Bewegung und der Form von Umfeldobjekten erfolgt auf der Basis des in Kapitel 6.3 eingeführten Szenenbaums. Grundsätzlich sind alle dort angegebenen Relativbeziehungen zu schätzen, was im ersten Ansatz auf einen ‘‘Gesamt’’-Schätzer führt, der alle Formparameter und Relativlagen in einem Schritt gleichzeitig schätzt. Dieser ‘‘Gesamt’’-Schätzer Ansatz ist über die Verkopplung der zu schätzenden Größen über den Szenenbaum motiviert. Auf die Abbildung eines Objektmerkmals auf die Kamerabildebene haben nach Gl. 6.26 alle dazwischenliegenden Transformationen Einfluß, können also demzufolge im Umkehrschluß auch aus der Vermessung dieses Merkmals geschätzt werden. Die Verkopplung erfolgt also über das Meßgleichungssystem, wie im folgenden für den Videosensor angegeben:

$$\begin{bmatrix} \underline{y}_{\text{K,O}_1} \\ \vdots \\ \underline{y}_{\text{K,O}_n} \end{bmatrix} = \begin{bmatrix} \underline{C}_{\text{K,O}_1} & \cdots & \underline{0} & \underline{C}_{\text{K,EGO}}^{\text{O}_1} & \cdots & \underline{C}_{\text{K,KB}}^{\text{O}_1} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \underline{0} & \cdots & \underline{C}_{\text{K,O}_n} & \underline{C}_{\text{K,EGO}}^{\text{O}_n} & \cdots & \underline{C}_{\text{K,KB}}^{\text{O}_n} \end{bmatrix} \begin{bmatrix} \underline{x}_{\text{O}_1} \\ \vdots \\ \underline{x}_{\text{O}_n} \\ \underline{x}_{\text{EGO}} \\ \underline{x}_{\text{FS}} \\ \underline{x}_{\text{S}} \\ \underline{x}_{\text{KA}} \\ \underline{x}_{\text{KB}} \end{bmatrix} \quad (6.27)$$

Wie zu sehen ist, bilden die linken Spalten der  $C$ -Matrix eine Blockdiagonalstruktur - der Zustand der Fremdobjekte ist nur aus deren eigenen Merkmalen zu schätzen. Der rechte Teil hingegen beinhaltet die Zustände der Transformationen vom Welt-Koordinatensystem bis zum Kamerabildebenen-Koordinatensystem, welche alle Merkmale beeinflussen.

Weiterhin lassen sich zwei Kategorien von Lage- und Formparametern unterscheiden:

1. Die an sich zeitlich “nichtvariablen“ Beziehungen zwischen festen Objekten. Dies sind in der Hauptsache die Sensoren und ihre Lage bzgl. des Sensorkoordinatensystem und weiter bzgl. des EGO-Koordinatensystems. Diese sind nach der Montage bzw. Änderungen am Sensor oder Fahrzeug in einem Kalibriergang “offline“ zu bestimmen.
2. Die zeitlich variablen Relativlagen, die laufend “online“ zu ermitteln sind. Dies sind die Form und Lage der Objekte, die EGO-Lage zum Welt-Koordinatensystem und die EGO-Lage zur Fahrspur sowie die Formparameter der Fahrspur.

Dieser Kategorisierung folgend erfolgt in dieser Arbeit eine Entkopplung der “offline“ zu bestimmenden Zustände von den “online“ zu schätzenden.

$$\begin{bmatrix} \underline{y}_{K,O_1} \\ \vdots \\ \underline{y}_{K,O_n} \end{bmatrix} = \begin{bmatrix} \underline{C}_{K,O_1} & \cdots & \underline{0} & \underline{C}_{K,EGO}^{O_1} & \underline{C}_{K,FS}^{O_1} \\ \vdots & \ddots & \vdots & \vdots & \vdots \\ \underline{0} & \cdots & \underline{C}_{K,O_n} & \underline{C}_{K,EGO}^{O_n} & \underline{C}_{K,FS}^{O_n} \end{bmatrix} \begin{bmatrix} \underline{x}_{O_1} \\ \vdots \\ \underline{x}_{O_n} \\ \underline{x}_{EGO} \\ \underline{x}_{FS} \end{bmatrix} \quad (6.28)$$

Jetzt sind neben den Objekten selbst nur noch die Eigenfahrzeugbewegung sowie die Fahrspurparameter im Meßsystem enthalten.

Auf den ersten Blick haben die Formparameter der Fahrspur zwar nichts mit der Objektvermessung zu tun, da diese über die Videovermessung aus den Fahrspurbegrenzungen ermittelt wird. In Fällen, in denen die Vermessung der Fahrspurbegrenzungen jedoch instabil ist oder widersprüchliche Ergebnisse bringt, kann es durchaus sinnvoll sein, die Schätzung des Fahrspurverlaufs durch die Hereinnahme von ausgezeichneten Objekten (z. B. Leitplanken, Leitpfosten, vorausfahrende Fahrzeuge) zu stützen. Diese Schätzung ist aber eine Ergänzung des obigen Szenenbaums, der unabhängig von der sonstigen Zustandsschätzung der Objekte verläuft und daher auch von diesem entkoppelt ist.

Weiterhin ergibt eine Schätzung der Eigenfahrzeugbewegung aus der Vermessung bewegter Fremdobjekte jedoch i. a. keinen Sinn, da hierzu bereits die Fahrzeugsensorik in Form von Raddrehzahlgeber, Lenkwinkelgeber und Gierratensensor vorhanden ist und andererseits noch die Relativlage zur Fahrbahn geschätzt wird. Diese Möglichkeit der Schätzung wird daher auch nicht genutzt, d. h. die entsprechenden Elemente der Jacobi-Matrix werden für die eigentliche Schätzung zu 0 gesetzt. Nicht zu vernachlässigen ist allerdings die Abhängigkeit der Meßwerte von den Unsicherheiten der Eigenbewegung, die im Rahmen der Datenassoziation zu größeren Akzeptanzintervallen führen. Es folgt, daß die zur Eigenbewegung gehörigen Matrixelemente zunächst berechnet und zur Prädiktion der Innovationskovarianz herangezogen werden. Vor dem an die Assoziation anschließenden Innovationsschritt werden sie dann allerdings zu 0 gesetzt.

$$\underline{P}^* = \underline{\Phi}(T) \hat{\underline{P}} \underline{\Phi}^T(T) + \underline{Q} \quad (6.29)$$

$$\underline{P}^* = \underline{\Phi}(T) \begin{bmatrix} \hat{\underline{P}}_{OO} & \hat{\underline{P}}_{O,EGO} \\ \hat{\underline{P}}_{O,EGO} & \hat{\underline{P}}_{EGO,EGO} \end{bmatrix} \underline{\Phi}^T(T) + \begin{bmatrix} \underline{Q}_O & \underline{0} \\ \underline{0} & \underline{Q}_{EGO} \end{bmatrix} \quad (6.30)$$

Da wir aber anhand dieses Meßgleichungssystems die Fahrzeugeigenbewegung nicht schätzen, sind auch die damit verbundenen Schätzfehlermatrizen  $\hat{\underline{P}}_{O,EGO}$  und  $\hat{\underline{P}}_{EGO,EGO}$  hier zu



0 zu setzen, so daß wir  $\underline{P}^*$  zu

$$\underline{P}^* = \begin{bmatrix} \underline{\Phi}_O(T)\hat{\underline{P}}_{OO}\underline{\Phi}_O^T(T) + \underline{Q}_O & \underline{0} \\ \underline{0} & \underline{Q}_{\text{EGO}} \end{bmatrix} \quad (6.31)$$

erhalten. Für die Innovationskovarianz  $V$  ergibt sich mit  $\underline{V} = \underline{C}\underline{P}^*\underline{C}^T + \underline{R}$  und unter Benutzung von 6.29:

$$\underline{V}_{K,O} = \begin{bmatrix} \underline{C}_{K,O} & \underline{C}_{K,\text{EGO}} \end{bmatrix} \begin{bmatrix} \underline{P}_O^* & \underline{0} \\ \underline{0} & \underline{P}_{\text{EGO}}^* \end{bmatrix} \begin{bmatrix} \underline{C}_{K,O}^T \\ \underline{C}_{K,\text{EGO}}^T \end{bmatrix} \quad (6.32)$$

Die Elemente der Jacobi'schen Meßmatrix werden durch numerische Differentiation mittels des beidseitigen Differenzenquotienten berechnet. Für die in die Relativlagen eingehenden Zustände werden hierfür die homogenen Transformationsmatrizen bei einer Variation der Zustandsgröße  $x_{Di}$  um  $\Delta x_{Di}$  bestimmt. Die Transformationsbeziehung für das Merkmal vom Objektkoordinatensystem in das Koordinatensystem des Sensors zerfällt dabei i. a. in drei Teile:

$$T_G = T_P T_\Delta T_A \quad (6.33)$$

worin  $T_G$  die Gesamttransformation,  $T_\Delta$  die den zu differenzierenden Zustand betreffende Transformation und  $T_P$  und  $T_A$  die vor respektive hinter  $T_\Delta$  befindlichen Relativlagen sind.

$$\frac{\partial P_M}{\partial x_{Di}} = \frac{T_P T_\Delta(x_{Di} + \Delta x_{Di}) T_A P_O(\underline{x}_F) - T_P T_\Delta(x_{Di} - \Delta x_{Di}) T_A P_O(\underline{x}_F)}{2\Delta x_{Di}} \quad (6.34)$$

Für die numerische Differentiation der Formparameter  $x_{Fi}$  erfolgt die Variation um  $\Delta x_{Fi}$  über die Berechnung der daraus resultierenden Merkmalpositionen in Objektkoordinaten und anschließende Transformation mit  $T_G$  in das Koordinatensystem des Sensors:

$$\frac{\partial P_M}{\partial x_{Fi}} = \frac{T_G P_O(x_F + \Delta x_{Fi}) - T_G P_O(x_F - \Delta x_{Fi})}{2\Delta x_{Fi}} \quad (6.35)$$

# Kapitel 7

## Verfahren der Bildverarbeitung

Die in Kapitel 6 vorgestellten Repräsentationsformen für Objekte wie für die Fahrbahn basieren primär auf der Repräsentation der Objektkonturen. Diese Form der Repräsentation wurde bereits im Hinblick auf die zur Verfügung stehenden Verfahren zur Bildmessung ausgewählt, die in dieser Arbeit auf konturbasierten Verfahren beruht. Konturbasierte Meßverfahren ermitteln Punkte mit hohen Grauwertgradienten im Vergleich zur Umgebung, die Kandidaten für Objektkonturen bilden. Über die Einbeziehung der Grauwerte der angrenzenden Flächen und Maßzahlen für die Homogenität von Flächen entlang von Konturen werden indirekt aber auch Flächenmerkmale berücksichtigt.

Es wird zunächst auf das reine Meßverfahren eingegangen, gefolgt von den Möglichkeiten, durch Merkmalsaggregation von den primären Konturmeßwerten zu höheren Merkmalen in Form von Kanten oder Linien zu gelangen.

### 7.1 Konturvermessung

#### 7.1.1 Bildverarbeitungs-Operatoren

Die Verfahren zur Konturvermessung basieren in der Regel auf dem Einsatz von Korrelationsoperatoren, d. h. einer Faltung des Bildes (oder eines Bildbereiches) mit einer Filtermaske. Die ersten Verfahren dieser Art der Kantenvermessung stammen von [Roberts 63]. Heute weitverbreitete Verfahren sind der Sobel- und der Prewitt-Operator (Abb. 7.1). Aufgrund der kleinen Faltungsmaske ist die Richtungsselektivität beider Operatoren gering. Durch Anwendung beider orthogonaler Maskenrichtungen kann ein vollständiges Betrags- und Richtungsbild des Gradienten ermittelt werden. Demgegenüber steht die aufgrund der Operatorgröße geringe Glättung und demzufolge hohe Rauschempfindlichkeit. Diese läßt sich durch eine vorherige Filterung des Bildes, beispielsweise durch einen Gauß'schen Tiefpaß verbessern. Die Ermittlung von Konturen aus Gradientenbetrags- und Gradientenrichtungsbild erfolgt etwa über ein Ausdünnen des Gradientenbetragsbildes, das nur noch die Gradienten der Pixel enthält, an denen die jeweils betragsmäßig größten Gradientenwerte einer Kontur - die Extremwerte - auftreten. Aufgrund ihrer Einfachheit und des festen

<table><tr><td>-1</td><td>0</td><td>+1</td></tr><tr><td>-2</td><td>0</td><td>+2</td></tr><tr><td>-1</td><td>0</td><td>+1</td></tr></table>	-1	0	+1	-2	0	+2	-1	0	+1	<table><tr><td>-1</td><td>-2</td><td>-1</td></tr><tr><td>0</td><td>0</td><td>0</td></tr><tr><td>+1</td><td>+2</td><td>+1</td></tr></table>	-1	-2	-1	0	0	0	+1	+2	+1	<table><tr><td>-1</td><td>0</td><td>+1</td></tr><tr><td>-1</td><td>0</td><td>+1</td></tr><tr><td>-1</td><td>0</td><td>+1</td></tr></table>	-1	0	+1	-1	0	+1	-1	0	+1	<table><tr><td>-1</td><td>-1</td><td>-1</td></tr><tr><td>0</td><td>0</td><td>0</td></tr><tr><td>+1</td><td>+1</td><td>+1</td></tr></table>	-1	-1	-1	0	0	0	+1	+1	+1
-1	0	+1																																					
-2	0	+2																																					
-1	0	+1																																					
-1	-2	-1																																					
0	0	0																																					
+1	+2	+1																																					
-1	0	+1																																					
-1	0	+1																																					
-1	0	+1																																					
-1	-1	-1																																					
0	0	0																																					
+1	+1	+1																																					
horizontaler	vertikaler	horizontaler	vertikaler																																				
Sobeloperator		Prewittoperator																																					

Abbildung 7.1: Standardmaskenoperatoren zur Gradientenbestimmung

Ablaufschemas ist diese Vorgehensweise besonders in Systemen mit starrer Bildvorverarbeitungsstufe, beispielsweise durch Signalprozessoren oder FPGA's, beliebt. Bei anderen Systemen ist die notwendige Rechenzeit nicht unbeträchtlich und der Nutzen vor allem für die modellbasiert "prädizierten" Merkmale verfolgter Objekte begrenzt. Ganz abgesehen davon ist die erforderliche Nachbearbeitung in Form einer Konturpunktverkettung/Hough-Transformation noch beträchtlich.

Am ISF werden daher bevorzugt Verfahren eingesetzt, die sich im wesentlichen durch zwei Hauptmerkmale beschreiben lassen:

1. Es werden, um den Echtzeitanforderungen zu genügen, keine kompletten Bilder auf der ikonischen Ebene ausgewertet, sondern lediglich Bildausschnitte, welche verfolgte Objekte oder relevante Szenenbereiche enthalten, in denen Objekte erwartet werden.
2. Es wird eine größere Vielfalt an Maskenoperatoren eingesetzt, welche eine differenziertere, dedizierte Vermessung gestatten.

Die allgemeinen Korrelationsmasken wurden am ISF mit den Tools Kronos [D. Dickmanns 94] bzw. Cronos [Fürst 97] in einer - auf [Mysliwetz 90] basierenden, sehr effizienten Implementierung für General Purpose Prozessoren umgesetzt. Die Maskenoperatoren haben die in Abb. 7.2 dargestellte Gestalt mit den folgenden Parametern:

- Maskenhöhe (3, 5, 9, 17)
- Maskenwinkel in Schritten von  $[90/(\text{Maskenhöhe} - 1)]^\circ$
- Anzahl der Spalten gleich Null (0...5)
- Anzahl der Spalten gleichen Vorzeichens ungleich Null (1...10)
- Besetzung der Spalten (ternär, "Rampe", "Hügel", "Tal" [Fürst 97])

Beim Einsatz derart parametrierter Korrelationsmasken muß die Bewertung der Ergebnisse stets im Hinblick auf die Maskenauslegung erfolgen. (Anm. : Diese Bemerkung gilt im Sinne der Meßtechnik naturgemäß für jede Art von Messung, da in die Beobachtung stets

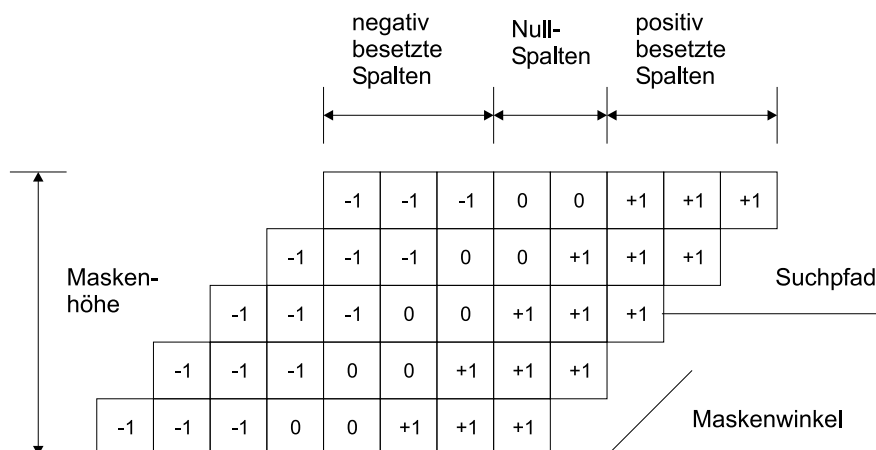


Abbildung 7.2: Allgemeiner Korrelationsoperator

nicht nur die beobachtete Größe, sondern auch die Meßvorrichtung miteingeht.) Der maximale Korrelationswert ergibt sich nur bei Signalverläufen, für die der Operator ein sog. “Matched Filter“ bildet. Je größer die Abweichung des Bildinhalts von dieser “Meßerwartung“ ist, desto niedriger fällt der Korrelationswert aus. Generell läßt sich festhalten, daß eine Korrelationsmaske desto selektiver ist, je größer die Maskenhöhe und je geringer die Maskenbreite sind.

Im Sinne der Echtzeitbildverarbeitung und abgestimmt auf die erwartungsbasierte Merkmalsuche erfolgt die Anwendung der Maskenoperatoren entlang sog. Suchpfade. Diese Suchpfade dürfen horizontal, vertikal oder diagonal unter Winkeln von  $45^\circ$  verlaufen, da aufgrund der diskreten Geometrie nur dann gewährleistet ist, daß kein Pixel mehr als einmal in die Messung eingeht. Als Ergebnis der durch Suchpfadposition und Suchpfadwinkel kommandierten Messungen erhält man nicht direkt das Faltungsergebnis, sondern vielmehr die anhand des Faltungsergebnisses ermittelten Extremwerte, die betragsmäßig oberhalb einer vorgebbaren (Rausch-)Schwelle liegen. Zusätzlich zu diesen Diskontinuitätsinformationen werden auch die Grauwerte der von dieser Diskontinuität begrenzten Flächen geliefert. Damit ist - in begrenztem Maße - eine Art flächenbasierter Bildverarbeitung möglich, die - wie man noch sehen wird - sowohl im Hinblick auf räumliche Kontinuitätsbedingungen bei der Linienbildung als auch bzgl. der zeitlichen Zuordnung von Meßwerten in aufeinanderfolgenden Bildern einer Bildfolge ausgesprochen nützlich sind. Beim Einsatz der o. g. Werkzeuge ergaben sich mehrere Problempunkte, die zu den im folgenden gezeigten Erweiterungen führten.

### 7.1.2 Bestimmung der Extremwerte

Die Bestimmung der betragsmäßig oberhalb einer Schwelle gelegenen Extremwerte ist mathematisch gesehen eine triviale Operation, bei der für die Maskenantwort  $r$  bei gegebener Schwelle  $t$  gelten muß:

$$r(i) \text{ ist Extremum, wenn } (r(i) > t) \wedge (r(i) > r(i-1)) \wedge (r(i) > r(i+1)).$$

Obwohl diese Methode analytisch vollkommen korrekt ist, liefert sie unter der Annahme, daß das Extremum als Diskontinuität zwei homogene Flächen voneinander trennt nicht zwangsläufig das gewünschte Ergebnis. Wenn man die Grauwerte der begrenzten Flächen als wesentliches Attribut verwertet, muß gleichzeitig die Homogenität dieser Flächen auch sichergestellt sein. Um die Einhaltung dieser Nebenbedingung zu prüfen, gibt es zwei Möglichkeiten:

1. Die vordere und hintere Maskenhälfte wird an den Extremwertpositionen auf Homogenität geprüft, und das Ergebnis dieser Prüfung wird beim Meßergebnis vermerkt.
2. Es werden nur Extremwerte als Meßergebnisse akzeptiert, welche die Homogenitätsbedingung für beide Maskenhälften erfüllen.

Im ersten Fall wird die Entscheidung, welche Extremwerte wie auszuwerten sind, auf einen späteren Zeitpunkt verschoben. Diese Methode bietet den Vorteil, daß eine explizite Gruppenbildung von mehreren Extrema erfolgen kann, die als Gruppe dann einen Grauwertübergang zwischen homogenen Flächen darstellen. Der zweite Fall kommt einer Vorfilterung der Meßergebnisse gleich, die zwar nur die Extrema zwischen homogenen Flächen liefert, allerdings auch nur bezogen auf die gewählte Maskenparametrierung.

### 7.1.3 Maskenoperatoren mit variablen Nullspalten

Alternativ zu diesem Verfahren kann man zusätzlich eine Homogenitätsbetrachtung in die Konturpunktbestimmung einbeziehen. Das hierzu verwendete Verfahren entspricht einem Maskenoperator mit einer variablen Anzahl an Nullspalten. Das eigentliche Bezugsobjekt bildet dann die Homogenität des unter der vorderen und hinteren Maskenhälfte befindlichen Flächenausschnitts. Deren Homogenität kann dabei durch:

- eine Maskenantwort unterhalb einer zu wählenden Schwelle, d. h. nur in Suchpfadrichtung, oder durch
- eine Varianz der Maskenhälfte unterhalb einer zu wählenden Schwelle bestimmt werden.

Im Hinblick auf die Rechenleistung bietet es sich an, auf die erste Möglichkeit zurückzugreifen, da die dafür benötigten Werte sowieso im Rahmen der Vermessung bestimmt werden.

Die vordere Maskenhälfte wird bei diesem Verfahren solange in Suchpfadrichtung verschoben, bis die Homogenitätsbedingung verletzt wird. Es wird dann weiter in Suchpfadrichtung fortgeschritten, bis diese Schwelle (betragsmäßig) wieder unterschritten wird. Damit sind zwei benachbarte, unter der gewählten Bedingung homogene, Flächenausschnitte in Größe einer Maskenhälfte gefunden. Im Anschluß läßt sich nun die Art der Diskontinuität bestimmen. Zur Charakterisierung der Diskontinuität gibt es folgende Möglichkeiten:

- Es wird der stärkste Extremwert innerhalb des bestimmten Intervalls gewählt.
- Es wird der mittlere Gradient in dem bestimmten Intervalls gewählt.
- Es wird die Grauwertdifferenz zwischen den beiden bestimmten homogenen Flächenausschnitten vor und hinter der Diskontinuität gewählt.

Da nur im zweiten und dritten Fall alle Werte innerhalb des Intervalls berücksichtigt werden, sind diese beiden Möglichkeiten vorzuziehen.

Als Beispiel ist in Abb. 7.3 ein verschliffener, nicht monotoner Grauwertübergang zwischen zwei Flächen dargestellt. Dieser wird mit den in der Abbildung gezeigten Korrelationsoperatoren auf einem horizontal von links nach rechts gerichteten Suchpfad vermessen. Man kann erkennen, daß bei Verwendung von Masken mit 3 Nullspalten und 3 Einsspalten, der Übergang am schärfsten und am eindeutigsten als *ein* Übergang gefunden wird, während in anderen Fällen gleich mehrere Extrema entdeckt werden.

Es soll aber nochmals darauf hingewiesen werden, daß mit diesem Ansatz lediglich von einem anderen Modell ausgegangen wird: Statt der Diskontinuität der Kanten steht beim Verfahren der variablen Nullspalten die Homogenität der Flächen im Mittelpunkt. Das Vorgehen ist ansonsten ebenso modellgetrieben: Das Verfahren bevorzugt stets die durch die Parametrierung bevorzugten Grauwertverläufe.

#### 7.1.4 Vermessung von Konturkanten

Die im vorherigen Abschnitt gezeigten Verfahren dienen der Bildvermessung an bestimmten Bildpositionen oder in bestimmten Bildregionen bis hin zum kompletten Bild. Da es sich bei den zu vermessenden Bildmerkmalen um Konturkanten bzw. -linien handelt, benötigt man je Kontur mehrere solcher Bildmessungen zu einer sicheren Bestimmung der Lage und Form dieser Kontur. Die einfachste Methode besteht in der Vermessung der relevanten Bildregionen durch Maskenoperatoren geringer Selektivität. Die einfachen Operatoren der Größe  $3 \times 3$  sind allerdings aufgrund ihres Nachbearbeitungsaufwands und ihrer Rauschempfindlichkeit nur bedingt geeignet. Darüberhinaus besteht die Gefahr, daß man schwach ausgeprägte Merkmale aufgrund der geringen Selektivität der Operatoren durch die Wahl der Rauschschwelle unterdrückt. Daher gilt es, folgende Fälle zu unterscheiden:

- Vermessung vorhergesagter Konturen, d. h. Merkmalen bereits instanziierteter Objekthypothesen ("Tracking"-Messung)
- Vermessung nicht besetzter Raumbereiche, d. h. Bildbereichen in denen Objekte erwartet werden ("Detektions"-Messung)

Im letzten Fall werden - basierend auf Hintergrundwissen über die Szene - bestimmte Szenenbereiche auf das Auftauchen "neuer" Objekte hin überwacht. Auf der Basis der geometrischen Informationen aus dem Szenenbaum werden die zu überwachenden Bereiche

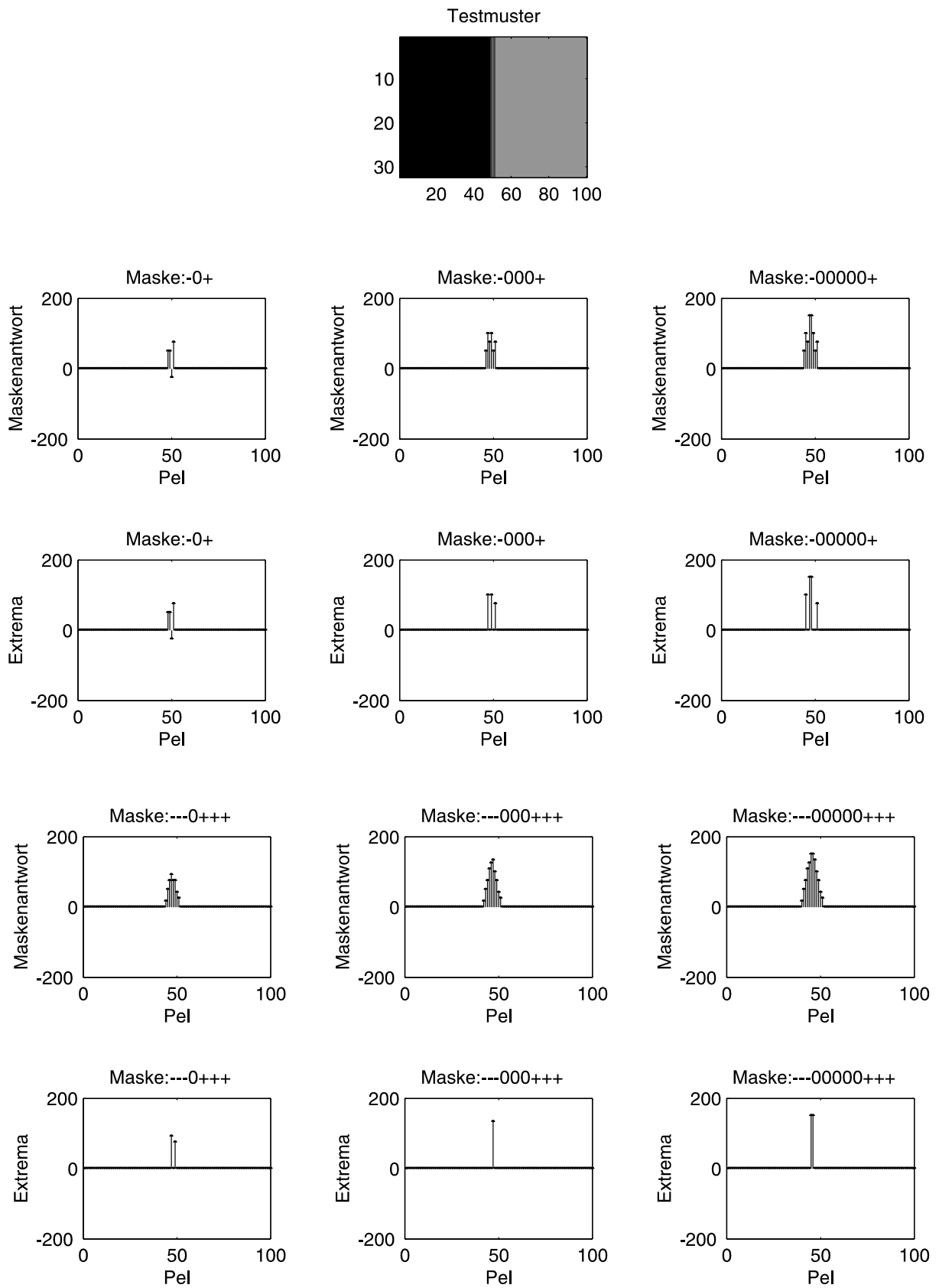


Abbildung 7.3: Diskontinuitätenbestimmung mit unterschiedlichen Maskenbesetzungen

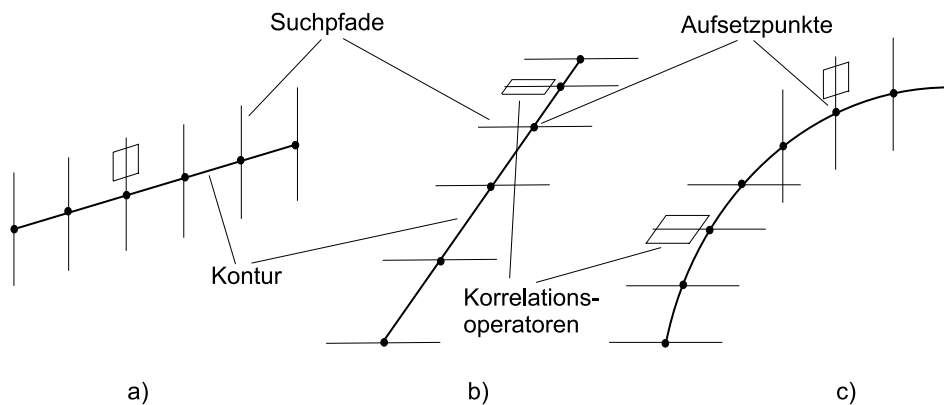


Abbildung 7.4: Konturen mit Aufsetzpunkten und Suchpfaden von Maskenoperatoren

in Ermangelung besseren Wissens mit einem horizontalen und vertikalen Gitter von o. g. unselektiven Bildmessungen der Operatorgröße  $3 \times 3$  bzw.  $5 \times 5$  vermessen.

Im Falle vorhergesagter Konturen erfolgt hingegen eine gerichtete Vermessung. Bei den Konturlinien kann es sich sowohl um Kanten im Sinne der Bildverarbeitung (ein Grauwertübergang) als auch um Linien im Sinne der Bildverarbeitung (Kombination aus zwei Grauwertübergängen entgegengesetzten Vorzeichens) handeln. Sowohl die Vermessung von Linien als auch die Vermessung von Flächen werden zurückgeführt auf die Vermessung einfacher Kanten, und einer nachgeschalteten Merkmalsaggregation. Bei den genannten linienhaften Strukturen kann es sich nach Kapitel 6 sowohl um gerade als auch um gekrümmte Formen in 3D - folglich auch in 2D - handeln. Mit Ausnahme von Geraden lassen sich Kurven in 3D, die einer perspektivischen Projektion unterworfen werden nicht analytisch in 2D darstellen. Die gängige Praxis besteht in der Beschreibung dieser Kurven in 2D durch einen Polygonzug. Die notwendige Zahl an Polygonpunkten ist abhängig von der Krümmung der Kontur. Je nach Form und Lage der Kontur in 3D ergibt sich in Abhängigkeit von den Aspektbedingungen die Ausprägung des die Kontur beschreibenden Polygonzugs in 2D.

Je nach Länge des Polygonzugs im Bild liegen die Polygonpunkte enger oder weiter voneinander entfernt. Daher wird unter den Polygonpunkten eine Auswahl zur Vermessung in Kombination mit der Maskenhöhe getroffen. Aus dem Wissen heraus, daß es sich um äquidistante Punkte in 3D handelt, besteht das Ziel der Vermessung darin, die Messungen so zu parametrieren, daß an den einzelnen Meßpunkten in 2D möglichst Maskenoperatoren eingesetzt werden, die in 3D einer gleichen Modellvorstellung der Kante entspringen. D. h. beispielsweise, daß die Masken jeweils in etwa die gleiche Fläche in 3D überstreichen. Dieser Tiefeneffekt wird allerdings aus Effizienzgründen nur in Fällen berücksichtigt, in denen die Kontur eine merkliche Tiefenausdehnung aufweist. Im Rahmen der Fahrumgebungserfassung sind in der Regel die einzigen Objekte, die eine solche nicht zu vernachlässigende Tiefenausdehnung besitzen, jene, welche die Fahrspur bzw. allgemein den Bodenbereich repräsentieren. Falls darüberhinaus auch stärkere Konturkrümmungen auftreten, muß bei der Auswahl der Maskenhöhe auch die lokale Krümmung berücksichtigt werden, was dann zur Senkung der Richtungsselektivität und zur Auswahl geringerer Maskenhöhen und/oder einer höheren Anzahl Nullspalten führen kann, um den Meßfehler in Konturlängsrichtung



gering zu halten. Das kann im Zweifelsfall bis zur Unterdrückung der Messung bei sehr starken Krümmungen führen.

Die Auswahl des Suchpfades erfolgt so, daß die Messung möglichst orthogonal zur Tangente an den Konturpunkt erfolgt. Da die Tangente selbst nicht vorliegt, wird ersatzweise die Sekante vom vorhergehenden zum nachfolgenden Punkt als Entscheidungskriterium herangezogen. Die Auswahl besetzter Spalten erfolgt zunächst anhand von Lookup-Tabellen, die vom Formmodell referenziert werden. Die Länge des Suchpfades kann aus der Innovationskovarianz  $v$  des Meßpunktes in Suchpfadrichtung bestimmt werden. Die Innovationskovarianz ergibt sich wie in Abschnitt 5 beschrieben als Modellunsicherheit im Merkmalsraum als Summe des vorhergesagten Schätzfehlers und des Meßrauschens. Zu ihrer Bestimmung werden die Elemente der Jacobi'schen Matrix benötigt. Eine zu starke modellgetriebene Einschränkung des Suchraums führt allerdings auch dazu, daß bereits kleinste, kurzfristige Störungen zu einem Merkmalsverlust führen, da diese dann nicht mehr innerhalb des Meßbereichs liegen. Unabhängig davon ist eine Begrenzung der Suchpfadlänge nach unten (5 Pixel) sinnvoll, um die Messung nicht ad absurdum zu führen. Eine schematische Darstellung verschiedener Konturformen, der Aufsetzpunkte und Suchpfade der Messung sind in Abb. 7.4 dargestellt.

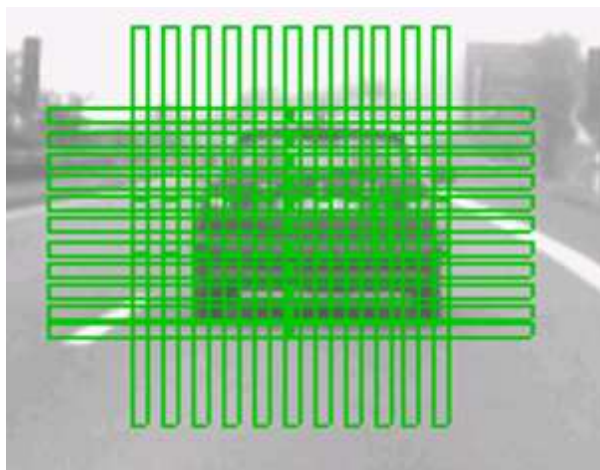
Aufgrund der gewählten Meßverfahren in Kombination mit den weiter unten beschriebenen Techniken zur Merkmalsaggregation gibt es Untergrenzen für die Dimensionen der meßbaren Konturen und damit auch der Objekte im Bild. In Kombination der Objektgröße in 3D sowie der Kameraeigenschaften ergeben sich mit der hier gewählten Anordnung eine maximale Entfernung von ca. 80 m für Fahrzeuge sowie etwa 60 m für Fahrspurmarkierungen.

In Abbildung 7.5 ist die Vermessung eines PKW in der Heckansicht dargestellt. Im Bild 7.5 a) sind die horizontalen bzw. vertikalen Suchpfade der Maskenoperatoren der Höhe 5 bei einem Winkel von  $0^\circ$  bzw.  $90^\circ$  dargestellt. Die anderen Bilder zeigen jeweils das Meßergebnis bei einer unterschiedlichen Parametrierung entweder des Maskenoperators oder aber einer unterschiedlichen Wahl der Extremwertschwelle. Wie man erkennt, besitzt der PKW viele ausgeprägte horizontale innere Strukturen. Im Gegensatz dazu gibt es relativ wenige vertikale innere Strukturen, die zu Extremwerten führen. In Abhängigkeit der Meßparameter sind einzelne Strukturen - insbesondere innere - von der Messung erfaßt oder nicht. Insbesondere die unterschiedliche Wahl der Extremwertschwelle hat massiven Einfluß auf das Meßergebnis, wie man an der gegenüber den Bildern b) und d) deutlich höheren Schwelle in Bild c) erkennt.

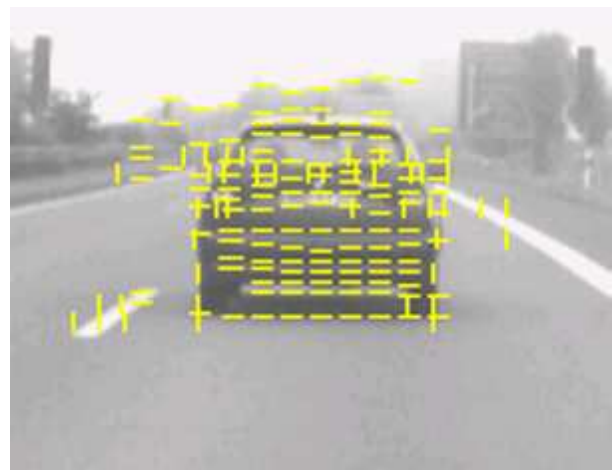
Ebenso ist die Anzahl der inneren Strukturen natürlich von Objekt zu Objekt und in Abhängigkeit der Umgebungsbedingungen stark schwankend.

## 7.2 Linienaggregation

Das Ergebnis der Bildvermessung besteht in den auf den Suchpfaden gefundenen Extrema und deren Attributen. Um aus diesen Konturpunkten die eigentlich in den Modellen repräsentierten Merkmale - die Konturen - zu erhalten, müssen diese Konturpunkte einer Merkmalsaggregation, in diesem Fall einer Linienbildung unterzogen werden.



a) Suchbereiche der Kantenoperatoren  
(Maskenhöhe=5)



b) Gemessene Extrema  
(Schwelle=10, Maske=1-1-0-1-1)



c) Gemessene Extrema  
(Schwelle=30, Maske=1-1-0-1-1)



d) Gemessene Extrema  
(Schwelle=10, Maske=1-1-0-0-1-1)

Abbildung 7.5: Kantenorientierte Bildvermessung an einem Beispiel bei unterschiedlicher Parametrierung des Maskenoperators

Für die erwartungsbasiert vorgenommenen Messungen besteht auch die Möglichkeit der sog. "Nächste-Nachbar-Klassifikation". Dabei wird jeweils das auf dem Suchpfad der prä-dizierten Merkmalsposition nächstgelegene Extremum diesem als Meßresultat zugeordnet. Das Attribut "nächstgelegene" ist hier natürlich im Sinne von "geringstem Abstand" zu verstehen, d. h. nicht zwangsläufig nur den geometrischen Abstand sondern möglicherweise auch weitere Attribute in die Klassifikation einbeziehend. Diese Art der Klassifikation ist zwar schnell, abgesehen von wenigen robusten Anwendungen (z. B. der Fahrspurerkennung) aber zu erwartungsgetrieben. Nur bei sehr eindeutigen Merkmalen und bei einer mit großer Wahrscheinlichkeit zutreffenden Hypothese wird dieses Vorgehen zum Erfolg führen. Man ist aber hier gut beraten, vor allem wenn entsprechende Rechenleistung verfügbar ist, auf die im folgenden geschilderten Methoden zur Merkmalsaggregation überzugehen.

Durch die Merkmalsaggregation erfolgt die Berücksichtigung der Gestaltmerkmale in Form lokaler Nachbarschaften oder lokaler und globaler Merkmalsattribute. Um diese Kriterien zu berücksichtigen, bedarf es einer expliziten Primitivengenerierung, welche die Meßergebnisse einer Kante über alle Suchpfade gemeinsam betrachtet und sowohl nach lokalen Kriterien (Zuordnung benachbarter Konturelemente) als auch nach globalen Aspekten (Gestalt des Merkmals) eine Zuordnung trifft. Bei dieser Primitivengenerierung gibt es folgende Punkte zu beachten:

- Positions-, Orientierungs- und Längenunsicherheit auf Basis des Modellwissens
- Abweichungen vom stetigen Konturverlauf aufgrund von Störeinflüssen (nicht modellierte “Störkanten”)
- “Fehl“-Messungen im Sinne fehlender Messungen (Lücken) im Konturverlauf aufgrund von Störeinflüssen (photometrische Eigenschaften)
- Homogenitätsannahmen über die von den Konturen begrenzten Flächen
- Grauwertprofile als Basis zeitlicher Zuordnung.

Diese qualitativen Kriterien lassen eine Reihe von quantitativen Kriterien zur Bewertung der Messungen und als Ausgangspunkt der Konturgenerierung finden:

- die maximale Anzahl mit Konturpunkten besetzter Suchpfade
- die maximale Anzahl nicht mit Konturpunkten besetzter Suchpfade innerhalb von Konturen, d. h. also “Konturlücken”
- maximale Positionsabweichung von benachbarten Konturpunkten
- maximale Grauwertabweichung der vorderen/hinteren Grenzfläche benachbarter Konturpunkte
- maximale Varianz des mittleren Grauwerts der vor/hinter der Kontur gelegenen Flächen

Im Rahmen dieser Arbeit werden zwei Gruppierungsverfahren eingesetzt. Im einen Fall handelt es sich um ein Verfahren, welches sich am ehesten als lokale, gerichtete Hough-Transformation bezeichnen läßt, im anderen Fall handelt es sich um ein Konturverkettungs/-verfolgungsverfahren.

### 7.2.1 Lokale gerichtete Hough-Transformation

Die Bezeichnung lokale, gerichtete Hough-Transformation leitet sich aus ihrer Verwandtschaft zur klassischen Hough-Transformation ab, denn es handelt sich hierbei um ein Verfahren, das mittels Akkumulation von Konturpunkten eine Häufungsanalyse durchführt,

um auf das Vorhandensein einer Kontur zu schliessen. "Lokal" ist das Verfahren deswegen zu nennen, da es sich i. d. R. nur auf begrenzte Bereiche und nicht auf das gesamte Bild bezieht. Als "gerichtet" ist das Verfahren deshalb zu bezeichnen, da es nicht nach Konturen beliebiger Lage sucht, sondern nur eine bestimmte Auswahl in Länge und Winkellage um die erwartete Lage herum, analog zur Wahl der Maskenoperatoren. Das Verfahren sei anhand einer Skizze erläutert (Abb. 7.6).

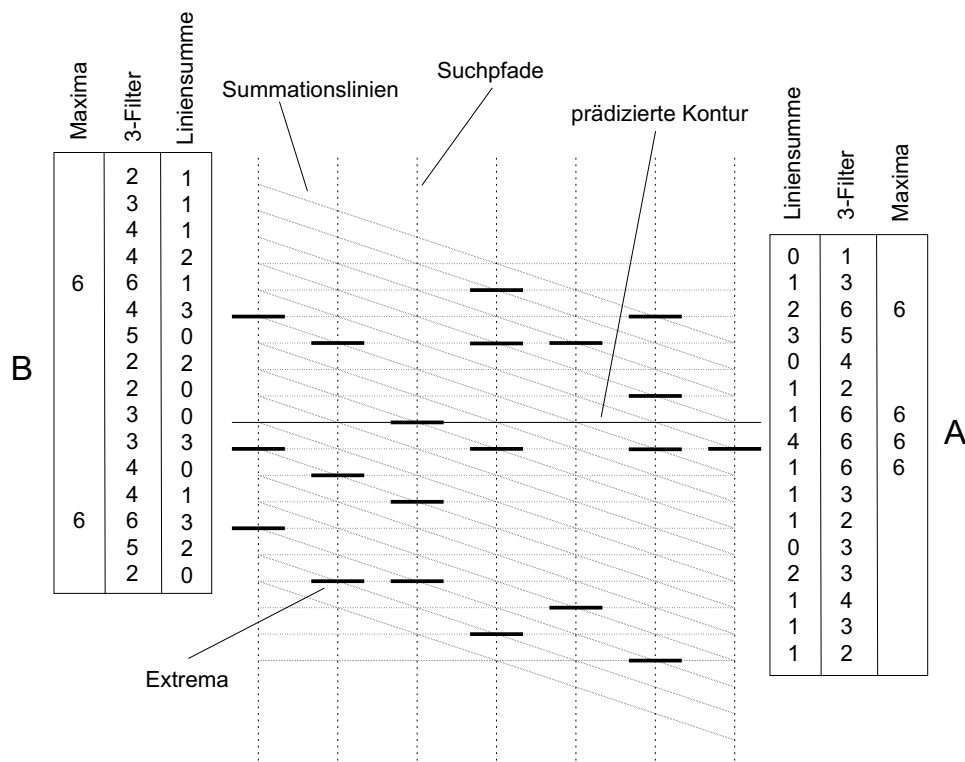


Abbildung 7.6: Linienaggregation mittels lokaler, gerichteter Hough-Transformation

Die erwartete Kontur wird mittels Korrelationsoperatoren auf den dargestellten vertikalen Suchpfaden vermessen. Das Ergebnis der Vermessung sind die angegebenen Extrema auf den Suchpfaden. Wie zu erkennen ist, befindet sich in Nähe der prädiizierten Kontur eine Anhäufung von Extrema. Daneben finden sich aber noch eine weitere Häufung parallel dazu im oberen Meßbereich sowie eine schräg verlaufende Häufung im unteren Meßbereich. Aufgrund der Meß- und Modellfehler und der Modellierungsungenauigkeiten wird man nicht erwarten, daß die Messung eine gerade, geschlossene Kontur parallel zur Erwartung liefert. Die Lageungenauigkeit wird dadurch berücksichtigt, daß man zum einen einen Unschärfbereich von mehreren Pixeln vorsieht und zum anderen auch alternative Winkellagen der Messung berücksichtigt. Zunächst werden die Extrema "zeilenweise" entsprechend der erwarteten Winkellage aufsummiert, im Bild gezeigt für die Winkellagen 'A' für einen horizontalen Verlauf und 'B' für einen schiefen Verlauf (1. Spalte der Summationstabellen). Anschließend werden die Ergebnisse gemäß dem vorgegebenen Unschärfbereich aufsummiert (2. Spalte der Summationstabellen), im Beispiel mit einem Unschärfbereich von 3 Pixeln. Alle Häufungen von Extrema, die oberhalb der für die Konturbildung angesetzten Mindestanzahl Extrema liegen, sind Konturkandidaten (Beispiel: Schwelle = 6 Pixel). Im

Anschluß daran werden alle Konturkandidaten auf die Einhaltung der sonstigen Nebenbedingungen, wie etwa das Vorhandensein von Lücken oder inkompatible Flächengrauwerte geprüft.

Die Extrema werden also entsprechend ihres Abstandes von der Referenzlage - der erwarteten Lage - aufsummiert oder gehäuft, und durch den vorgegebenen Unschärfbereich wird die Quantisierung des Akkumulatorfeldes der Hough-Transformation in gleitender Version nachempfunden. Durch Überlagerung eines "Offsets" von Suchpfad zu Suchpfad können auch Summierungen in einem Winkel zur Referenzlage erfolgen. Daher ist die Namensgebung als lokale, nämlich an der Position der Referenzkontur gelegene und auf diese gerichtete Häufungsanalyse als einer Form der auf einen Unterraum begrenzten Hough-Transformation durchaus zutreffend.

Eine weitere Eigenschaft tritt in obigem Beispiel nicht deutlich zutage, da es sich dort um eine gerade Kontur handelt. Der Bezug auf die Aufsetzpunkte der Referenzkontur bei der Summierung entspricht in gewisser Hinsicht der verallgemeinerten Hough-Transformation [Ballard, Brown 82], denn damit wird die Summierung anderer Konturverläufe in die von Geraden transformiert und kann so ebenfalls schnell und effizient ausgeführt werden.

### 7.2.2 Konturpunktverkettung

Im Gegensatz zur im vorhergehenden Abschnitt besprochenen Häufungsanalyse wird bei der Konturpunktverkettung direkt von den lokalen Zusammenhängen benachbarter Konturpunkte ausgegangen, welche dann durch sukzessive Verkettung zu Konturkurven verknüpft werden. Der grundsätzliche Ablauf sei auch in diesem Fall anhand der o. a. Konturvermessungsskizze - hier für die Erläuterung der Konturpunktverkettung aufbereitet - erläutert (Abb. 7.7).

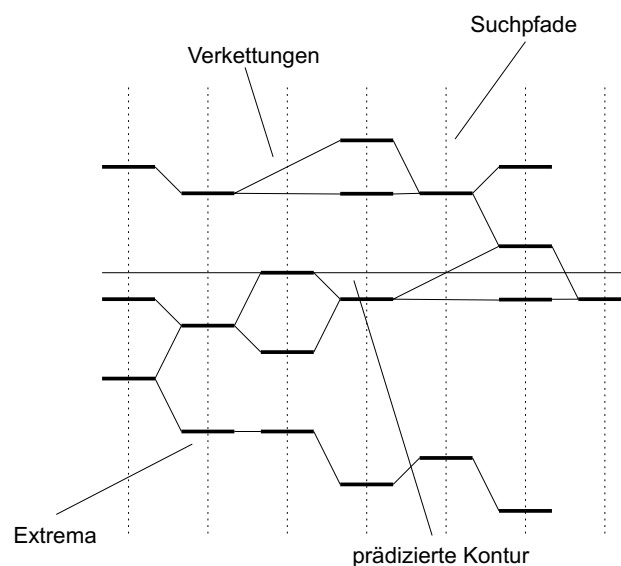


Abbildung 7.7: Linienaggregation durch Konturpunktverkettung

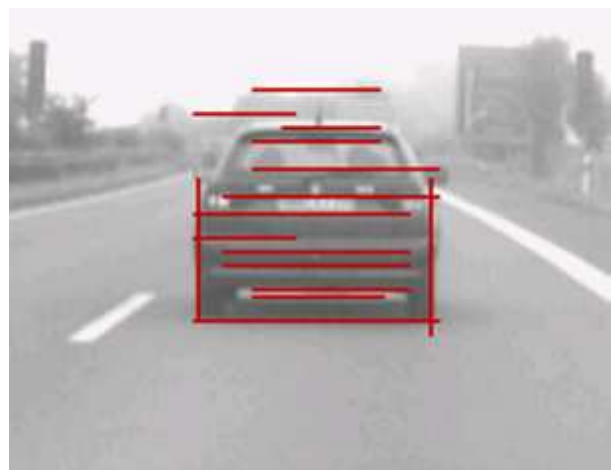
Im ersten Schritt werden alle benachbarten Extrema einem Kompatibilitätstest unterzogen. Dabei müssen sie den vorgegebenen Randbedingungen genügen, damit sie als Verkettungskandidaten in Frage kommen. Hierzu gehört sicher der geometrische Abstand der Extrema voneinander; entsprechendes Modellwissen vorausgesetzt zählen hierzu aber auch kompatible Grauwerte der Grenzflächen oder gleichartige Grauwertübergänge. Die in der Abbildung gezeigte Prinzipskizze geht von einem tolerierbaren Abstand von  $\pm 2$  Pixeln in Suchpfadrichtung aus. Die damit möglichen Verknüpfungen der Extrema sind dünn markiert. Als Kandidaten werden im ersten Schritt lediglich Extrema auf benachbarten Suchpfaden betrachtet. Wenn es innerhalb des Toleranzbereiches jedoch keinen Nachbar-Konturpunkt gibt, werden - bei entsprechender Vorgabe - auch Extrema auf den folgenden Suchpfaden betrachtet. Im zweiten Schritt werden dann alle möglichen Verknüpfungen von Konturpunktfolgen ermittelt und anhand ihrer globalen Eigenschaften auf Einhaltung der vorgegebenen Gütekriterien geprüft.

Bei Einsatz auf detailreichen realen Szenen besteht die Tendenz zu einer kombinatorischen Explosion der möglichen Konturpunktfolgen. Daher bietet es sich an, die Einhaltung der Gütekriterien direkt nach jeder Verkettung zu testen, um die Konturpunktkette entweder zu verwerfen, oder aber in die (endliche) Liste der am besten beurteilten Kandidaten einzuordnen.

Das Ergebnis einer Konturpunktverkettung ist am Beispiel der Szene aus Abschnitt 7.1.4 in Abb. 7.8 zu sehen. Wie zu erkennen ist, führen insbesondere die zahlreich gemessenen Konturpunkte horizontaler Strukturen zu vielen horizontalen Konturkandidaten. Aus diesen gilt es, durch ein entsprechendes Auswahlverfahren diejenigen auszuwählen, die im Formmodell repräsentiert sind. Im Gegensatz dazu fällt die Auswahl der vertikalen Konturen relativ leicht, da hier nur wenige Kandidaten ermittelt wurden.



a) Extrema nach der Kantenvermessung  
(siehe Abb. 7.5)



b) Konturkandidaten nach  
Konturpunktverkettung

Abbildung 7.8: Kantenorientierte Bildvermessung und Konturverkettung an einem Beispiel

## 7.3 Bewertung von Messung und Gruppierung

Sofern eine Messung auf der Basis der Erwartung einer bereits bestehenden Objekthypothese erfolgt, lassen sich für jede aufgesetzte Messung die Abstände der Attribute zu ihrer erwarteten Ausprägung bestimmen. Die Attribute, zu denen sich sinnvolle Abstandsmaße definieren lassen, können sowohl 4D-basiert (geometrische Attribute), also aus dem räumlich-zeitlichen Modell stammend, als auch 2D-basiert (photometrische Attribute), d. h. aus den Merkmalsattributen der vorangegangenen Bildmessungen stammend, sein. Aus den Abständen werden über angenommene Wahrscheinlichkeitsverteilungen die Wahrscheinlichkeiten berechnet, daß ein Merkmal auf der Basis der einzelnen Attribute der Erwartung entspricht. In der Folge läßt sich dann durch Kombination der zu einer Linie aggregierten Extrema aus deren Gütemaßen ein Gesamtgütemaß für die Kontur bestimmen. Bevor aber die Berechnung der Gütemaße für die Einzelextrema wie für die Kontur beschrieben werden, wird zunächst die Bewertung der ausgewählten Attribute sowie die Bestimmung eines Rauschabstandes für die Meßwerte dargestellt.

### 7.3.1 Bewertung der Attribute eines Extremums

#### Zugehörigkeitswahrscheinlichkeit

Die Wahrscheinlichkeit, daß ein gemessener Extremwert dem erwarteten Vorhersagewert entspricht, wird auf der Basis des geometrischen Abstandes sowie der photometrischen Attribute bestimmt. Beide Wahrscheinlichkeiten werden als normalverteilt angenommen:

- Geometrischer Abstand auf dem Suchpfad:

$$P_d = \frac{1}{\sqrt{2\pi}v_d} \exp\left(-\frac{\Delta_d^2}{2v_d^2}\right)$$

mit

$\Delta_d$  = Positionsdivergenz der gemessenen zur erwarteten Position  
des Extremums auf dem Suchpfad

$v_d$  = Innovationskovarianz entlang des Suchpfades

- Photometrische Abstände der Grauwerte:

$$P_{Brt} = \frac{1}{\sqrt{2\pi}\sigma_{Brt}} \exp\left(-\frac{\Delta_{Brt}^2}{2\sigma_{Brt}^2}\right)$$

mit

$Brt$  = Helligkeit (Brightness) des Flächenausschnitts

$\sigma_{Brt}$  = angenommene Varianz der Helligkeiten

Bei diesen Wahrscheinlichkeiten handelt es sich um bedingte Wahrscheinlichkeiten im Bayes'schen Sinne, denn sie geben jeweils die Wahrscheinlichkeit dafür an, ob ein gemessenes Extremum einem erwarteten Merkmal zugeordnet werden kann unter der Bedingung des Eintretens bestimmter geometrischer oder photometrischer Attributwerte.

### Bestimmung eines Rauschabstands

Neben diesen für die Abstandsberechnung relevanten Attributen kann der Gradient eines Extremums dazu genutzt werden, ein Signal-Rausch-Verhältnis (SNR) zu ermitteln. Hohe Korrelationswerte entsprechen einem hohen SNR. Ein solches Extremum ist zeitlich i. a. gut zuzuordnen, da es meßtechnisch stabil zu detektieren ist, im Gegensatz zu Extrema, die knapp oberhalb der Rauschschwelle liegen und welche daher leicht bei der folgenden Messung unter diese Schwelle fallen können. Diese Eigenschaft sollte auch in der Gewichtung der Extrema Eingang finden. Der Betrag des normierten Gradienten  $G$  liegt im Bereich  $0 \dots 255$  (bei 8 Bit Grauwerten). Durch Einführung der Rauschschwelle  $G_{\min}$  kann er nur Werte im Intervall  $G_{\min} \dots 255$  annehmen. Unter der Annahme, daß alle oberhalb einer bestimmten Schwelle liegenden Gradienten relativ sichere Messungen bedeuten und nur die im Bereich der Rauschschwelle befindlichen Messungen unsicher sind, bietet es sich an, eine logarithmische Kennlinie für das SNR anzusetzen:

$$d_{\text{SNR}} = \frac{\ln G + 1}{\ln 256}$$

Neben diesem, den Abstand zur Rauschschwelle charakterisierenden SNR, kann noch ein zweites Signal-Rausch-Verhältnis definiert werden, welches den Vorteil besitzt, als Teil des Meßrauschens betrachtet zu werden und als solches Eingang in die Kalman-Filterung der Meßdaten zu finden. Betrachtet man innerhalb des Suchpfades eines Meßpunktes alle für eine Zuordnung in Frage kommenden Extrema, so soll deren Anzahl innerhalb des Suchbereichs, der wie o. a. abhängig von der Innovationskovarianz gewählt wird, als eine Art Meßrauschen im Sinne der Unsicherheit der Zuordnung betrachtet werden. Liegt nur ein gemessenes Extremum vor, so ist das Unsicherheitsrauschen minimal, andernfalls erhöht es sich proportional zur Anzahl der gefundenen Extrema.

Nun muß auch bei Vorliegen eines einzelnen Extremums dieses noch lange nicht korrekt im Sinne der Modellerwartung sein. Diese Unsicherheit ist hier jedoch nicht gemeint, sondern allein die Unsicherheit, die entsteht durch den Zwang, sich zwischen mehreren Meßwerten zu entscheiden. Im übrigen wird die Entscheidung ja nicht auf der Basis der einzelnen Suchpfade getroffen, sondern über die weiteren Gestaltannahmen.

Das Rauschen der Zuordnung wird als proportional zur Anzahl der im Suchbereich gefundenen Extrema angesetzt (Abbildung 7.9):

$$\sigma_Z = k_{\sigma_z}(N_{\text{Extrema}} - 1) \quad \text{für} \quad N_{\text{Extrema}} > 0$$

Bei Vorliegen entsprechenden Modellwissens (z. B. Gradientenrichtung) kann die Mehrdeutigkeit und damit die Anzahl zu beachtender Extrema möglicherweise eingeschränkt werden und führt somit zu einem geringeren Rauschanteil.

### 7.3.2 Gesamtgütemaß eines Extremums

Um die für die Zuordnung in Frage kommenden Extremwerte vergleichen zu können, muß auf der Basis der aus den Einzelattributen bestimmten normierten Zugehörigkeitswerte



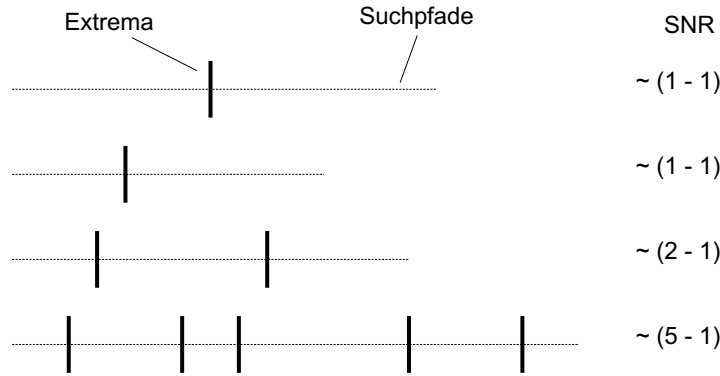


Abbildung 7.9: “Zuordnungsrauschen“ aufgrund ambivalenter Extrema auf Suchpfaden

ein Gesamtzugehörigkeitswert bzw. Gütemaß bestimmt werden. Zur Bestimmung dieses Gütemaßes folgende Überlegungen:

1. Die Art der Verknüpfung der Attribute ist eine Konjunktion, da bereits ein nichtzutreffendes Attribut zum Ausschluß des jeweiligen Meßwertkandidaten führen soll.
2. Die Attribute sollen entsprechend ihrer jeweiligen Relevanz gewichtet werden.
3. Unter der Annahme, daß ein wiederholt zugeordnetes Merkmal mit größerer Wahrscheinlichkeit ein korrekt zugeordnetes Merkmal (im Sinne des jeweiligen zugrundeliegenden Modells) ist, sollte eine zusätzliche zeitliche Gewichtung eingeführt werden.

Um die 3. Forderung zu erfüllen, werden die o. g. Zugehörigkeitswerte der Attribute zunächst mit einer zeitlichen Wichtung  $w_{t,i}$  versehen. Bei den photometrischen Attributen kann diese zeitliche Gewichtung im Bereich  $0 \dots 1$  liegen und wird auf dieser Ebene wie u. a. adaptiert. Beim geometrischen Abstand ist die zeitliche Gewichtung *auf dieser Ebene* stets 1, da sie bereits im räumlich-zeitlichen 4D-Modell berücksichtigt wird. Zusätzlich erhalten alle Attributzugehörigkeiten einen absoluten Wichtungsfaktor  $w_{\text{abs},i}$ . Somit ergibt sich als Gesamtgewicht für die Zugehörigkeitswerte der Attribute

$$w_i = w_{\text{abs},i} w_{t,i}$$

Die zeitliche Gewichtung muß zwingend für alle Attribute eines Merkmals separat geführt werden, da aufgrund sich ändernder Aspektbedingungen die Anzahl gültiger, d. h. verwertbarer Attribute variieren kann. Genauso selbstverständlich ist es, daß diese Gewichtungen und Zugehörigkeitswerte auf der 2D-Ebene für jedes (Kamera-)Bild getrennt geführt werden. Als Verknüpfung, die eine gewichtete Konjunktion durchführt, wird auf der Basis empirischer Untersuchungen das gewichtete geometrische Mittel gewählt:

$$d_e = \left( \prod_i d_i^{w_{\text{abs},i} w_{t,i}} \right)^{\frac{1}{\sum_i w_{\text{abs},i} w_{t,i}}}.$$

Das Gesamtgewicht eines Extremums  $w_e$  ergibt sich aus der Summe der Attributgewichtungen:

$$w_e = \sum_i w_{\text{abs},i} w_{t,i}.$$

### 7.3.3 Gesamtgütemaß einer Kontur

Um von den Gütemaßen für ein Extremum zu einem Gütemaß für eine Kontur, die mit einem der o. g. Verfahren gebildet worden ist, zu gelangen, bieten sich folgende Aspekte als Grundlage an:

- Die für die Extrema auf den einzelnen Suchpfaden (isoliert) ermittelten Gütemaße, z. B. in Form einer gewichteten Summe.
- Nachbarschaftseigenschaften der Extrema auf benachbarten Suchpfaden, z. B. mittlerer Grauwert, Varianz der Grauwerte.
- Abstandsmaße bzgl. der zur Konturbildung herangezogenen Referenz. Dies ist allerdings i. a. nur bei Geraden (mittlere quadratische Abweichung) sinnvoll möglich, da für andere Konturen keine analytische Beschreibung in 2D vorliegt.

Aufgrund der Beschränkung auf gerade Konturen wird der geometrische Abstand zur Referenz nicht weiter betrachtet. Das ergibt auch insofern einen Sinn, als durch eine entsprechende Bewertung eine Idealisierung der Erwartung stattfindet. Die zu erfassenden realen Objekte sind durch die Objektmodelle bereits idealisiert dargestellt. Mithin können also etwaige Meßabweichungen sowohl auf Meßfehler als auch auf Modellierungsfehler zurückgehen. Es macht folglich keinen Sinn, diese nicht zu vermeidende Abweichung als Gütekriterium heranzuziehen.

Mehr Aufmerksamkeit verdienen die Nachbarschaftseigenschaften. Die beiden wesentlichen Nachbarschaftseigenschaften sind bereits genannt:

1. Der Mittelwert der Flächengrauwerte beider von der Kontur begrenzten Flächen  $G_{m,Frt/Bck}$ , der sich aus den Grauwerten der einem Konturkandidaten angehörenden Extrema ergibt. Dieser Wert ist nur sinnvoll bei a-priori Kenntnis der Flächengrauwerte und Vorhandensein eines expliziten Beleuchtungsmodells und wird damit nur in den wenigsten Fällen benötigt.
2. Die Varianz der Flächengrauwerte  $G_{v,Frt/Bck}$ , die sich analog der Mittelwerte aus den Extrema der Konturkandidaten ergeben. Deren Berechnung ist in den Fällen sinnvoll, in denen Kenntnisse über die Homogenität der vermessenen Flächen getroffen werden können; da dies oft der Fall ist, erweist sie sich als ungleich relevanter als die Berechnung der Mittelwerte.

## 7.4 Zusammenfassende Wertung

Abschließend soll noch einmal die qualitative Vorgehensweise der Merkmalvermessung und Primitivengenerierung zusammengefaßt werden, um im folgenden Kapitel über relationale Merkmalszuordnung die Grundideen klar vor Augen zu haben.

1. Konturen werden, zerlegt in Konturelemente als Merkmale basierend auf Objektmodellen und Szenenwissen, gezielt anhand einfacher, effizienter Verfahren vermessen.
2. Die Messungen werden bezogen auf die Merkmalsattribute bewertet und über Gruppierungsverfahren zu Kandidaten für die zugrundeliegenden Konturen zusammengefaßt.
3. Für die Konturkandidaten werden Abstandsmaße für die photometrischen Attribute des Mittelwerts und der Varianz der Flächengrauwerte berechnet. Die sich daraus ergebenden Gütemaße werden zusammen mit den zuvor bestimmten Gütemaßen der Merkmalattribute zu einem Gesamtgütemaß für jede Konturlinie zusammengefaßt. Die Zusammenfassung von Gütemaßen erfolgt über empirisch zu bestimmende Gewichtungen anhand des gewichteten geometrischen Mittels, um die konjunktive Eigenschaft der Gütemaße zu unterstreichen.
4. Das so bestimmte Gesamtgütemaß hat folgende wesentliche Eigenschaften:
  - (a) Ausgeprägtere Konturen (Länge im Bild, größere Anzahl gültiger Attribute) erfahren eine höhere Gewichtung.
  - (b) Zeitlich konsistentere Merkmale werden über zeitliche Gewichtungsfaktoren höher gewichtet.

# Kapitel 8

## Hopfield Netze

Wie im Kapitel 5 über rekursive Schätzverfahren bereits erwähnt, ist für die Form- und Bewegungsschätzung mittels Kalman-Filter die eindeutige und korrekte Zuordnung von gemessenen zu erwarteten Meßwerten eine unabdingbare Voraussetzung für die Konvergenz des Schätzverfahrens. Nun stellt im Falle der umfelderfassenden Sensorik aber gerade dieser Assoziationsschritt eine der Hauptproblempunkte dar. Sowohl die Entscheidung, welchem erwarteten Merkmal ein Meßwert zuzuordnen ist, als auch die Entscheidung - insbesondere bei Bildmerkmalen - ob es sich bei einem gemessenen Merkmal überhaupt um einen Meßwert für das modellierte Merkmal handelt oder aber um ein "Stör"-Merkmal - sind nicht trivial.

Als Abstandmaß zwischen Erwartungswert und Meßwert in Verbindung mit rekursiven Schätzverfahren wird oft die Mahalanobis-Distanz genutzt und in erster Näherung auf eine Zuordnung nach dem "Nächste-Nachbarn"-Prinzip (engl. "Next-Neighbour") zurückgegriffen. Bei diesem Zuordnungsprinzip wird jeweils dasjenige gemessene Merkmal zugeordnet, welches den geringsten Abstand zur Erwartung hat. Allerdings konkurrieren i. a. mehrere Merkmale um die Zuordnung, beispielsweise wenn ein gemessenes Daten-Merkmal mehreren erwarteten Modell-Merkmalen zugeordnet werden kann und für mehrere dieser Erwartungswerte gleichzeitig den jeweils geringsten Mahalanobis-Abstand unter allen möglichen Meßwerten darstellt. In diesem Fall ist eine Zuordnungsstrategie gefragt, welche unter Annahme eines Optimierungskriteriums eine globale, lokale oder andersgeartete Zuordnung durchführt. Eine "brute-force"-Zuordnungsstrategie kommt aus Gründen der Rechenzeit meist nicht in Frage; stattdessen gelangen verschiedene Zuordnungsstrategien zum Einsatz, welche i. d. R. eine suboptimale Zuordnung bei begrenzter Rechenzeit ermöglichen.

Ein solches Verfahren stellt z. B. der Auktionsalgorithmus [Bertsekas 01] dar, bei welchem die Meßwerte unter Kostengesichtspunkten im Sinne einer Auktion "meistbietend" versteigert werden. Interessant an den verschiedenen Ausprägungen des Auktionsalgorithmus ist die Möglichkeit, aus suboptimalen Verteilungen der Zuordnungen durch Rückgabe von bereits "ersteigerten" Meßwerten wieder herauszukommen.

Im Bereich der Bildverarbeitung kommt der Zuordnungsproblematik - im Vergleich zu anderen umfelderfassenden Sensoren - aufgrund der Vielzahl und Vielfalt der Merkmale

eine nochmals gesteigerte Bedeutung zu. Vielfach kommen graphbasierte Verfahren zum Einsatz, welche im ersten Schritt aus der Konstruktion eines Attributierten Relationalen Graphen (bzw. Hypergraphen) bestehen. Die Knoten des Attributierten Relationalen Graphen (ARG) werden von den zuzuordnenden gemessenen “Daten“-Merkmale gebildet, während die Kanten den zwischen diesen Primitiven bestehenden Relationen entsprechen. Die Kompatibilitäten oder Abstände der gemessenen “Daten“-Merkmale zu den erwarteten “Modell“-Merkmale bilden die unären Attribute der Knoten, die Kompatibilitäten zwischen den (unterschiedlichen “Modell“-Merkmale zugeordneten) “Daten“-Merkmale führt zu den binären Attributen der Kanten des ARG. Im zweiten Schritt wird dann mittels geeigneter Verfahren versucht, eine im Sinne der jeweiligen Problemstellung optimale Zuordnung zu ermitteln. Eine Vergleichbarkeit oder vergleichende Bewertung der vielen unterschiedlichen zum Einsatz gelangenden Verfahren erscheint kaum möglich - zu unterschiedlich sind deren Voraussetzungen und Ziele. Die Charakterisierung eines Verfahrens kann nach folgenden Gesichtspunkten erfolgen:

- der Art und Bestimmung der verwendeten Merkmalsprimitive
- der Art und Bestimmung der verwendeten unären und binären Attribute
- dem Zuordnungsverfahren

Ein wesentliches Kriterium bildet in allen Fällen der gezielte Einsatz von Vorwissen und Modellwissen, um die ansonsten nahezu unvermeidliche kombinatorische Explosion zu verhindern. D. h. aber nichts anderes, als daß Zuordnungsverfahren einer auf die jeweilige Anwendung abgestimmten Wissensbasis bedürfen.

Die zur Lösung der Zuordnung als eines Optimierungsproblems in Frage kommenden Verfahren lassen sich grob in zwei Kategorien einteilen:

- wahrscheinlichkeitsbasierte Ansätze
- energiebasierte Ansätze

Bei wahrscheinlichkeitbasierten Ansätzen handelt es sich in der Regel um sog. Bayes-Netze [Kruse et al. 91], bei denen die Attribute als bedingte Wahrscheinlichkeiten im Sinne des Satzes von Bayes modelliert werden, was insbesondere eine entsprechende Normierung der Wahrscheinlichkeiten und damit der Attributwerte bedeutet. Anhand entsprechender Propagationsmechanismen lassen sich die Wahrscheinlichkeiten aller Knoten eines Graphen ermitteln. Abgesehen von der Problematik, ein konsistentes, lösbares Bayes-Netz zu konstruieren, ist die Anwendbarkeit des Bayes’schen Ansatzes auf vielen Gebieten nicht unumstritten [Kosko 92]. Vor allem aber der notwendige Formalismus spricht gegen einen Einsatz in Echtzeit-Anwendungen.

Im Gegensatz dazu fassen die energiebasierten Methoden den Graphen als ein energiespeicherndes Element auf, dessen Energie es mittels geeigneter Relaxationsverfahren zu minimieren oder zu maximieren gilt. Dabei wird jedem Knoten ein Zustand oder Potential

zugeordnet und aus diesen Potentialen und den Kantengewichten als Verbindungen zu anderen Knoten eine Energie definiert. Neben verschiedenen diskreten und kontinuierlichen Relaxationsverfahren [Kruse et. al. 94] zählen u. a. die sog. “Markoff Random Fields“ und die “Hopfield“-Netze zu dieser Kategorie.

Im Rahmen dieser Arbeit ist der Einsatz von Hopfield-Netzen zur Optimierung der Merkmalszuordnung im Zusammenhang mit den Polyedermodellen für Fahrzeuge (Kapitel 6) untersucht worden [von Holt 94]. Weitere Arbeiten, welche sich dieser Methodik zur Merkmalszuordnung bedienen sind [Suganthan et. al. 93] und [Suganthan et. al. 94]. An dieser Stelle soll zunächst die prinzipielle Funktionsweise von Hopfield-Netzen vorgestellt werden; im Anschluß daran wird dann auf die Anwendung im Rahmen der Assoziation von Bildmerkmalen eingegangen.

## 8.1 Hopfield-Netze zur Lösung von Optimierungsproblemen

Hopfield-Netze werden zur Klasse der neuronalen Netze gezählt. Es wurde allerdings bewußt vermieden, den Ausdruck “neuronales Netz“ in die Abschnittsüberschrift einzubeziehen, um nicht die vielfach mit diesem Begriff assoziierten Vorurteile - im positiven wie im negativen Sinn - aufkommen zu lassen. Gemeinhin werden unter dem Begriff der neuronalen Netze meist die sog. Backpropagation-Netzwerke [Kosko 92], [Kruse et. al. 94] verstanden, die im Rahmen der Entdeckung neuer Lernverfahren in den 80er Jahren in Kombination mit den Fortschritten in der Rechnertechnologie sowohl in der Wissenschaft als auch in der Industrie viel Beachtung fanden. Der Umstand, daß diese künstlichen neuronalen Netze in Anlehnung an biologische Nervenzellen konstruiert werden und ihr Einsatz als Klassifikatoren auf einem vorausgehenden Training beruht, ließ bei vielen nicht tiefer mit der Materie Vertrauten den Gedanken an eine Nachbildung des menschlichen Gehirns durch entsprechend trainierte neuronale Netze aufkommen. Mittlerweile ist aber in diesem Bereich eine gewisse Ernüchterung eingetreten, da sich die Aufgaben bei neuronalen Netzen von der Findung eines sonst erforderlichen Algorithmus hin zur Bestimmung der Netztopologie und der Trainingsmethoden verschieben. So bleibt es letztlich eine Frage der jeweiligen Problemstellung, ob diese effizienter anhand “klassischer“ Methoden oder mithilfe neuronaler Netze gelöst werden sollten.

Als wesentliche Kennzeichen neuronaler Netze sollen folgende Eigenschaften gelten:

- Die Grundfunktionalität der Berechnungsmethoden läßt sich an physiologischen Kenntnissen über biologische Nervenzellen und deren Vernetzung motivieren.
- Es handelt sich um i. a. engvermaschte Netze einfacher nichtlinearer Berechnungselemente, die entsprechend der verwendeten Berechnungsfunktion der Elemente nahezu beliebige Funktionen nachbilden können.
- Wissen wird in neuronalen Netzen nicht in den Berechnungselementen sondern in deren Verbindungen gespeichert.

- Wesentlicher Teil des Entwurfs eines neuronalen Netzes ist der Entwurf der Gesamtstruktur, d. h. welche Neuronen miteinander verbunden sind, z. B. durch Vollvermaschung oder Schichtenbildung.

Die hier eingesetzten Hopfield-Netze zählen zur Klasse der rückgekoppelten neuronalen Netze, wobei die Rückkopplungen in Form ungerichteter Verbindungen zwischen den Neuronen bestehen. Aufgrund dieser symmetrischen Verbindungen können sich die Neuronen sowohl gegenseitig anregen (positive Verbindungsgewichte) als auch gegenseitig hemmen (negative Verbindungsgewichte). Als Eingangsfunktion der Neuronen wird die gewichtete Summe der Eingänge verwendet, die im Fall eines diskreten (binären) Hopfield-Netzes auf eine lineare Schwellwertfunktion (Abb. 8.1(a)) oder im Fall eines kontinuierlichen Hopfield-Netzes auf eine sigmoidenförmige Ausgangsfunktion (Abb. 8.1(b)) wirken.

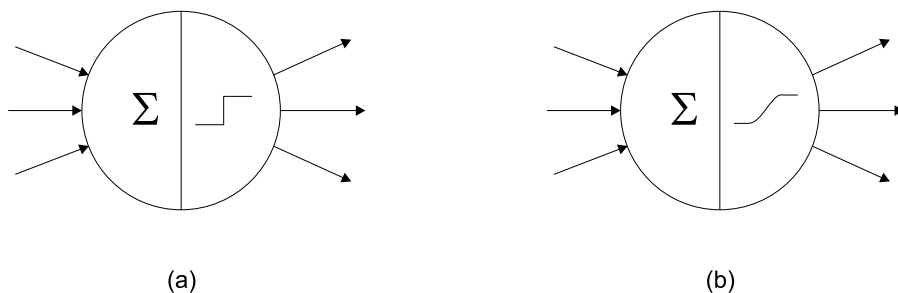


Abbildung 8.1: Neuron eines binären (a) und eines kontinuierlichen (b) Hopfield-Netzes

Im Gegensatz zu vorwärtsgerichteten neuronalen Netzen ist die Anzahl der Berechnungsschritte nicht definiert, sondern die Zustände der Neuronen müssen solange neu berechnet werden, bis das Netz in einen stabilen Ruhezustand konvergiert, in dem sich keine Änderung mehr ergibt. Die Konvergenz des Netzes läßt sich nachweisen, da die Änderung des Aktivierungszustandes eines Neurons nur bei gleichzeitiger Energieabnahme des Gesamtnetzes erfolgen kann. Dabei kann das Netz allerdings leicht in ein lokales anstelle des globalen Energieminimums konvergieren.

Die Struktur von Hopfield-Netzen läßt sich auf zwei Arten motivieren. Einerseits aus physiologischen Überlegungen heraus, daß auch das Gehirn eher einen rückgekoppelten Aufbau mit gegenseitiger Anregung bzw. Hemmung besitzt. Andererseits lassen sich Hopfield-Netze auch als vereinfachtes Modell des Verhaltens sog. Spingläser deuten, d. h. von Materialien deren Atome sich wie Dipolmagnete verhalten. Die Neuronen entsprechen den Dipolmagneten und ihr Aktivierungszustand der Ausrichtung (Spin) im Magnetfeld. Aufgrund der Verknüpfungsstruktur der Neuronen, welche die magnetischen Wechselwirkungen beschreibt, nehmen diese nach einiger Zeit einen Zustand minimaler Energie ein, der übertragen auf die Ausrichtung der Dipolmagnete auch einem minimalen Energiezustand der Spingläser entspricht. In diesem Verhalten sind Hopfield-Netze ähnlich dem vollverknüpften Ising-Modell [Ising 25], das zunächst bei der theoretischen Betrachtung von magnetischen Materialien verwendet wurde.

### 8.1.1 Berechnungsvorschriften für Hopfield-Netze

Die Berechnungsvorschriften für Hopfield-Netze sollen zunächst für binäre und anschließend für kontinuierliche Netze angegeben werden.

Der Eingangswert des Neurons  $i$  eines binären Hopfield-Netzes ergibt sich nach Abb. 8.2 zu

$$H_i = \sum_{j \neq i} t_{ij} x_j + I_i$$

wobei  $x_j$  den Ausgangswert des Neurons  $j$ ,  $t_{ij}$  das Gewicht der Verbindung zwischen den Neuronen  $j$  und  $i$  und  $I_i$  den externen Eingangswert des Neurons  $i$  darstellt.

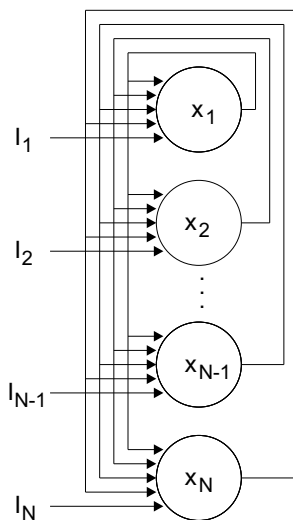


Abbildung 8.2: Rückgekoppelte Struktur eines Hopfield-Netzes

Der Ausgangswert des Neurons  $i$  ergibt sich dann zu

$$x_i = V_i^0 \text{ wenn } H_i < U_i \quad (8.1)$$

$$x_i = V_i^1 \text{ wenn } H_i \geq U_i \quad (8.2)$$

Die binären Werte  $V_i^0$  bzw.  $V_i^1$  können darin z. B. 0 oder 1 oder auch  $-1$  und  $+1$  sein.

Die Neuberechnung aller Neuronenzustände geschieht dabei seriell, wobei das jeweilige Neuron in stochastischer Art zufällig ausgewählt wird.

Durch die Wahl der Verbindungsgewichte, der Neuronenanzustände und der Neuroneneingangswerte konvergiert das Netz in ein lokales Minimum der Energie, die mit der Ljapunov Funktion

$$E = -\frac{1}{2} \sum_{i \neq j} \sum t_{ij} x_i x_j - \sum_i I_i x_i + \sum_i U_i x_i$$

beschrieben wird. Die mit der Zustandsänderung eines Neurons verbundene Änderung der Energie beträgt dann

$$\Delta E = - \left( \sum_{j \neq i} t_{ij} x_j + I_i - U_i \right) \Delta x_i$$



Da gemäß Gl. 8.1.1 das Produkt aus  $\Delta x_i$  und dem Klammerausdruck stets positiv ist, nimmt die Energie bei jeder Zustandsänderung ab, d. h. die Zustandsänderungen der Neuronen sind von Hopfield so definiert worden, daß eine solche nur dann erfolgt, wenn damit eine Energieabnahme verbunden ist.

In ähnlicher Weise wurde von Hopfield ein Modell vorgeschlagen, bei dem die Neuronen kontinuierliche Aktivierungszustände

$$x_i = g_i(u_i) \text{ mit } V_i^0 \leq x_i \leq V_i^1$$

annehmen können, wobei  $g_i(u_i)$  eine kontinuierliche, monoton steigende Funktion des Eingangswertes ist. Typischerweise werden dafür sigmoidenförmige Funktionen, bei denen  $V_i^0$  und  $V_i^1$  die Asymptoten bilden, gewählt, z. B.

$$g_i(u_i) = \frac{1}{1 + \exp\left(\frac{-u_i}{T}\right)} \quad \text{oder} \quad g_i(u_i) = -\frac{1}{2} \tanh\left(\frac{u_i}{T}\right)$$

Der Parameter  $T$  bestimmt dabei, ob der Verlauf der Sigmoiden eher abrupt oder eher sanft ist. Er kann auch in Analogie zur Temperatur gesehen werden und wird z. B. bei den sog. Boltzmann-Maschinen oder dem Simulated Annealing von großen zu kleinen Werten hin variiert mit dem Ziel, möglichst eine Konvergenz in das globale Minimum der Energiefunktion zu erreichen.

Aufgrund der asynchronen Berechnungsvorschrift ergibt sich eine Phasenverschiebung zwischen den Eingangswerten der Neuronen und den Ausgangswerten der mit ihnen verbundenen Neuronen. Dieses Verhalten läßt sich vergleichen mit dem Aufladen eines Kondensators  $C_i$ , der dann über einen Widerstand  $R_i$  entladen wird, was sich in Anlehnung an elektrische Schaltkreise ausdrücken läßt

$$C_i \left( \frac{du_i}{dt} \right) = \sum_j t_{ij} x_j - \frac{u_i}{R_i} + I_i$$

In diesem Sinn können die Eingangswerte der Neuronen als elektrische Spannungen interpretiert werden. Die damit verbundene Energiefunktion ist durch

$$E = -\frac{1}{2} \sum_{ij} \sum_j t_{ij} x_i x_j + \sum_i \left( \frac{1}{R_i} \int_0^{x_i} g_i^{-1}(x) dx \right) + \sum_i I_i V_i$$

gegeben. Auch hierfür läßt sich der Nachweis der Konvergenz in lokale Energieminima erbringen [Hopfield 82].

### 8.1.2 Einsatzgebiete von Hopfield-Netzen

Hopfield-Netze können einerseits als autoassoziative Speicher zum Einsatz kommen, indem anhand bestimmter Lernregeln (z. B. Hebb-Regel [Kruse et. al. 94]) die Verbindungsgewichte entsprechend vorbesetzt werden. Wenn anschließend die Neuronen mit einem bestimmten Muster besetzt werden, kann festgestellt werden, in welches lokale Minimum das Netz

konvergiert und damit welchem der im Netz gespeicherten Muster das angelegte Muster ähnlich kommt (wegen der lokalen Minima nicht zwangsläufig am ähnlichsten).

Daneben haben sich als weiteres Einsatzgebiet kombinatorische Optimierungsaufgaben herausgebildet, wie etwa das bekannte Problem des Handlungsreisenden. Bei diesen Problemen gilt es, die Randbedingungen des Optimierungsproblems in die Energiefunktion aufzunehmen. Auch bei dieser Aufgabe stellt sich das Problem der lokalen Minima erneut, da abhängig vom Problem entschieden werden muß, ob eine erreichte suboptimale Lösung des Problems als akzeptabel gelten kann oder auf weitergehende Verfahren wie etwa die Boltzmann-Maschine übergegangen werden muß.

## 8.2 Hopfield-Netze im Rahmen der relationalen Merkmalszuordnung

Welche Erwartung wird nun mit der relationalen Zuordnung verbunden? Es wird erwartet, daß über die Einbeziehung der Gestalt von Objekten die Merkmalszuordnung stabilisiert wird und auf diese Weise weniger widersprüchliche Hypothesen aufgestellt und verfolgt werden. Einer der Grundgedanken hinter der relationalen Zuordnung wird darin gesehen, daß signifikanteren und ausgeprägteren Merkmalen in der Zuordnungsphase ein höheres Gewicht beigemessen wird, während die weniger signifikanten bzw. ausgeprägten Merkmale sich in deren Muster einfügen oder aber die Inkonsistenz des Musters bestätigen. Unter Signifikanz soll hier die Bedeutung eines Merkmals für das Formmodell im Allgemeinen sowie für die Diskriminierbarkeit eines konkreten Formmodells im Besonderen verstanden werden. Im Gegensatz dazu soll der Begriff Ausprägung hier die Qualität der Messung bezeichnen, derzufolge beispielsweise eine lange, unterbrechungsfreie Kontur mit großem Gradienten (SNR) eine hohe Ausprägung besitzt. Da wir hier nur Konturen betrachten wollen, wird jetzt auch die Bewertung der Bildmessungen im letzten Kapitel deutlich. Die Abstands- bzw. Zugehörigkeitsmaße in Verbindung mit der Gewichtung ergeben gerade diese Ausprägung. Über den zeitlichen Gewichtungsfaktor erfolgt eine Kopplung der Wertung einzelner Merkmale über die Zeit, so daß nicht nur die aufgrund der momentanen Ausprägung maßgebenden Merkmale, sondern auch die zeitlich konsistentesten stärker gewichtet werden.

Analog zu dem für die Merkmale als Knoten des Zuordnungsgraphen Gesagten werden auch die Relationen als Kanten des Graphen durch ihre Signifikanz und ihre Ausprägung gekennzeichnet. Die Signifikanz ist wiederum ein aus dem Modell abgeleitetes Maß für die Aussagekraft bzw. die Diskriminierungseigenschaft einer Relation, während die Ausprägung sich aus der Relation zweier gemessener Merkmale zueinander ergibt.

Die Signifikanz eines Knotens wird also bezogen auf ein Modell-Merkmal bestimmt und ist demnach für alle Daten-Merkmale gleich, die diesem Modell-Merkmal zugeordnet werden können. Ebenso ist die Signifikanz für eine Kante aus der Modellrelation abgeleitet und also für alle Relationen zwischen den Daten-Merkmalen zweier in Relation miteinander stehender Modell-Merkmale gleich. Die Ausprägung bezeichnet hingegen im Fall der Knoten die individuelle Bewertung der Güte eines jeden Daten-Merkmals bezogen auf

die Modellerwartung. Entsprechend bezeichnet die Ausprägung einer Relation die Bewertung der Kompatibilität zweier Daten-Merkmale, die zu miteinander in Relation stehenden Modell-Merkmalen gehören.

Voraussetzung zur relationalen Zuordnung von Merkmalen ist die Möglichkeit einer Abbildung der Merkmale und deren Relationen zueinander auf einen Attribuierten Relationalen Graph (ARG). D. h. die Formrepräsentationen der Objekte müssen diese relationalen Angaben beinhalten. Da es sich bei den Formmodellen um Repräsentationen in 3D handelt, sind auch die enthaltenen Relationen zunächst in 3D gegeben. Für die Zuordnung von Bildmerkmalen in 2D müssen darum aus den Relationen in 3D solche in 2D abgeleitet werden.

Hierzu gibt es folgende Möglichkeiten:

- Die Relationen können unter bestimmten Bedingungen direkt vom 3D- in den 2D-Raum übernommen werden: Die Parallelität zweier Kanten im 3D-Raum bleibt unter perspektivischer Abbildung auch im 2D-Raum in weiten Grenzen erhalten, sofern diese weitestgehend senkrecht zur Blickrichtung verlaufen.
- Aus den Relationen in 3D können für einen Zeitpunkt Relationen in 2D abgeleitet werden, z. B. in Form von (Aspekt-)Schnittpunkten von Kanten im Bild.

Die in dieser Arbeit verwendeten Relationen zählen sämtlich zur erstgenannten Kategorie. Als Merkmale dienen lediglich Kanten in Form von Geraden. Die entsprechenden Relationen sind bei der Repräsentation der Formmodelle hinterlegt. Nachfolgend werden die unären Abstandsmaße für die Knoten und die binären Abstandsmaße für die Kanten des ARG näher erläutert.

### 8.2.1 Unäre Abstandsmaße

Bei den unären Abstandsmaßen handelt es sich um die Bewertung der gemessenen Merkmale, wie sie im vorhergehenden Kapitel 7 über die Bildvermessung bereits behandelt wurden. Für die konkrete Beschreibung der Bewertung sei auf dieses Kapitel verwiesen. Hier sei nur der Vollständigkeit halber die Liste der Bewertungskriterien nochmals zusammengestellt:

1. Signal-Rausch-Verhältnis (SNR) abgeleitet aus dem Grauwertgradienten
2. Mittelwert und Varianz des Grauwertverlaufs vor bzw. hinter der Kontur
3. Geometrischer Abstand der Meßpunkte von der Erwartung
4. Photometrischer Abstand der Meßpunkte von der Erwartung

## 8.2.2 Binäre Abstandsmaße

Bei Verwendung der hauptsächlich konturbasierten Formmodelle bieten sich in der Hauptsache zwei Relationen an:

- Die Vertex-Relation als Schnittpunkt/Treffpunkt mindestens dreier Kanten (bei 3D-Formmodellen). Die Endpunkte zweier Daten-Kanten sollten möglichst benachbart sein, um als Kandidaten für die sich in einem Vertex treffenden Modell-Kanten in Frage zu kommen. Problematisch an dieser Relation ist die schlechte Konditionierung bei vielen realen Objekten, da z. B. bei modernen Fahrzeugen die Kanten und damit auch die Vertexpunkte abgerundet und in Verbindung mit den reflektierenden Eigenschaften der metallischen Oberflächen nicht vermeßbar sind, sondern nur als virtueller Schnittpunkt der Kanten existieren. Ebenso problematisch ist die Aspektabhängigkeit der Kantenschnittpunkte in 2D, etwa in Form schleifender Schnitte. Diese Gesichtspunkte sollten in die Signifikanz der Vertex-Relation für die betreffenden Objekte eingehen.
- Die Parallel-Relation bewertet die aus dem Modell vorgegebene Parallelität zweier Modell-Kanten, die zu einer Fläche gehören. Bei ihr handelt es sich um ein sogenanntes “nicht zufälliges“ (engl. “non accidental“) Ereignis, d. h. eines, das nur mit sehr geringer Wahrscheinlichkeit zufällig zwischen zwei Daten-Kanten im Bild entsteht, sondern i. a. aufgrund eines systematischen relationalen Zusammenhangs beider Kanten. Entsprechend signifikant ist diese Relation zu bewerten. Die Aspektabhängigkeit ist bei den üblichen Abbildungsbedingungen wenig ausgeprägt, im Einzelfall jedoch zu prüfen.

Im folgenden werden für beide Kategorien von Relationen die notwendigen Schritte zu ihrer Berechnung weiter ausgeführt.

### Vertex-Relation

Ein Vertex bildet bei 3D-Formmodellen stets den Treffpunkt von drei oder mehr Bildkanten. Da jede Bildkante einem Knoten des Graphen entspricht, kann eine solche Relation nicht direkt in den relationalen Graphen umgesetzt werden, da eine Kante des Graphen stets nur zwei Knoten desselben miteinander verbindet. Stattdessen werden, wie in Abb. 8.3 gezeigt, die sich am Vertex treffenden direkt benachbarten, zu einer gemeinsamen Fläche gehörenden Kanten paarweise zueinander in Relation gesetzt. Um nun aus der Lage der Modellkanten(-Kandidaten) zueinander deren Kompatibilitätskoeffizienten  $k_{V,a/b}$  zu bestimmen, wird zunächst der - in Abhängigkeit der Kantenlage - tatsächliche oder virtuelle Schnittpunkt beider Kanten bestimmt. Mit den in Abb. 8.4 gegebenen Bezeichnungen ermittelt man dann das Verhältnis der Strecke der den Vertex bildenden Kantenendpunkte zur Gesamtlänge der Strecke:

$$t_{a/b} = \frac{s_{a/b}}{l_{a/b}}$$

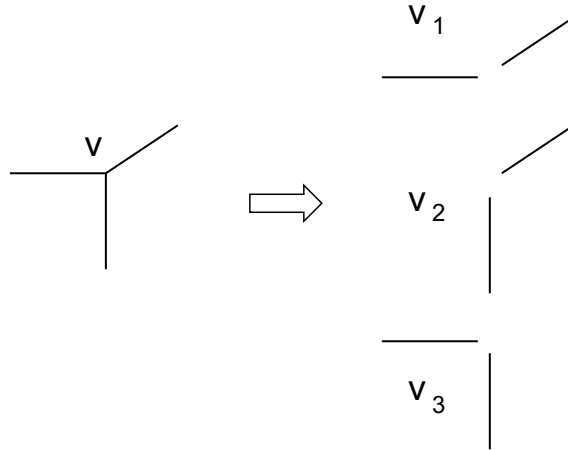


Abbildung 8.3: Umsetzung einer Vertex-Relation in Kanten des ARG

Hieraus wird dann das zwischen  $+1$  für höchste Kompatibilität und  $-1$  für höchste Inkompatibilität beider Modellkanten liegende  $k_{V,a/b}$  bestimmt:

$$k_V = \begin{cases} 0 & \text{falls } |t| > 1/2 \\ 1 - 2|t| & \text{falls } |t| \leq 1/2 \end{cases}$$

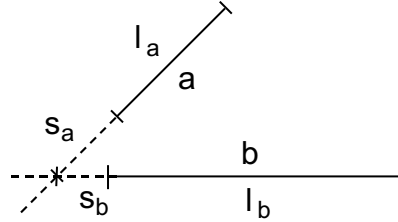


Abbildung 8.4: Notation einer Vertex-Relation

### Parallel-Relation

Der Kompatibilitätskoeffizient zweier Kanten in bezug auf ihre Parallelität wird in Abhängigkeit des zwischen beiden Kanten liegenden Winkels bestimmt. Mit den in Abb. 8.5 gegebenen Bezeichnern ergibt sich der wiederum zwischen  $+1$  für höchste Kompatibilität und  $-1$  für höchste Inkompatibilität liegende Kompatibilitätskoeffizient  $k_P$  zu:

$$k_P = \begin{cases} -1 & \text{falls } |\alpha| > \alpha_g \\ 2(1 - \frac{\sin \alpha}{\sin \alpha_g}) - 1 & \text{falls } |\alpha| \leq \alpha_g \end{cases}$$

Durch Wahl des (Grenz-)Winkels  $\alpha_g$  wird mithin der Toleranzbereich der Relation festgelegt. Als sinnvolle Werte haben sich Winkel im Bereich  $\alpha_g = 22.5^\circ - 45^\circ$  erwiesen.

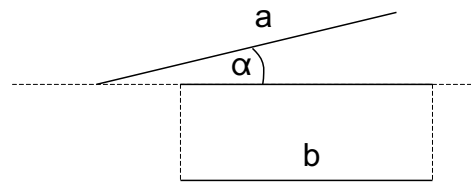


Abbildung 8.5: Notation einer Parallel-Relation

### 8.2.3 Bildung eines Attributierten Relationalen Graphen am Beispiel

In diesem Abschnitt soll anhand eines schematischen Beispiels die Abbildung der gemessenen Merkmalskandidaten eines Objekts auf dessen korrespondierenden ARG und das Vorgehen bei der eigentlichen Zuordnungsoptimierung veranschaulicht werden.

In Abb. 8.6 ist zunächst die Konstruktion des ARG aus einem Formmodell unter den zu einem Zeitpunkt gegebenen Aspektbedingungen zu sehen. Ausgehend vom Formmodell bilden die sichtbaren Kanten als Modell-Merkmale die Merkmals-Knoten. Die im Formmodell hinterlegten relationalen Beziehungen zwischen den Modell-Kanten führen zu den dargestellten Vertex- bzw. Parallel-Relationen. Der ARG resultiert dann aus der Zusammenschau der Merkmals-Knoten und der Relationen. Damit haben wir jetzt das Basisgerüst konstruiert, auf dem das Verfahren zur optimierten Zuordnung aufsetzen kann.

Im nächsten Schritt erfolgt - wie in Abb. 8.7 dargestellt - die Abbildung der aus der Messung resultierenden Merkmals-Kandidaten auf den ARG. Jedem Merkmals-Knoten des ARG werden "Kandidaten"-Subknoten entsprechend der Anzahl gemessener Daten-Merkmale zugeordnet. Jeder dieser "Kandidaten"-Subknoten besitzt zum einen ein unäres Gewicht, welches seine Kompatibilität mit dem korrespondierenden Merkmals-Knoten ausdrückt (wie gut paßt die gemessene Kante zur Erwartung?). Zum anderen werden für sämtliche Relationen, über die ein Merkmals-Knoten mit anderen Merkmals-Knoten verbunden ist, die Kompatibilitätsmaße zwischen allen "Kandidaten"-Subknoten beider Merkmals-Knoten berechnet. Diese bilden die Matrix der binären Verbindungsgewichte zwischen den Merkmals-Knoten. In der Abbildung ist z. B. zu erkennen, daß die über die Vertex-Relation *b* verbundenen Merkmals-Knoten 1 und 4, welche 2 bzw. 3 "Kandidaten"-Subknoten besitzen, in der Folge mit einer  $2 \times 3$  Gewichtsmatrix miteinander gekoppelt sind. Wie der Abb. 8.7 ebenfalls zu entnehmen ist, führt das Nichtvorhandensein von Daten-Merkmalen für ein Modellmerkmal dazu, daß der entsprechende Merkmals-Knoten und die zu diesem führenden Verbindungen aus dem Graphen entfernt werden können.

Als nächstes folgt dann der Optimierungslauf zur Zuordnung maximal eines besten "Kandidaten"-Subknoten zu einem Merkmals-Knoten im Sinne des ARG. In diesem Zusammenhang ist wichtig, die hierarchische Struktur des Netzes zu betonen. In einem nicht hierarchischen Hopfield-Netz wären alle Knoten gleichberechtigt. In diesem Fall würden als Knoten des Hopfield-Netzes lediglich die "Kandidaten"-Subknoten auftreten. Die Merkmals-Knoten des ARG wären dann überflüssig. Nach dem Optimierungslauf ergäbe sich eine aus den Gewichten des ARG resultierende "optimale" Zuordnung, d. h. Aktivierung bestimmter Kno-

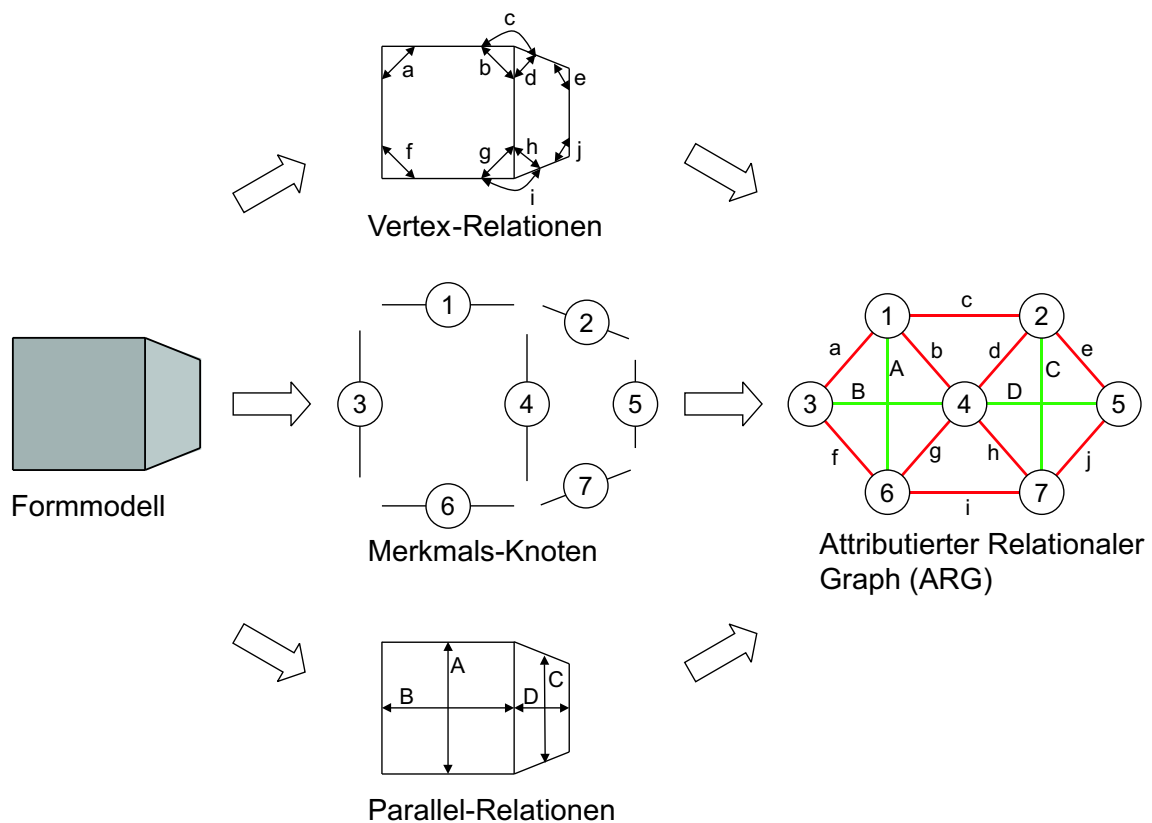


Abbildung 8.6: Bildung eines ARG anhand der Topologie des Formmodell

ten. Da aber je Merkmals-Knoten nur maximal ein aktivierter Knoten zulässig ist wegen der Eindeutigkeit der Zuordnung, müßten die “Kandidaten“-Subknoten eines Merkmals-Knotens so untereinander verbunden sein, daß diese Randbedingung gewährleistet wäre. Diese Lösung ist hier nicht gewählt worden. Vielmehr treten als Knoten des Hopfield-Netzes nur die Merkmals-Knoten als “Hauptknoten“ auf. Als positiver Nebeneffekt wird dadurch die Anzahl der Knoten als Neuronen des Hopfield-Netzes deutlich reduziert und damit in der Folge auch die Anzahl notwendiger Iterationen bis zum Erreichen eines Endzustands. Der Ablauf eines solchen Optimierungslaufs soll anhand der in Abb. 8.8 dargestellten Iterationen verdeutlicht werden.

### 1. Iteration

Zu Beginn wird für jeden Knoten zufällig eine Zuordnung aus der Menge seiner “Kandidaten“-Subknoten als aktiver Subknoten ausgewählt. Damit der ausgewählte Subknoten wirklich aktiviert wird, muß seine Eingangsaktivierung oberhalb der Aktivierungsschwelle liegen. Die Aktivierungsschwelle ergibt sich bei dieser initialen Zuordnung allein auf der Basis der unären Zuordnungsgüte der Subknoten zum Merkmalsknoten (Hauptknoten). Wenn die Aktivierungsschwelle nicht überschritten wird, erhält dieser Hauptknoten keinen initial aktiven Subknoten. Im Beispiel besitzen die Hauptknoten 1, 3, 4 und 7 aktivierte Subknoten; bei den Hauptknoten 5 und 6 wurde hingegen kein Subknoten ausgewählt.

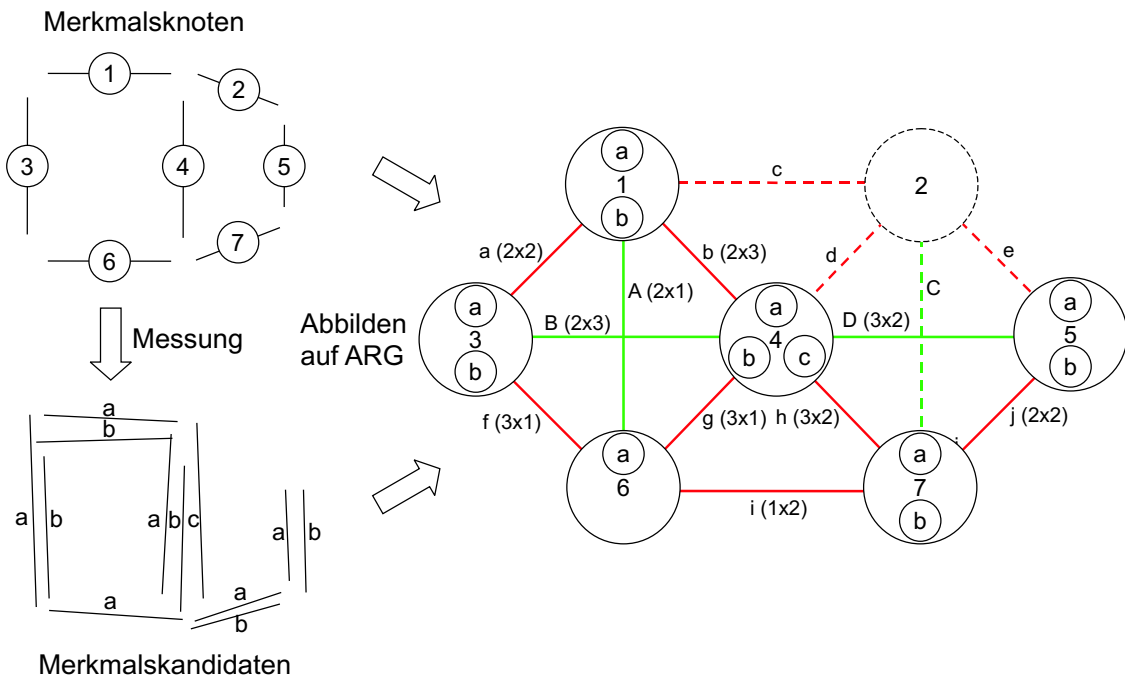


Abbildung 8.7: Abbilden der gemessenen Daten-Merkmale auf den ARG

## 2. Iteration

Unter den Haupt-(Merkmals-)Knoten wird ein Knoten ausgewählt (farblich hervorgehoben). Für diesen Hauptknoten wird auf der Basis der momentan aktiven Subknoten der Nachbarknoten und ihrer binären Verbindungsgewichte zu den eigenen Subknoten sowie den unären Gewichten der eigenen Subknoten ein neuer aktiver Subknoten bestimmt. Dazu wird für jeden Subknoten die Summe seiner Eingangsaktivierung berechnet. Derjenige Subknoten mit der höchsten Eingangsaktivierung wird als neuer aktiver Subknoten ausgewählt, sofern die Eingangsaktivierung oberhalb der Aktivierungsschwelle liegt. Im Beispiel wurde die Zuordnung für den Hauptknoten 3 aktualisiert. Hierbei wurde statt des bislang aktiven Subknoten b der Subknoten a aktiviert.

## 3. Iteration

Analog zu Iterationsschritt 2 wird wiederum ein Hauptknoten (4) ausgewählt und die Zuordnung durch Berechnung aller Eingangsgewichte von den Nachbarknoten bestimmt. Auch in diesem Fall erfolgt ein Wechsel des aktiven Subknotens (c zu b).

## 4. Iteration

Für den Hauptknoten 6, der nur über einen aktivierbaren Subknoten verfügt, ergibt sich aufgrund der aktiven Nachbarknoten, daß nun auch dieser - bislang unterhalb der Aktivierungsschwelle liegende - Subknoten aktiv gesetzt wird.



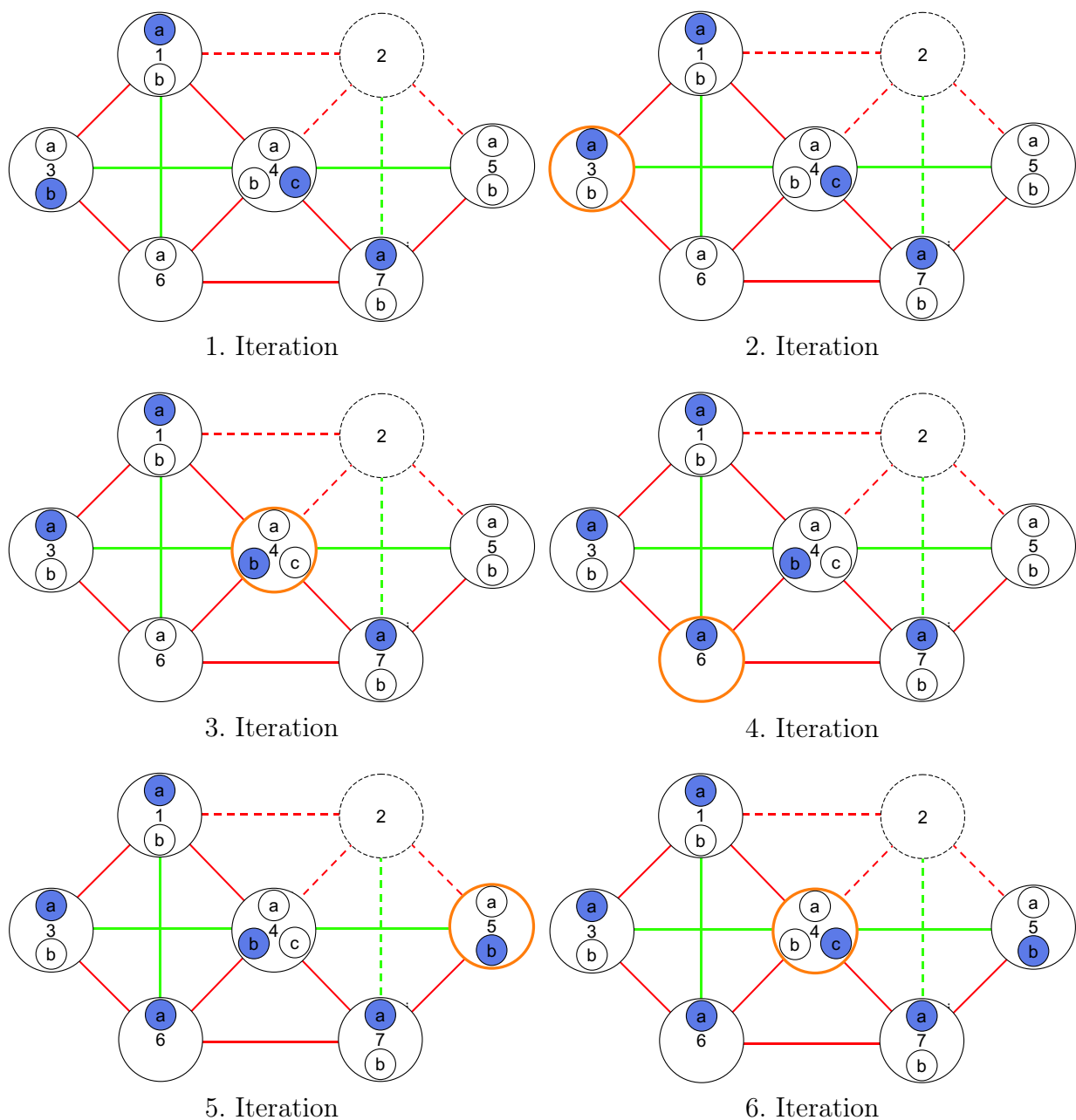


Abbildung 8.8: Beispielhafter Ablauf der Iterationen eines ARG

## 5. Iteration

Ähnlich wie für Hauptknoten 6 in der vorhergehenden Iteration ergibt sich auch für Hauptknoten 5 aufgrund der Konstellation der Nachbarknoten nunmehr ein aktiver Subknoten (b).

## 6. Iteration

In der letzten - hier beispielhaft dargestellten Iteration - wird erneut per Zufallsverfahren der Hauptknoten 4 aktualisiert. Der im 3. Iterationsschritt erfolgte Wechsel

des aktiven Subknotens von Subknoten *c* zu Subknoten *b* wird aufgrund der nunmehr geänderten Zuordnung der Subknoten in den benachbarten Hauptknoten wieder umgekehrt, so daß wiederum Subknoten *c* aktiviert ist.

Das Iterationsverfahren kann entweder nach Ablauf einer bestimmten Zahl von Iterationen oder nach Überschreiten einer Zeitschwelle abgebrochen werden. Das absolute Abbruchkriterium besteht darin, daß alle Hauptknoten in Folge einmal ausgewählt wurden, ohne daß bei einem von ihnen die Zuordnung geändert wurde. In diesem Fall befindet sich das Hopfield-Netz in einem lokalen oder globalen Energieminimum.

### 8.3 Alternative Zuordnungsstrategien

Sollte aufgrund der gegebenen Szenenkomplexität oder der verfügbaren Rechenleistung gar kein Optimierungslauf möglich sein, so kann in gewissen Grenzen dennoch ein Gewinn aus dem relationalen Modell gezogen werden, indem eine der folgenden Heuristiken zur Anwendung gelangt:

1. Es erfolgen nur bei den Merkmalen Meßwertzuordnungen, die mit einer vorzugebenen Zahl anderer Merkmale in Relation stehen, für die überhaupt Kandidaten gefunden wurden. Es werden dann für sich genommen die Merkmale isoliert, d. h. ohne Kompatibilitätstest, zugeordnet. Damit lassen sich Zuordnungen isoliert stehender Merkmale zugunsten von Merkmalgruppierungen unterdrücken.
2. In Erweiterung der soeben beschriebenen Darstellung, wird die Kompatibilität miteinbezogen. Dabei arbeitet man die Merkmale in der Rangfolge ihrer unären Gewichtung ab. D. h., das am stärksten ausgeprägte Merkmal wird als erstes zugeordnet. Bei jedem folgenden Merkmal wird die Kompatibilität zu bereits zugeordneten Merkmalen in der Nachbarschaft, also solchen die eine höhere unäre Gewichtung haben, geprüft. Wenn das binäre Gewicht der Kante zwischen beiden Merkmalen negativ ist, unterbleibt die Zuordnung. Nach Abarbeitung aller Merkmale müssen die Merkmale noch auf Isoliertheit geprüft und ggf. eliminiert werden.



# Kapitel 9

## Multisensorielle Objekterkennung

In den vorangegangenen Kapiteln wurden verschiedene zentrale Gesichtspunkte der Umfeldwahrnehmung unabhängig voneinander betrachtet. In diesem Kapitel zur Multisensoriellen Objekterkennung und dem anschließenden Kapitel zur Szeneninterpretation geht es nun um die Zusammenschau dieser isolierten Aspekte. An dieser Stelle werden zunächst die Aspekte der Vorhersage, Messung, Assoziation und Initialisierung in den Meßmodellen für den Lidar- und den Videosensor sowie die Schätzung von Form- und Bewegungsgrößen mit den Objektmodellen der Szenenrepräsentation zusammengeführt. Zur Einordnung dieser Beiträge zur Objekterkennung wird einleitend die Struktur der Informationsfusion vorgestellt.

### 9.1 Allgemeines Vorgehen

Das Eintreffen der Informationen von den verschiedenen Sensoren ist in der hier verwendeten Sensorkonfiguration und Systemarchitektur asynchron. Darüberhinaus verfügen die Sensoren über unterschiedliche Zykluszeiten und Latenzzeiten zwischen Messung und Meßwertausgabe. Diese Anordnung legt eine ebenfalls asynchrone Verarbeitung der Informationen nahe. Anderfalls müßte eine Prädiktion bzw. Retrodiktion von Meßdaten erfolgen, welche aber erst bei sehr großen Unterschieden in den Zyklus- und/oder Latenzzeiten erforderlich ist. Hier dagegen erfolgt die Objekterkennung stets auf Einzelsensorbasis, unabhängig davon, ob “zufällig” zwei oder mehr Sensorinformationen (nahezu) zeitgleich eintreffen.

### Struktur der Sensordaten-/Informations-Fusion

In Anlehnung an die in Kapitel 3.2.2 gegebene Darstellung des 4D-Ansatzes zur rekursiven Szeneninterpretation ist in Abbildung 9.1 die modulare Struktur der Informationsfusion dargestellt.

Bei Eintreffen neuer Sensorinformationen werden die in der UMFELDMODELL genannten Szenenrepräsentation enthaltenen Objekte durch das Modul UDT auf den Meßzeitpunkt

prädiziert. Das darauf folgende Modul TSF übernimmt die Abbildung in den Merkmalsraum des Sensors. Die Assoziation der vorhergesagten Merkmale mit den vom Sensor tatsächlich gelieferten Daten erfolgt im Modul ASO. In Abhängigkeit des Erfolges der Assoziation werden nun entweder die zugeordneten Merkmale als Meßwerte zum Aktualisieren des Objektzustands im Modul UDM genutzt, oder die nicht mit vorhergesagten Objektmerkmalen assoziierten Meßmerkmale dienen zum Instanzieren neuer Objekthypothesen durch das Modul INS. Am Ende eines Fusionsschrittes werden sämtliche aktualisierten, nicht aktualisierten oder instanziierten Hypothesen einer Plausibilitätsprüfung unterzogen (UDE). Die drei Module TSF, ASO und INS beinhalten die sensorabhängigen Teile der Informationsgewinnung und werden zusammen als Meßmodell des jeweiligen Sensors bezeichnet.

Die zum Sensormodell zählenden sensorspezifisch ausgeprägten Module TSF, ASO und INS führen dazu, daß sich bei Mehrsensoranordnungen eine Struktur ergibt, wie sie in Abbildung 9.2 am Beispiel einer 2-Sensor-Anordnung zu sehen ist. Durch die Zusammenfassung der sensorspezifischen Module zu sogenannten FUSIONSENSOREN tritt die Modularität deutlich zutage, die auch schon im Abschnitt über das verallgemeinerte 4D-Modell dargestellt wurde.

In der weiteren Darstellung werden die wesentlichen Verarbeitungsschritte genauer erläutert, beginnend mit dem rekursiven Teil der Objekterkennung, der die sensorspezifischen Meßmodelle des Video- und des Lidarsensors beinhaltet, sowie dem allgemeinen Block der Form- und Bewegungsschätzung. Anschließend wird auf die wiederum sensorspezifische Initialisierung von Objekten bei deren erstmaliger Detektion eingegangen. Die Verknüpfung

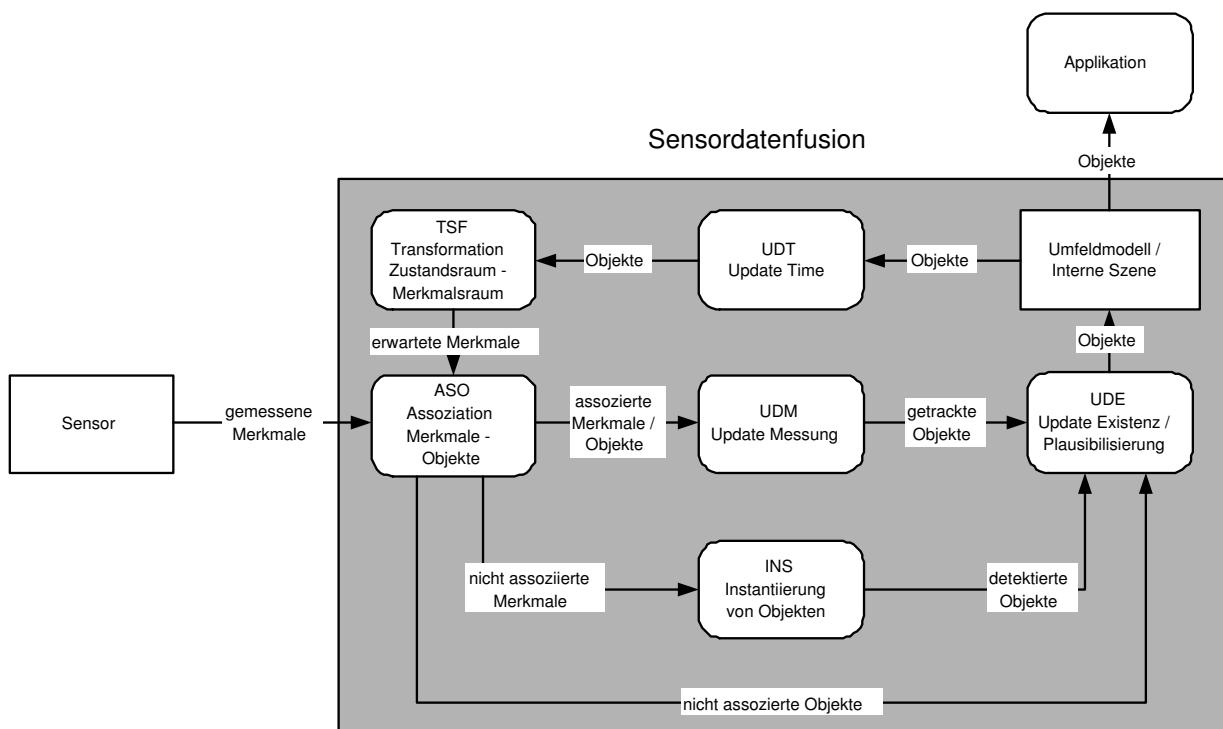


Abbildung 9.1: Module der Informationsfusion für eine 1-Sensor-Anordnung

der einzelnen Funktionsblöcke und die Hypothesenbewertung sind Inhalt des nachfolgenden Kapitels 10 zur Szeneninterpretation.

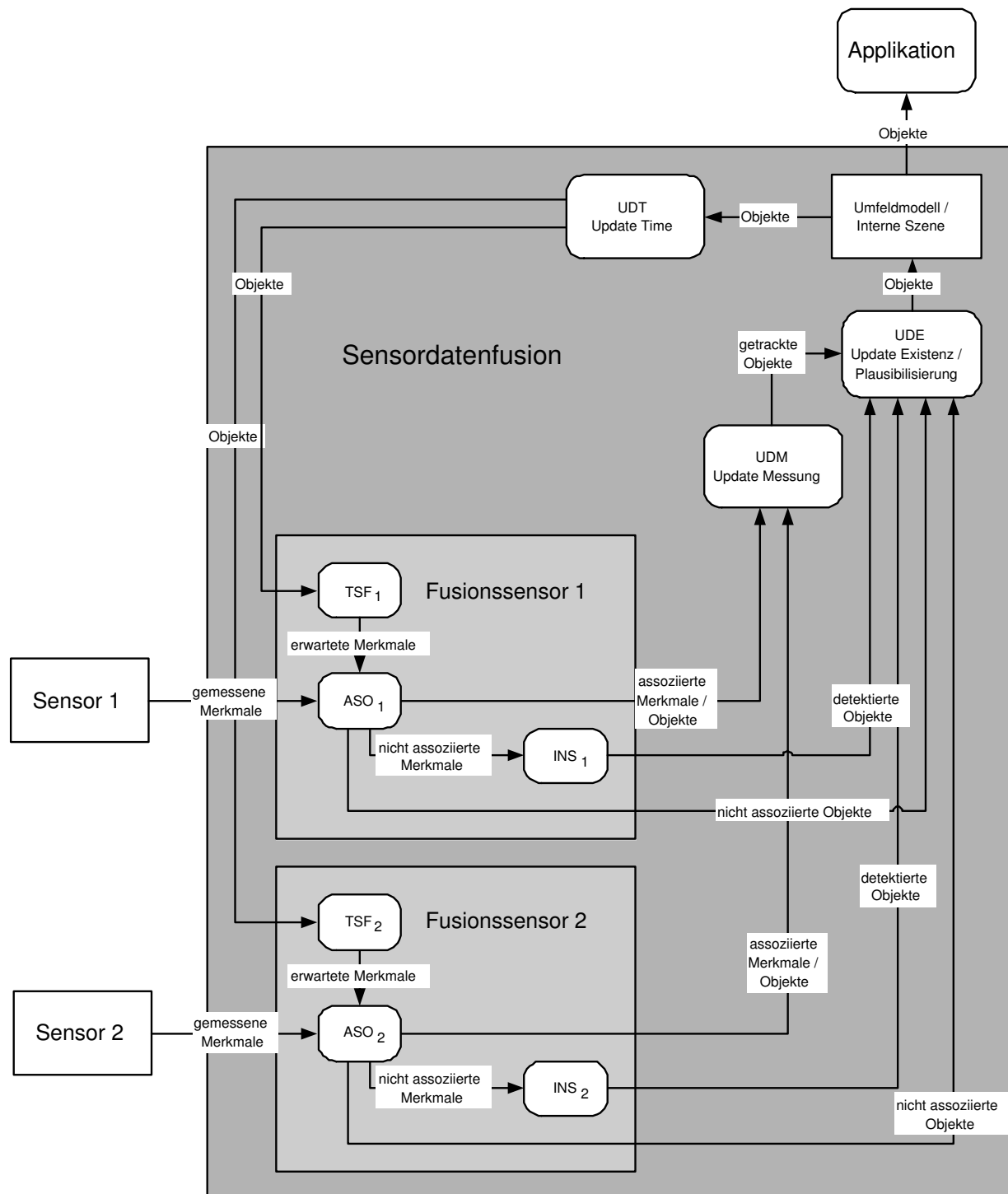


Abbildung 9.2: Module der Informationsfusion für eine 2-Sensor-Anordnung

## 9.2 Meßmodell Video

Das Meßmodell des Videosensors beinhaltet, Abb. 9.1 folgend, die Vorhersage der Bildmerkmale, deren Vermessung im Bild und schließlich die Assoziation der gemessenen Merkmale mit den prädizierten Merkmalen. Hierbei ist vor allem der letzte Schritt der Assoziation mit hohen Entscheidungsunsicherheiten verbunden und als wesentlich für die Qualität der Objekterkennung anzusehen.

### 9.2.1 Vorhersage der Merkmale

Die Vorhersage der Bildmerkmale basiert auf der Prädiktion der Zustände aller in der Szene enthaltenen Objekte zum Zeitpunkt der Messung, d. h. der Bildaufnahme des Video-Chips. Entsprechend der in Kapitel 4 angegebenen Sensoreigenschaften wird die Ausprägung der modellierten Objektmerkmale zum Meßzeitpunkt durch die Transformation der Merkmale aus dem Zustands- in den Merkmalsraum ermittelt. Neben der reinen Transformation sind der Sensorsichtbereich sowie die sich aus Sensorsicht ergebenden gegenseitigen Verdeckungen der Objekte maßgebliche Faktoren in der Vorhersage der Merkmale.

#### Objekte

Die allgemeine Sichtbarkeitsüberprüfung von Objekten ist selbst bei einer polyedrischen Darstellung eine auch heute noch sehr rechenzeitintensive Aufgabe, die in Abhängigkeit der Zahl und der Komplexität der Objekte in der Szene exponentiell ansteigt. Daher werden für den ersten Schritt der Sichtbarkeitsüberprüfung Objekte zunächst durch ihren einhüllenden Quader beschrieben (Abb. ??).

Da in dieser Arbeit darüberhinaus alle Objekte als auf der Fahrbahnebene befindlich angenommen werden, wird das Problem in einem zweiten Schritt auf ein 2-dimensionales Problem in der Grundebene reduziert. Es ergeben sich einhüllende Rechtecke für die Objekte, die zur Sichtbarkeitsüberprüfung herangezogen werden. Dieser Vereinfachung liegt die Annahme zugrunde, daß sich gegenseitig in der Ebene verdeckende Objekte auch in der Höhe mit den derzeit eingesetzten Meßverfahren nicht sicher zu trennen sind, zumal derzeit auch die Fahrbahn als ebene Fläche ohne Vertikalkrümmung angenommen wird. Diese Annahme muß bei Hereinnahme der Fahrbahn-Vertikalkrümmung oder bei Einsatz verfeinerter Meßverfahren natürlich gegebenenfalls wieder revidiert werden. Erste Betrachtungen zur Erkennung verdeckter Objekte unter Nutzung des 4D-Ansatzes hat [Schmid 94] durchgeführt.

Bei der Berechnung der Verdeckungen in der Ebene werden die verwendeten Meßverfahren ebenfalls berücksichtigt: Als sichtbar, vermeßbar und schätzbar werden nur solche Objekte eingestuft, die nicht durch andere verdeckt bzw. teilverdeckt werden. Insbesondere teilverdeckte Objekte und die diese verdeckenden Objekte werden aber gekennzeichnet, da sie durch ihre Merkmale zu einer erschwerten Vermessung und Assoziation der Vordergrund-Objekte führen können. Auf der Basis der so bestimmten Sichtbarkeiten der einhüllenden

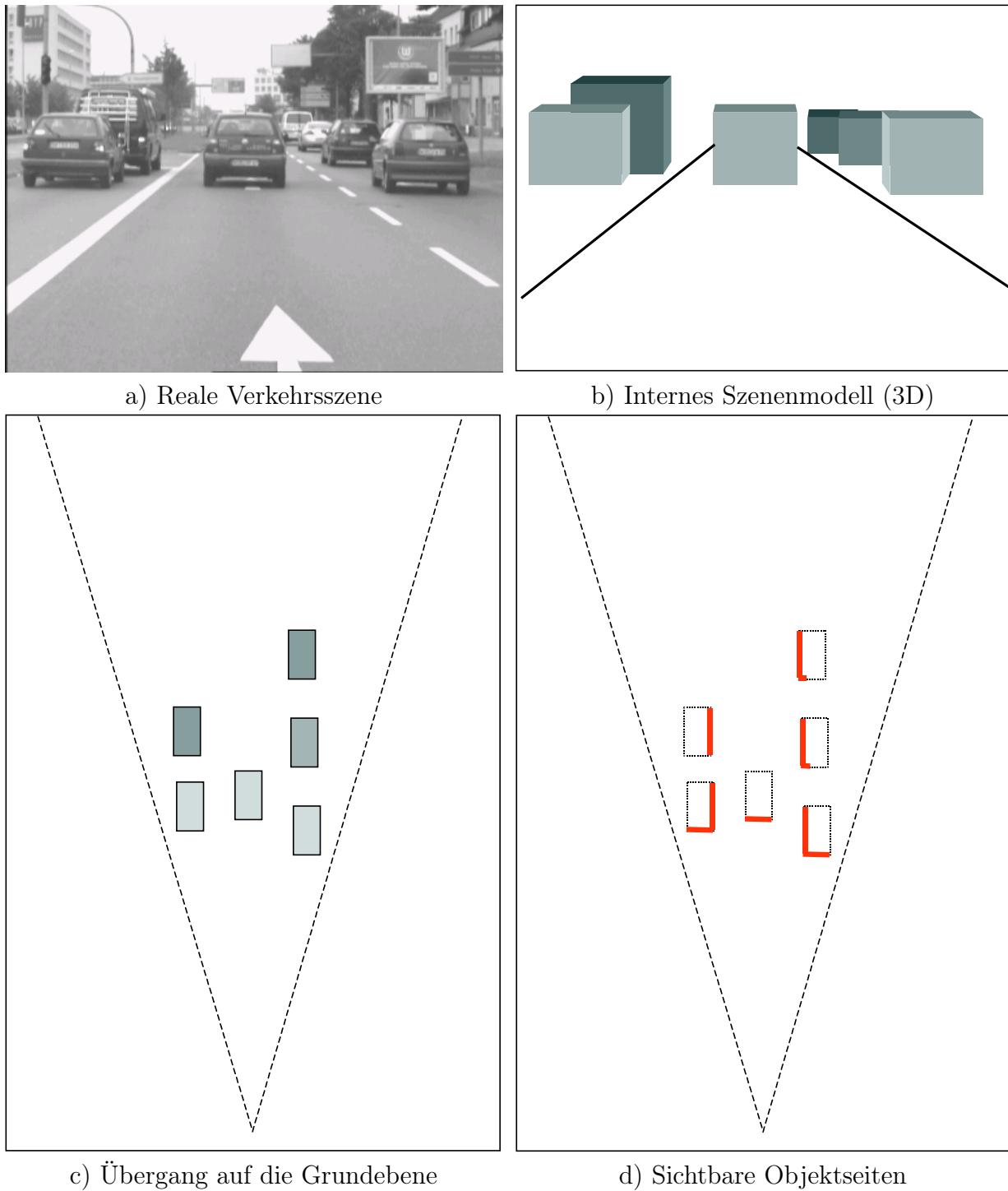


Abbildung 9.3: Reduktion der Verdeckungsanalyse auf ein ebenes Problem

Rechtecke bzw. Quader läßt sich dann die Sichtbarkeit der einzelnen Objektmerkmale bestimmen. Diese Überprüfung ist bei Objekten trivial, die direkt in Quaderform modelliert sind; für Fahrzeuge in der Polyederdarstellung nach Abschnitt 6.2.5 ist die Überprüfung aufgrund der möglichen teilkonkaven Form aufwendiger. Die Überprüfung erfolgt nach Standardverfahren, die aus der Computergrafik bekannt sind.



In Abb. 9.3 ist der Ablauf am Beispiel einer realen Szene erläutert. In Bild 9.3 a) ist die reale Szene zu erkennen, im Bild 9.3 b) daneben die korrespondierende interne Modellvorstellung. Im Bild 9.3 c) ist auf die Projektion in die Grundebene übergegangen worden und Bild 9.3 d) schließlich sind die sichtbaren Kanten der einhüllenden Rechtecke in der Grundebene zu entnehmen.

## Fahrspur

Die Sichtbarkeitsüberprüfung für Fahrspur-Merkmale erfolgt in zwei Schritten:

1. Im ersten Schritt werden die Fahrspur-Merkmale ins Bild transformiert. Die Beschreibung für die Konturvermessung erfolgt analog zu den Objekten durch eine Reihe von Konturpunkten. Bei der Annahme einer ebenen Fahrbahn kann es in Abwesenheit von Objekten nicht zu (Eigen-)Verdeckungen kommen, so daß alle Punkte sichtbar sind, welche im Sichtbereich des Sensors liegen. Im Fall vertikaler Fahrbahnkrümmungen sind hingegen auch Eigenverdeckungen möglich.
2. Im zweiten Schritt erfolgt dann die Berücksichtigung der Objekte. Objekte werden als bis auf die Fahrbahn reichend modelliert, so daß sie sämtliche hinter ihnen liegenden Markierungen verdecken.

### 9.2.2 Vermessung der Merkmale

Im Anschluß an die Vorhersage der Merkmale im Bild werden diese vermessen. Dabei gelangen im Rahmen dieser Arbeit die in Kapitel 7 dargestellten konturbasierten Methoden zur Vermessung von Konturpunkten und deren Aggregation zu Konturen zum Einsatz.

## Objekte

Die Formmodelle der Objekte nach Kapitel 6.2 beinhalten sämtliche Angaben zu ihren vermeßbaren Konturlinien, Punkten auf diesen Konturlinien und deren geometrische und photometrische Attribute. Anhand dieser Angaben werden die Bildmessungen parametrisiert, die Bildmessung durchgeführt und die Konturpunkte zu Konturen aggregiert. Als Ergebnis der Messung erhält man für jede sichtbare Objektkontur eine Liste möglicher Kantenkandidaten. Die Zuordnung eines Kantenkandidaten aus dieser Liste erfolgt im Rahmen der Assoziation (Abschnitt 9.2.3).

## Fahrspur

Im Gegensatz zur Vermessung von Objektkonturen, bei denen in der Trackingphase das Wiederauffinden bereits gesehener ortsfester Merkmale im Mittelpunkt steht, liegt es bei der Vermessung der Fahrspur in der Natur der Sache, daß aufgrund der eigenen Fortbewegung eine ständige "Umwälzung" der Merkmale erfolgt. D. h. es werden diejenigen

Merkmale verloren, über die man “hinweggefahren“ ist, während in den entfernten Vorausschaubereich neue, bislang nicht vermessene Merkmale eintreten. Die Merkmalsvermessung ist damit eine ständige Mischung aus Tracking bereits gesehener Merkmale und der “Detektion“ neuer Merkmale.

Ein weiterer Unterschied besteht darin, daß auf markierten Wegen die Merkmale nicht einfache Konturkanten sind, sondern daß es sich in aller Regel um Konturlinien mit einem charakteristischen Dunkel-Hell-Dunkel-Übergang handelt (unter normalen Witterungs- und Beleuchtungsbedingungen). Dieser Charakteristik wird Rechnung getragen, indem als aggregierter Merkmalstyp ein Linienelement, d. h. eben jene Kombination eines Dunkel-Hell- und eines Hell-Dunkel-Übergangs gebildet wird. Basierend auf diesen Merkmalen werden eigene Gütemaße für Linienmeßwerte definiert. Zur reinen Bildvermessung werden auch für die Fahrspur die in Kapitel 7.1 erläuterten Verfahren genutzt. Eine an die Vermessung anschließende explizite Linienaggregation erfolgt jedoch nicht. Die Fahrspurmarkierungen sind i. a. so ausgeprägt, daß die Zuordnung der Meßwerte auf der Basis der Einzelmeßwerte erfolgen kann, mit einer nachgeschalteten Plausibilisierung.

### 9.2.3 Assoziation der Merkmale

Auf die Kernproblematik Datenassoziation bei umfelderfassenden Sensoren - insbesondere in der Bildverarbeitung - ist bereits wiederholt hingewiesen worden.

Für Objekte ist in Kapitel 8 mit den Hopfield-Netzen auch bereits ein allgemeines Verfahren zur Bearbeitung dieser Problematik beschrieben worden. Die notwendigen Merkmalsattribute und -relationen sind in den Formmodellen der Objekte repräsentiert, die notwendigen Maße werden im Rahmen der Bildvermessung bestimmt. Bei Verwendung des polyedrischen Fahrzeugmodells mit seiner Vielzahl an Merkmalen und deren Relationen ist der Einsatz eines dedizierten Zuordnungsverfahrens wie des Hopfield-Netzes auch in jedem Fall zu empfehlen. Bei dem in den praktischen Versuchen in Kapitel 12 verwendeten einfachen Quadermodell ist der Einsatz jedoch unnötig. Das für diesen “einfachen“ Fall entwickelte Zuordnungsverfahren wird im folgenden Abschnitt über die Assoziation von Objektmerkmalen beschrieben. Die Zuordnung der gemessenen Fahrspurmerkmale zu den Erwartungen werden im daran anschließenden Abschnitt behandelt.

#### Objekte

Die Assoziation der gemessenen zu den erwarteten Objektmerkmalen basiert auf den relationalen Beziehungen, die in jedem Formmodell hinterlegt sind. Wie in Kapitel 8 beschrieben, bildet jedes Modell-Merkmal einen Knoten des Attributierten Relationalen Graphen (ARG). Die aus der Bildmessung erhaltenen Daten-Merkmale werden entsprechend ihrer Ausprägung im Bild und ihres Abstandes zum Modell-Merkmal bewertet und als Subknoten der Modell-Merkmals-Knoten in den ARG eingetragen. Damit hat man eine relationale Repräsentation und Bewertung der Messung in bezug auf das betrachtete Objekt aufgestellt. Die Auswertung des ARG, d. h. die Zuordnung eines Daten-Merkmals-Kandidaten zu einem Modell-Merkmals-Knoten kann jetzt über den Mechanismus des auf dem ARG

aufsetzenden Hopfield-Netzes erfolgen. Die Zuordnung mittels des Hopfield-Netzes garantiert, daß maximal ein Daten-Merkmal einem Modell-Merkmal zugeordnet wird. Bei einem entsprechend inkonsistenten ARG kann es aber auch sein, daß eine solche Zuordnung für manche Modell-Merkmals-Knoten unterbleibt, obwohl mögliche Kandidaten vorhanden sind. Im Anschluß an die so erhaltene Zuordnung erfolgt ein räumlicher Gütetest für die Gruppierung aller erwarteten Merkmale eines Objekts. Wenn ein bestimmtes Verhältnis erwarteter zu gemessenen und zugeordneten Merkmalen unterschritten wird, wird das Meßergebnis dieses Zyklus für die Hypothese verworfen. Typische Werte für dieses Verhältnis zugeordneter zu erwarteten Merkmalen liegen bei 60 - 80%. Sollte dieses Verhältnis bereits vor der Zuordnung aufgrund der zu geringen Anzahl an Merkmalen unterschritten werden, für die überhaupt mindestens ein Meßwert vorliegt, kann die Zuordnung natürlich gleich unterbleiben.

Bei Verwendung des einfachen Quadermodells und unter üblichen Aspektbedingungen bei Fahrten im Längsverkehr sind mit einer gewählten starren Kameraanordnung i. a. lediglich die Front- oder Heckansicht der Fahrzeuge sichtbar. Die sichtbaren Merkmale beschränken sich dann auf die beiden Seitenkanten, die Fahrzeugunterkante und -oberkante. Wie in Kapitel 12 im Rahmen der gezeigten Ergebnisse deutlich wird, ist das signifikanteste Merkmal die Fahrzeugunterkante, das am wenigsten aussagefähige Merkmal die Fahrzeugoberkante. Damit erhält man einen sehr eingeschränkten ARG mit festen Signifikanzen der Merkmale. Für einen solchen Fall - und ähnliche "Spezialfälle" - bietet es sich an, die Zuordnung nicht durch den intransparenten Mechanismus des Hopfield-Netzes vornehmen zu lassen, da der Aufwand für eine "handverdrahtete" Zuordnungslogik begrenzt ist, und darüberhinaus gestattet, für derart relevante Szenarien ggf. Sonderregeln für bestimmte Ansichten zu integrieren.

Das Verfahren für die "manuelle" Zuordnung der Heckansicht eines Quaders gestaltet sich wie folgt:

1. Bestimme den ARG inklusive aller Signifikanz- und Abstandsmaße.
2. Für jeden oberhalb einer Mindestgüte liegenden Zuordnungskandidaten der Fahrzeugunterkante evaluiere die Gruppierung mit möglichen Seitenkanten und bestimme die Güte dieser Gruppierung.
3. Wähle unter den evaluierten Gruppierungen diejenige mit der höchsten Güte aus.
4. Evaluiere, ob einer der Kandidaten für die Fahrzeugoberkante kompatibel ist und wähle den optimalen Merkmals-Kandidaten aus.
5. Eine Hypothese gilt nur dann als akzeptiert, wenn alle drei notwendigen Merkmale, also die Unterkante sowie beide Seitenkanten eine Zuordnung erhalten haben.

Wie man erkennt, ist die Fahrzeugunterkante das dominante Merkmal [Thomanek 96], während die Fahrzeugoberkante lediglich als "optionales" Merkmal betrachtet wird.

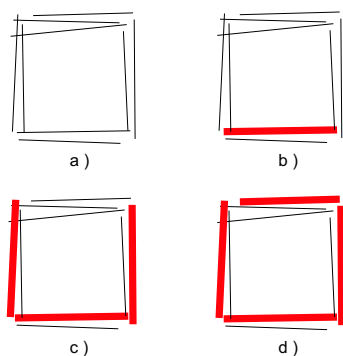
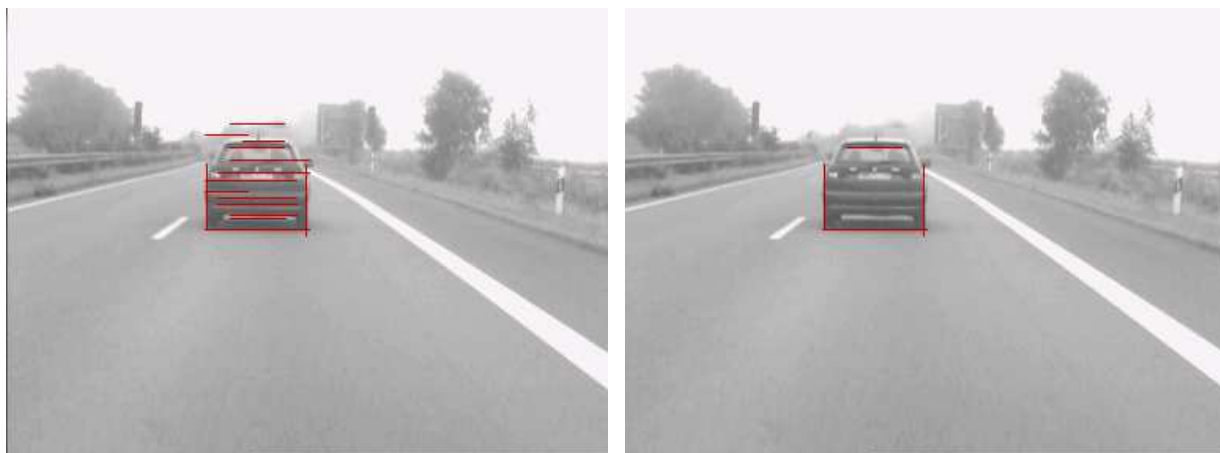


Abbildung 9.4: Beispielhafter Ablauf der "manuellen" Zuordnung

Der Ablauf des Verfahrens ist anhand der Beispielskizze in Abbildung 9.4 für eine rechteckförmige Konstellation veranschaulicht. Im Bild 9.4 a) sind für jedes Modellmerkmal mehrere Kandidaten zu erkennen. Durch die Reihenfolge der hervorgehobenen Merkmale unter allen Kandidaten in den Bildern 9.4 b) bis 9.4 d) soll die Priorisierung der Merkmale angedeutet werden. Zunächst wird als dominantes Merkmal in Bild 9.4 b) die "untere" Kante selektiert. In Bild 9.4 c) werden die hierzu kompatibelsten "Seitenkanten" ausgewählt und im letzten Schritt in Bild 9.4 d) schließlich eine dazu passende "Oberkante".

Ein Beispiel einer Zuordnung in einer realen Szene ist in Abb. 9.5 enthalten. Es handelt sich dabei um die Szene, die in Kapitel 7 bereits als Beispiel für die Bildvermessung diente. Es werden - bis auf die Fahrzeugoberkante - offensichtlich die "richtigen" Kanten selektiert. Für die Fahrzeugoberkante wird die Oberkante der Heckscheibe ausgewählt, da diese signifikanter ausgeprägt ist als das eigentliche, die Fahrzeugoberkante beschreibende Merkmal. Die Auswahl scheint in dieser Szene allerdings auch nicht besonders schwierig zu sein - aber ist sie das (für uns Menschen) nicht eigentlich immer ?



a) Durch Verkettung erhaltene Konturkandidaten

b) Durch relationale Zuordnung selektierte Konturkanten

Abbildung 9.5: Zuordnung von Merkmals-Kandidaten zu Modell-Merkmalen

## Fahrspur

Bei den Merkmalen der Fahrspur muß unterschieden werden, ob es sich um eine Fahrbahnmarkierung als Begrenzung handelt, oder ob die Begrenzung der Fahrspur durch eine einfache Kontur erfolgt. Bei Markierungen handelt es sich demzufolge an sich schon um eine einfache Aggregation zweier Konturen in Form eines Dunkel-Hell-Dunkel-Übergangs

(unter normalen Umgebungsbedingungen).

Das Vorgehen bei der Assoziation sieht im einzelnen wie folgt aus:

1. Bestimme für jedes Modell-Merkmal aus der Liste der vorhandenen Extremwerte Paare aufeinanderfolgender Dunkel-Hell-Dunkel-Übergänge (DHD).
2. Fasse die Extrema eines DHD-Übergangs zu einem "Linien-Merkmal" zusammen, wobei der Grauwert des mittleren "hellen" Segments sich als Mittelwert des hinteren bzw. vorderen Flächenstücks der beiden Extrema ergibt.
3. Ermittle den Abstand aller gefundenen Linien-Merkmale zum Erwartungswert des Modell-Merkmals
4. Wähle aus den Linien-Merkmalen dasjenige mit dem geringsten Abstand - also den "Next-Neighbour" - als Meßwert aus, sofern er einen Maximalabstand nicht überschreitet.
5. Bestimme - getrennt für jede Fahrspurmarkierung - die Güte der Messung über alle vermessenen Modell-Merkmale. Als Maße dienen die Geschlossenheit der Markierung und das Verhältnis erwarteter zu gemessenen und zugeordneten Merkmalen.
6. Wenn die Güte ausreichend ist, werden die Meßwerte für die entsprechende Markierung akzeptiert und die Attributwerte der Merkmale gemäß den Werten des Meßwerts aktualisiert. Die Messungen werden in den Meßvektor des Kalman-Filters eingetragen und das Zuordnungsrauschen entsprechend der Mehrdeutigkeit der Messung besetzt (Wie viele Linien-Kandidaten gab es ?).
7. Andernfalls werden die Messungen für die gesamte Markierung verworfen und die Attributwerte werden nur insofern aktualisiert als das Vertrauen in sie durch Herabsetzen der zeitlichen Gewichtung vermindert wird.
8. Im letzten Schritt wird das Gesamtvertrauen in jede Fahrspurbegrenzung aktualisiert, d. h. bei erfolgreicher Zuordnung erhöht, andernfalls erniedrigt.

### 9.3 Meßmodell Lidar

In gleicher Weise wie für den Videosensor beinhaltet das Meßmodell des Lidarsensors die sensorspezifischen Verarbeitungsschritte in der Objekterkennung nach Bild 9.1. Aufgrund der im Vergleich zum Videosensor sehr grob diskretisierenden horizontalen Strahlgeometrie des Lidarsensors ergeben sich gänzlich andere Bedingungen für die Merkmalsvorhersage und - verbunden mit den wenigen Meßpunkten - andere Möglichkeiten zur Merkmalsvermessung und -assoziation. Ein weiterer Unterschied besteht darin, daß mit Hilfe eines Lidarsensors die primären Merkmale der Fahrspur nicht zu vermessen sind, d. h. Fahrbahnmarkierungen oder Übergänge zwischen Fahrbahnbelag und Fahrbahnrand. In gewissem Rahmen kann über Objekte, die als zur Fahrbahninfrastruktur gehörend klassifiziert werden (insbesondere Leitpfosten und Leitplanken) eine Schätzung des Verlaufs des Fahrbahnbandes erfolgen.

Hierbei handelt es sich dann allerdings um eine Assoziation der Objekte mit der Fahrbahn auf höherer Ebene. An dieser Stelle wird es daher bzgl. des Meßmodells allein um die Objektvermessung gehen.

### 9.3.1 Vorhersage der Merkmale

Im Gegensatz zu den beim Videosensor eingesetzten Kanten-Merkmalen liefert der Lidar-sensor Entfernungspunkte als Merkmale. Dabei handelt es sich primär um Flächenmerkmale, da bei der Messung die Reflektivität der vom Laserstrahl getroffenen Flächen ausschlaggebend ist. Neben dem Material der Fläche spielt die Orientierung hierbei eine wesentliche Rolle. Generell ist die Reflektivität proportional zur effektiven Flächengröße, d. h. orthogonal zur Strahlrichtung des Lasers stehende Flächen reflektieren maximal. Es ist daher sinnvoll, neben der rein geometrischen Sichtbarkeitsvorhersage der Merkmale auch die Orientierung zum Laserstrahl in die Betrachtungen miteinzubeziehen. Allerdings ist die Streuung der Reflektivität in Abhängigkeit der Umweltbedingungen, des Materials und des Verschmutzungsgrades sehr hoch. Ein weiterer Aspekt ist die mit der Entfernung in der 4. Potenz abnehmende reflektierte Leistung, die analog zur auflösungsbedingten Sichtbarkeitsgrenze des Videosensors zu einer maximalen Reichweite führt.

Ausgangspunkt für die Vorhersage der Merkmale sind wie beim Videosensor die in den Merkmalsraum transformierten Objektmerkmale. Da die Sensoreigenschaften eine Feinvermessung von Formen nicht ermöglichen, wird hier sowohl für die Sichtbarkeitsberechnung, wie auch die eigentliche Vermessung auf den einhüllenden Quader respektive - da der Lidarsensor nur in einer horizontalen Ebene sensiert - auf das einhüllende Rechteck übergegangen. Ausgehend vom einhüllenden Rechteck sind nun drei Wege möglich, zu Merkmalen zu gelangen:

1. Es werden analog zum Videosensor die sichtbaren, nicht verdeckten Objekte bestimmt. Die einhüllenden Rechtecke bzw. deren Kanten werden als Merkmale verwendet. Dies erfordert eine im folgenden Abschnitt 9.3.2 beschriebene Weiterverarbeitung der vom Sensor gelieferten Entfernungsrohdaten hin zu ausgedehnten Segmenten, die mit den Objekten bzw. deren Kanten assoziiert werden können. Als Meßgrößen für die Schätzung dienen die für die Segmente ermittelten Attribute (Lage, Form).
2. Es wird direkt auf Basis der Lidarstrahlen gearbeitet. D. h. es werden die Schnittpunkte der Lidarstrahlen mit den Objektkanten bestimmt. Diese dienen direkt als Meßgrößen für die Schätzung.
3. Die beiden vorgenannten Möglichkeiten werden ähnlich dem Vorgehen bei der Bildvermessung kombiniert. Es werden sowohl die Einzelmerkmale auf Basis der Strahlen vorhergesagt als auch die Objektkanten. Es erfolgt dann eine Segmentierung im Sinne einer Gruppierung unter Berücksichtigung von Gestaltannahmen. In die nachfolgende Schätzung wiederum gehen dann aber nur die über die Gruppierung ausgewählten Einzelstrahlmeßwerte ein.

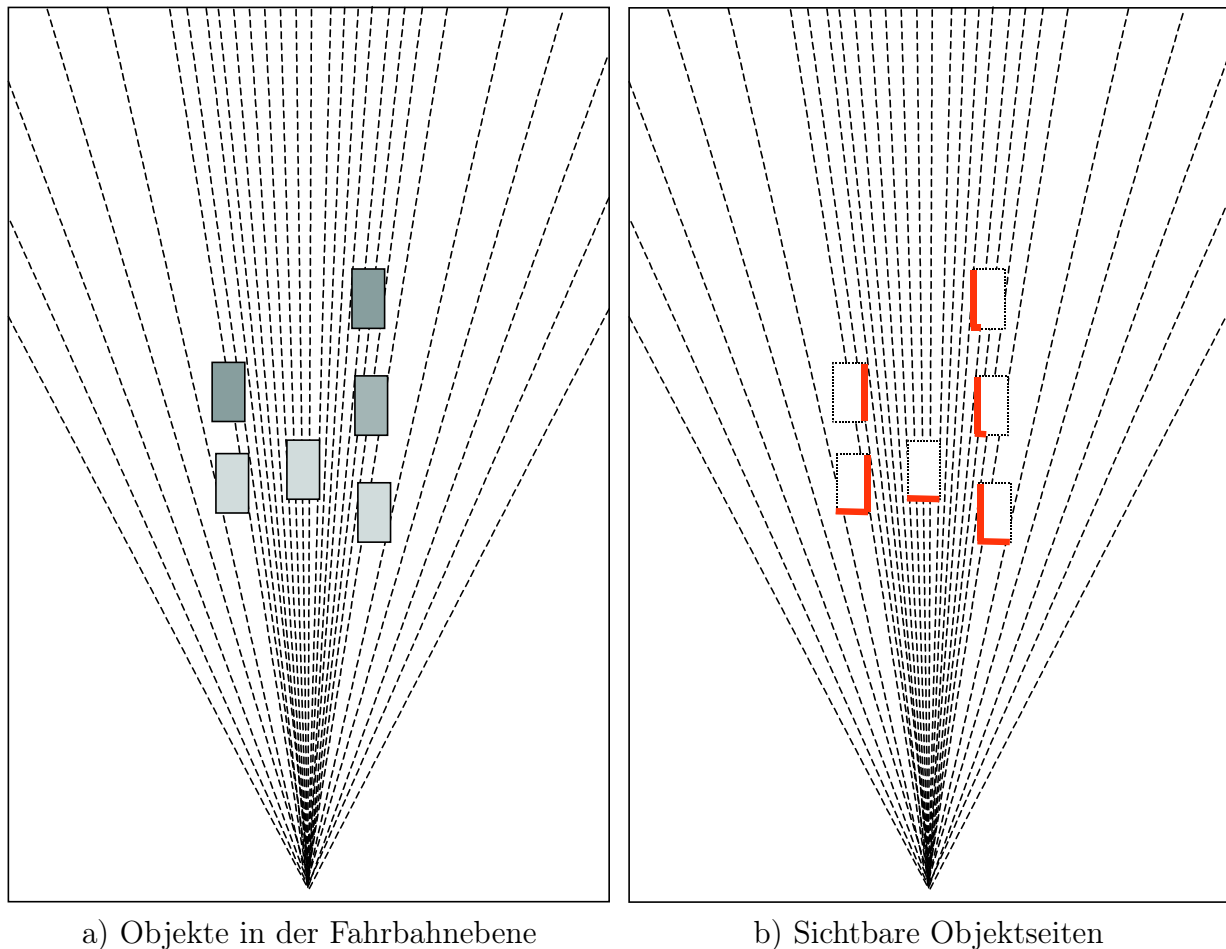


Abbildung 9.6: Verdeckungsanalyse einer Szene aus Sicht des Mehrstrahl lasers in der Draufsicht

Die Abb. 9.6 zeigt in Analogie zur Abb. 9.3 für den Videosensor dieselbe Szene, mit dem Unterschied, daß zusätzlich die Strahlen des Lidarsensors eingezeichnet sind, um einen Eindruck von der Auflösung der Sensors zu bekommen. Wie man sieht, treffen im zentralen Strahlenbereich jeweils mehrere Laserstrahlen auf die Objekte, die ca. 30 m entfernt sind. Objekte im peripheren Strahlenbereich werden hingegen nur von jeweils 1-2 Strahlen je Objektseite getroffen. Damit ist hier die Auflösungsgrenze erreicht, an der man sinnvoll ausgedehnte Objekte mit dem Lidarsensor messen und die Formparameter schätzen kann.

### 9.3.2 Vermessung der Merkmale

Die Vermessung der Merkmale besteht zum einen aus der Laufzeitmessung der Laserstrahlen. Diese liefert je Strahl bis zu 3 Entfernungspunkte. Die Bestimmung dieser Entfernungspunkte aus den Rückstreuereurven erfolgt im Steuergerät des Sensors. Das im Steuergerät durchgeführte Meßverfahren kann zwar in mehrfacher Hinsicht parametrisiert werden, jedoch geschieht dies im Rahmen eines separaten Offline-Kalibrierschritts; zur Laufzeit hingegen sind das Meßverfahren und seine Parameter derzeit fest.

Die vom Sensor gelieferten Entfernungsinformationen werden in einem ersten Schritt einer Vorfilterung unterworfen. Damit wird sichergestellt, daß zufällige oder auch systematische Fehlmessungen nicht in die weitere Auswertung eingehen. Dies betrifft im wesentlichen zwei Fälle:

1. Durch Verschmutzung der Linsen oder vor dem Sensor befindlicher (Streu-) Scheiben kann es zu einem konstanten Ziel in sehr naher Entfernung vor dem Sensor kommen. Dies kann auch bei dichtem Regen oder bei Nebel der Fall sein. Diese Informationen sind zwar für die Erkennung von Umweltinformationen oder Informationen über den Systemstatus durchaus relevant, haben aber mit der eigentlichen Objekterkennung nichts zu tun. Sie werden daher nach Möglichkeit herausgefiltert.
2. In Abhängigkeit des Fahrbahnbelags, insbesondere bei vertikaler Krümmung des Fahrbahnverlaufs oder vertikalen Strukturen kann es - verstärkt durch Fahrzeugnickbewegungen - zu "Fehlmessungen" kommen, da stets ein gewisser Teil der Laserenergie in Richtung des Bodens abgestrahlt wird. Um diese in der Regel nur in ein oder zwei aufeinanderfolgenden Meßzyklen auftretenden Störungen auszufiltern, werden die vom Sensor gelieferten Entfernungswerte anhand ihrer Signalstärke in Abhängigkeit der Entfernung vorgefiltert. Dies ist möglich, da reale Objekte unter regulären Umweltbedingungen i. a. recht hohe Signalstärken liefern. Um keine tatsächlichen Nutzinformationen herauszufiltern, wird die Schwelle allerdings so gelegt, daß es stets einige Fehlmessungen gibt, die nicht herausgefiltert werden. Diese gilt es dann über Regeln auf höherer Ebene zu eliminieren.

In der Abbildung 9.7 sind die vom Lidar gelieferten Rohdatenpunkte in Verbindung mit den angedeuteten Lidar-Strahlen veranschaulicht. Da der Videosensor einen geringeren Öffnungswinkel als das Lidar hat, sind dort nicht alle Rohdaten zu erkennen. Man beachte, daß allein im Zentralbereich mit einer Strahlaufweitung von  $0.9^\circ$  die Breite eines Strahls in ca. 100 m bereits der Breite eines Personenwagens entspricht. In der Abbildung deutlich zu erkennen sind das vorausfahrende Fahrzeug in der eigenen Spur bei ca. 45 m Entfernung, der LKW in der rechten Spur bei ca. 90 m sowie das Verkehrsschild am rechten Straßenrand bei ungefähr 100 m. Bis zu einer Entfernung von ca. 40 m lassen sich auch Rohziele einer Ablage von ungefähr 3 m erkennen. Hierbei handelt es sich um die linke Leitplanke oder auf dem Mittelstreifen befindliche Büsche.

## Segmentierung

Im Anschluß an die Vorfilterung folgt die Segmentierung der Rohdaten. Das Ziel der Segmentierung soll eine Datenkomprimierung für die folgende Assoziation sein. Es wird nicht zwangsläufig angestrebt, solche Segmente zu finden, die stets kompletten Objekten zugeordnet werden können. Dieser Fall kann in der Assoziation durch Zuordnungen mehrerer Segmente zu einem Objekt berücksichtigt werden. Ziel der Segmentierung ist vielmehr,



die Rohdaten datengetrieben zu gruppieren und so die “Bottom-Up“-Komponente in der Objekterkennung zu stärken.

Die Segmentierung erfolgt nach dem Schema:

1. Berechne für alle Rohdaten die maximal 2 nächsten Objekte unter Berücksichtigung einer maximalen Schranke.
2. Bestimme die Entfernung in x- und y-Richtung von Rohdaten auf benachbarten Strahlen, oder falls auf dem Nachbarstrahl keine Meßpunkte liegen, auf dem maximal übernächsten Strahl. Liegen sowohl x- wie y-Distanz  $\Delta x$  bzw.  $\Delta y$  unterhalb (getrennt) vorgegebener Schranken können die Meßpunkte zu einem Segment gehören.
3. Fasse alle im vorhergehenden Schritt ermittelten potentiellen Kandidaten zu Segmenten zusammen. Punkte nicht benachbarter Strahlen werden lediglich zusammengefaßt, wenn die Einzelpunkte jeweils dasselbe nächste Objekt besitzen.
4. Ermittle für die Segmente die Attribute durch Bildung eines einhüllenden Rechtecks mit Länge, Breite und x-/y-Position des Segments.

Bild 9.8 zeigt die auf der Basis der in Abbildung 9.7 gemessenen Rohdaten ermittelten Segmente. Die Segmentierung erfolgte bei Werten von  $\Delta x = 2$  m und  $\Delta y = 1$  m. Wie man erkennt, sind kompakte Objekte wie die Fahrzeuge oder auch das Verkehrsschild in ca. 100 m Entfernung durch die Segmentierung gut wiedergegeben. Die Leitplanke am linken Fahrbahnrand hingegen zerfällt in mehrere Segmente. Wenn die Fahrbahnbegrenzung als solches als geschlossenes Objekt gesucht wird, bietet sich ein weiterer Segmentierungsschritt an, der entweder erneut auf den Rohdaten, die keinen Objekten im Fahrbahnbereich zugeordnet wurden, oder auf den bereits gefundenen Segmenten in diesem Bereich aufsetzt.

### 9.3.3 Assoziation der Merkmale

Als Ergebnis der Segmentierung erhält man eine Segmentliste als Basis für die Assoziation mit getrackten Objekten. Die Liste enthält neben tatsächlichen Segmenten mit mindestens 2 Punkten auch die keinem Segment zugeordneten Rohdatenpunkte als “entartete“ Segmente, da auch diese mit Objekten zu assoziieren sind. Wie weiter oben bereits erwähnt, entspricht ein Personenwagen in ca. 100 m Entfernung vom Sensor gerade noch der Breite eines Laserstrahls, so daß dort i. a. nicht mit mehr als einem Entfernungspunkt je Objekt gerechnet werden kann.

Zur Assoziation der Segmente mit den Objekten wird wie folgt verfahren:

1. Es werden für alle Segmente die Mahalanobis-Abstände zu allen Objekten berechnet. Alle unterhalb eines definierten  $\chi^2$ -Abstands gelegenen Abstände werden in der Distanz-Matrix vermerkt.

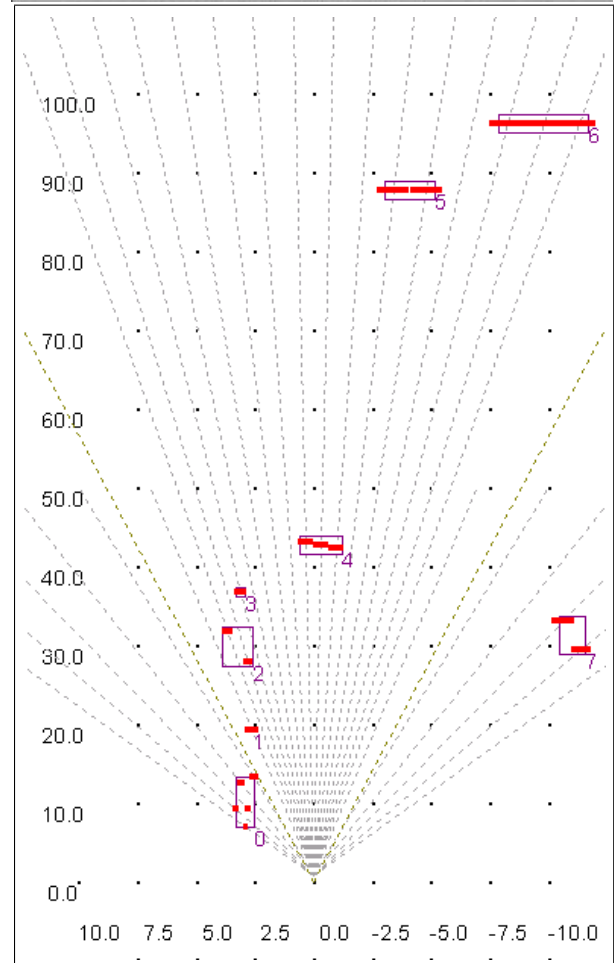
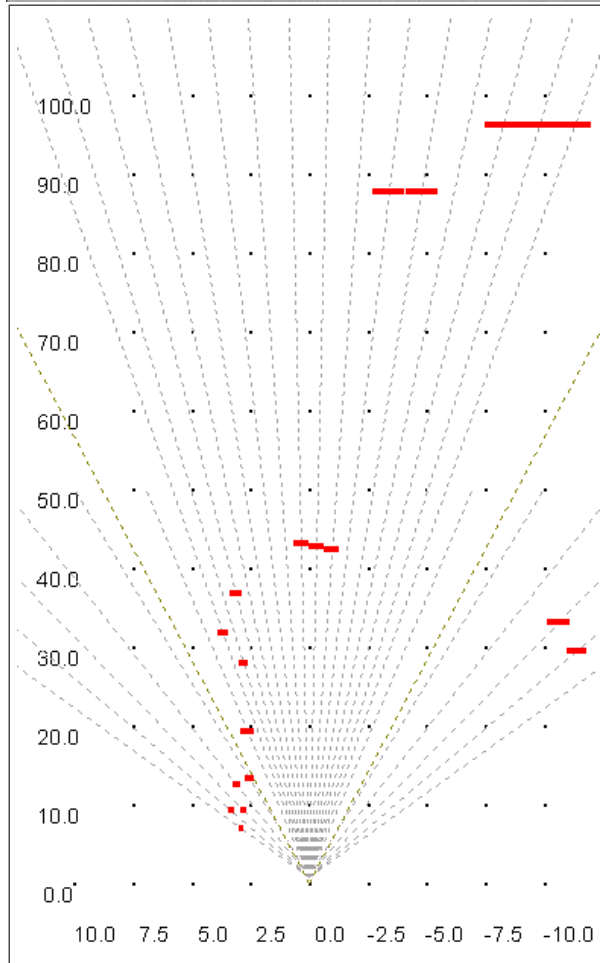
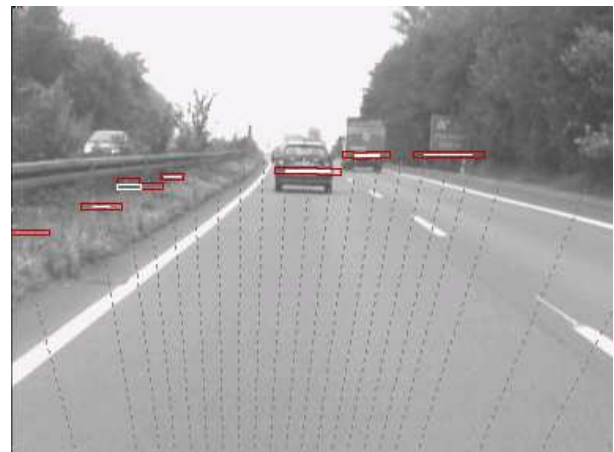
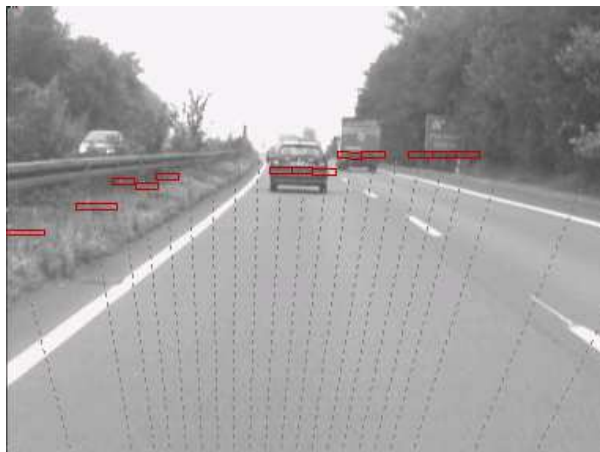


Abbildung 9.7: Lidar-Rohdaten im Video-  
bild und aus der Draufsicht

Abbildung 9.8: Lidar-Segmente im Video-  
bild und aus der Draufsicht

2. Es werden die global optimalen Zuordnungen von Segmenten zu Objekten vorgenommen. Anstelle der global optimalen Zuordnung sind andere Zuordnungsstrategien denkbar (z.B. Auktions-Algorithmus [Bertsekas 01]). Für die in dieser Arbeit untersuchten Aspekte hat sich die globale Strategie aber als ausreichend erwiesen. Falls die Zuweisungsstrategie dazu führt, daß mehrere Segmente einem Objekt zugewiesen

werden sollen, so wird geprüft, ob ein aus einer Vereinigung der Segmente entstehendes neues Segment innerhalb der maximalen Mahalanobis-Distanz liegt. Ist dies der Fall, wird das vereinigte Segment zugewiesen, ansonsten das optimale Einzelsegment.

3. Alle nicht zugewiesenen Segmente werden an die im Abschnitt 9.5.1 erläuterte Routine zur Initialisierung neuer Objekthypothesen übergeben.

## 9.4 Form- und Bewegungsschätzung

Die Schätzung von Form und Bewegung der in der Szene befindlichen Objekte, der Fahrspur sowie des Eigenzustands basiert auf den durch den Video- und Lidarsensor gemessenen Merkmalen, die mit den vorhergesagten Modell-Merkmalen assoziiert werden konnten. Die Meßwerte des Videosensors gehen sowohl in Fahrspur und Eigenzustandsschätzung wie in die Objekt-Form- und Bewegungsschätzung ein, wohingegen die Merkmale des Lidarsensors nur in die Objektschätzung eingehen. Im Falle des Videosensors erfolgt die Schätzung in der Reihenfolge:

1. Fahrspur und relativer Eigenzustand
2. Objekt-Form und -Bewegung

In den folgenden Abschnitten wird die Schätzung der Zustandsgrößen in der eben genannten Reihenfolge beschrieben.

### 9.4.1 Fahrspur und relativer Eigenzustand

Die Schätzung der Formparameter der Fahrspur sowie des relativen Eigenzustands bzgl. der Fahrspur lehnt sich im wesentlichen an das in [Behringer 96] gegebene Verfahren an. Dieses sieht eine kaskadierte Schätzung anhand mehrerer Schätzer vor. Die Schätzung erfolgt in der Reihenfolge:

1. Nickwinkel und Nickwinkelgeschwindigkeit,
2. Fahrspurbreite (und -breitenänderung),
3. Relativer Gierwinkel und Ablage,
4. Horizontale Krümmungsparameter (sowie gegebenenfalls die vertikalen Krümmungsparameter)

Um die Beobachtbarkeit des Gesamtsystems zu gewährleisten, wird bei der Schätzung einiger Zustände zwingend vorausgesetzt, daß links- wie rechtsseitig eine Fahrspurbegrenzung gefunden wurde. Dies betrifft die Schätzung der Nicklage wie der Fahrspurbreite. Die

anderen Zustandsgrößen werden auch bei Vorhandensein nur einer Fahrspurbegrenzung geschätzt.

Die Reihenfolge der Schätzer folgt aus der Beeinflussung der Merkmalsausprägung im Bild durch die jeweilige Zustandsgröße. Da diese bei der Nicklage aufgrund der perspektivischen Bedingungen sehr hoch ist, steht die Schätzung dieser Komponente am Anfang. Nach jedem Schätzschritt ergeben sich neue innovierte Werte für die jeweiligen Zustandsgrößen. Auf der Basis dieser neuen Werte erfolgt eine erneute Transformation der Merkmale in den Meßraum, auf deren Basis eine Neuberechnung der Residuen für die assoziierten Merkmal-Meßwert-Paare vorgenommen wird. Die Berechnung der Elemente der Jacobi'schen Meßmatrix erfolgt auf der Basis der zugeordneten Meßwerte numerisch mit dem in Abschnitt 6.3 gezeigten Verfahren. Im Anschluß an die Schätzung erfolgt eine Plausibilitätsprüfung der Zustandsgrößen. Der Gesamtablauf der Schätzung von Fahrspur und relativem Eigenzustand ist zusammenfassend in Abbildung 9.9 wiedergegeben.

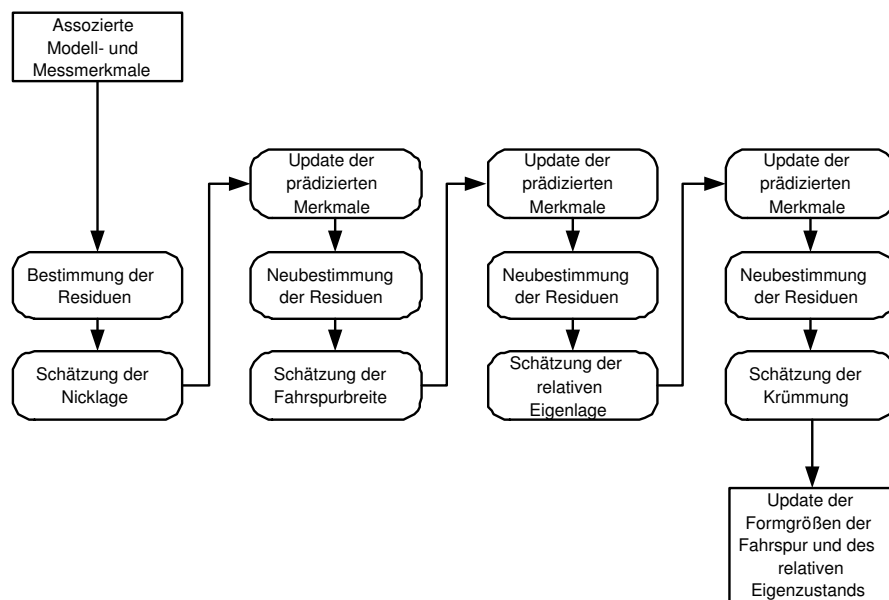


Abbildung 9.9: Ablaufdiagramm der Fahrspur- und Eigenzustands-Schätzung

### 9.4.2 Objekte

Wie in Abschnitt 6.3 ausgeführt, werden die Merkmale eines Objekts nur für die Schätzung der Form- und Bewegungsparameter des Objekts selbst genutzt. Die Berücksichtigung der Eigenbewegung erfolgt anhand der fahrzeugeigenen Sensorik. Da es sich bei den betrachteten Objekten um starre Formen handelt, sind die Formparameter - anders als beim Fahrspurmodell - feste, aber unbekannte Parameter.

Vor diesem Hintergrund gibt es mehrere Möglichkeiten zur Form- und Bewegungsschätzung, die in Abbildung 9.10 gezeigt sind. Im ersten Fall (a) werden aus den ermittelten Residuen in einem Schritt Form- und Bewegungsgrößen geschätzt. Im Fall (b) werden die

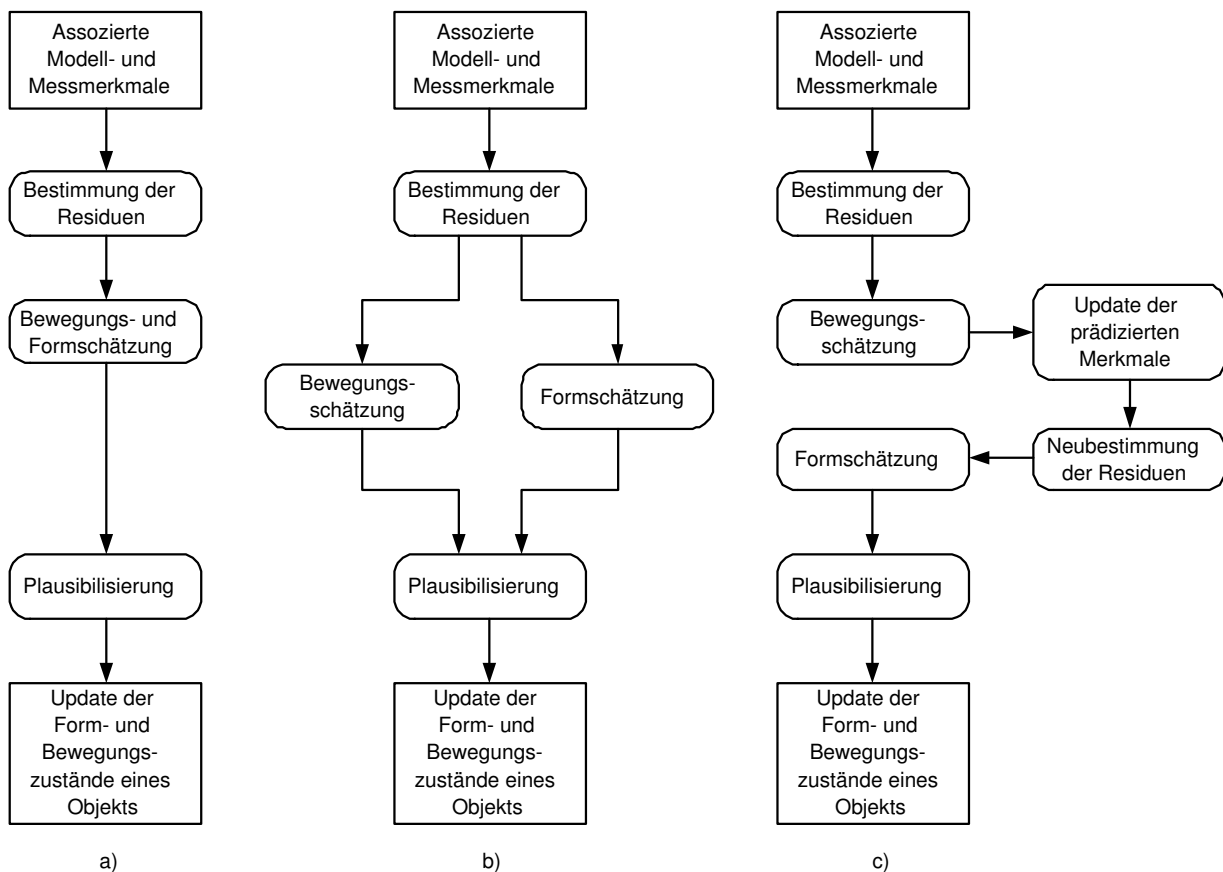


Abbildung 9.10: Möglichkeiten zur Form- und Bewegungsschätzung von Objekten

aus der Messung resultierenden Residuen parallel auf die Form- wie auf die Bewegungsschätzung gegeben. Im dritten Fall (c) werden sequentiell die Bewegungsgrößen und im Anschluß - nach einer Neuberechnung der Residuen - die Formparameter geschätzt. Im Fall (a) wird die "Verteilung" der durch das Residuum bedingten Innovation auf Form- bzw. Bewegungsgrößen in der Hauptsache durch die Systemkovarianzen  $\underline{Q}$  bestimmt. Um nicht eine zu hohe "Dynamik" in der Schätzung der an sich festen Formparameter zu erhalten, muß man dies bei der Besetzung von  $\underline{Q}$  berücksichtigen und verliert damit an Freiheit in der Wahl von  $\underline{Q}$ . Für Formen mit wenigen Formparameter wie etwa den hier verwendeten Quadern ist dies angesichts der überschaubaren Anzahl zu schätzender Größen noch akzeptabel. Bei komplexeren Formen wie beispielsweise dem Fahrzeug-Polyedermodell hat es sich als vorteilhaft erwiesen, zu der Schätzung nach (c) überzugehen, um die Schätzer in überschaubaren Dimensionen zu halten und auftretende Effekte beurteilen zu können. Die Schätzung nach (b), bei der die kompletten Residuen auch auf die Formschätzung gegeben werden, stellt dagegen nur für Fälle eine Alternative dar, bei denen eine Parallelisierung, z. B. aus Rechenzeitgründen, notwendig ist. Eine Darstellung verschiedener Möglichkeiten zur gleichzeitigen Schätzung von Form und Bewegung ist in [Schick 92] enthalten - unter Maßgabe der damals verfügbaren Rechenleistung. Dort wurden nicht nur Form und Bewegung separat geschätzt, sondern darüberhinaus die Formparameter für jede einzelne Kante in einem separaten Schätzer bestimmt und in einem nachgeschalteten Verarbeitungsschritt über alle Schätzungen gemittelt. Im Gegensatz dazu wurden in [Koller 92]

die Zustandsgrößen eines dem hier und in [Schick 92] ähnlichen Fahrzeug-Polyedermodells in einem Schätzer ermittelt. Die bei der Schätzung verwendeten Werte für die Systemkovarianzen werden aus der statischen Objektmodelldatenbank entnommen.

Ähnlich wie für die Fahrspur ist für alle Zustandsgrößen der Objekte in deren Modell hinterlegt, aus welchen Merkmalen diese geschätzt werden. Nur wenn die Beobachtbarkeit einer Zustandsgröße über die zur Verfügung stehenden Merkmale gegeben ist, wird diese auch geschätzt. Im Anschluß an die Form- und Bewegungsschätzung erfolgt ein Konsistenztest aller Zustandsgrößen auf Plausibilität, hinsichtlich:

1. des Betrags von Geschwindigkeiten und Beschleunigungen, die begrenzt sind, und
2. hinsichtlich Gültigkeitsintervallen für Formparameter für verschiedene Objektklassen.

Wesentliche Unterscheidungsmerkmale für verschiedene Objektklassen sind vor allem die Geschwindigkeit - hierbei ist allein schon die Information, daß ein Objekt überhaupt beweglich ist, relevant - sowie das Aufenthaltsgebiet eines Objektes in der Szene (im Fahrbahnbereich). Auf diese Thematik wird im Kapitel 10 noch näher eingegangen. Sofern ein Objekt Wertebereiche verletzt, erfolgt entweder eine Beschränkung der jeweiligen Zustandsgröße auf das Gültigkeitsintervall oder - bei entsprechend hohen Abweichungen von den zulässigen Werten oder bei wiederholten Abweichungen - wird die Objekthypothese verworfen. In diesem Fall werden die dem Objekt zugeordneten Merkmale wieder in die Liste der nichtzugeordneten Merkmale übernommen und dem im nächsten Abschnitt beschriebenen Initialisierungsverfahren übergeben.

## 9.5 Initialisierung

Wenn durch die Sensoren gelieferte Meßwerte nicht mit vorhergesagten Merkmalen getrackter Objekte assoziiert werden können, stehen diese Merkmale zur Instanziierung neuer Objekthypothesen zur Verfügung. Dies gilt insbesondere nach Aktivierung der Objekterkennung, wenn noch keine getrackten Objekthypothesen vorliegen. In diesem Abschnitt werden die jeweiligen sensorspezifischen Methoden erläutert, die dazu dienen, aus diesen Merkmalen neue Objekthypothesen zu generieren.

### 9.5.1 Lidar

Der Lidarsensor stellt aufgrund seiner Sensorcharakteristik den hauptsächlichen Sensor für die Instanziierung neuer Hypothesen für Objekte dar, da er sowohl die Reichweite wie den azimuthalen Winkel betreffend den weiteren Erfassungsbereich als der hier verwendete Videosensor besitzt. Auf die Strategie zur Sensorauswahl und -konfiguration in Verbindung mit der Objekterkennung wird im anschließenden Kapitel 10 eingegangen.

Als Grundlage der Initialisierung dienen die aus den Lidarrohdaten gebildeten Segmente, die keinem getrackten Objekt zugeordnet werden konnten. Im ersten Schritt wird aus

diesen Segmenten direkt eine Objekthypothese aufgesetzt, die mit der  $x - /y$ -Position des Segments und seiner Länge und Breite initialisiert wird. Da der Laser keine direkte Geschwindigkeitsmessung liefert, wird diese zunächst zu 0 gesetzt. Einen Hinweis auf die Objektklasse und damit verbundene Randbedingungen liefert der Szenenbereich, in dem sich neu instanziierte Objekte befinden. Dieses Wissen wird aber im ersten Schritt noch nicht ausgewertet, da zunächst ein weiterer Meßzyklus des Lidarsensors abgewartet wird. Kann in diesem Meßschritt ein Segment mit der neuen Objekthypothese assoziiert werden, werden erneut die Werte für Position und Ausdehnung des Segments in die Objekthypothese übernommen. Zusätzlich erfolgt auf der Basis der Positionsänderungen eine Initialisierung der Objektgeschwindigkeit. Dieses Vorgehen hat sich als vorteilhaft erwiesen, da ansonsten die Einschwingdauer des Kalman-Filters auf die Objektgeschwindigkeit und in der Folge davon die Meßwertassoziation in den ersten Meßzyklen nach der Instanziierung zu unsicher sind. Die große Variabilität der möglichen Szenen (stehende Objekte, in Eigenrichtung bewegte Objekte, entgegenkommende Objekte, querbewegte Objekte) ließe sich ansonsten nur durch einen aufwendigen Multihypothesenansatz zufriedenstellend lösen. Demgegenüber ist der hier verwendete Ansatz jedoch deutlich schlanker. Die Verfahrensschritte der Instanziierung von Objekthypothesen basierend auf Lidardaten ist in Abbildung 9.11 nochmals dargestellt. Im Anschluß an diesen abschließenden Initialisierungsschritt können dann die o. g. Nebenannahmen in Verbindung mit dem Szenenbereich, in welchem sich das Objekt aufhält, und den Objektattributen wie Länge, Breite und der Geschwindigkeit in die weitere Interpretation einfließen.

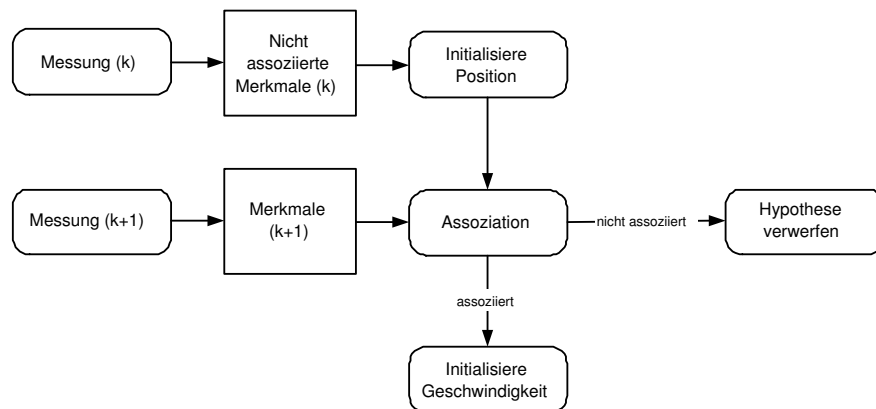


Abbildung 9.11: Lidarbasierte Instanziierung von Objekthypothesen

### 9.5.2 Video

Im Gegensatz zur lidarbasierten Instanziierung von Objekthypothesen, muß im Falle des Videosensors deutlich mehr Hintergrundwissen über die Szene und die in der Szene erwarteten Objekte berücksichtigt werden, um zu Hypothesen über das Umfeld zu gelangen. Insbesondere muß zum Aufsetzen einer Hypothese die Lage im Raum aufgrund der Nähe zu anderen Objekten im Bild oder durch andere Nebenannahmen gestützt angenommen werden. Daraus folgt auch für die Form eine entsprechende Skalierung. Zusätzlich zum

Lidarsensor ist durch den Videosensor auch das Fahrspurmodell zu instanziiieren. Auf der Basis einer vorhandenen Fahrspur kann dann die Initialisierung von Hypothesen über Objekte gestützt werden.

## Fahrspur

Die Initialisierung der Fahrspurahypothese stellt auf der höheren Ebene (Kapitel 10) einen deutlichen Unterschied zwischen der Anwendung beim autonomen Fahren oder im Bereich Fahrerassistenz dar. Zwar können in beiden Fällen Annahmen über die Umgebung getroffen werden (z. B., daß man sich auf befahrbarem Grund befindet), doch aus der Verantwortung des Fahrers für die Fahrzeugführung im Bereich Fahrerassistenz sind die Anforderungen an die Güte der Hypothesenüberprüfung deutlich geringer. Im Gegensatz dazu kann jede Fehlhypothese im Bereich des autonomen Fahrens zu kritischen Situationen führen. An dieser Stelle werden nur vereinfachte Ansätze zur Hypothesengenerierung im Bereich Fahrerassistenz betrachtet; Beiträge zu der beim autonomen Fahren vorliegenden komplexen Thematik kann den Arbeiten [Gregor 02] und [Lützel 02] entnommen werden.

Das Vorgehen zur Initialisierung der Fahrspurahypothese basiert bei fahrendem Fahrzeug auf der Annahme, diese in Fahrrichtung vor dem Fahrzeug zu erwarten. (Anmerkung: Im Falle eines stehenden Fahrzeugs ist die Relevanz der Information nicht gegeben, da kein direkter Handlungsbedarf besteht.) Es wird kein besonderes Initialisierungsverfahren eingesetzt, sondern stattdessen versucht, mit einer Start-Hypothese sofort in die rekursive Schätzung einzusteigen. Die Fahrspur wird hierzu zunächst ohne Krümmung mit einer aus dem Szenenkontext abgeleiteten Breite erwartet und einem relativen Gierwinkel  $\psi = 0$  sowie ohne Ablage ( $y = 0$ ) angenommen (siehe auch Abschnitte 6.1.2 und 6.2.5). Sodann wird die so instanziierte Hypothese über die o. a. Verfahren im Videobild vermessen und über rekursive Zustandsschätzung an die reale Fahrspur adaptiert. Parallel zu den Zuständen werden die in Abschnitt 9.2.2 angegebenen Gütemaße initialisiert und adaptiert. Sie und die Plausibilitätsprüfung der Zustandsgrößen führen zur Annahme oder zum Verwerfen der Fahrspurahypothese. Das gleiche Vorgehen findet auch bei einem Verlust der Fahrspur während der Laufzeit statt.

Als mögliche Erweiterungen dieses Vorgehens wäre eine Initialisierung der horizontalen Krümmung mit einer Fahrschlauchschätzung auf der Basis beispielsweise der Eigengierrate denkbar. Dies könnte in dynamischeren Szenen auf Landstrassen eventuell Vorteile versprechen, ist hier aber nicht näher untersucht worden.

## Objekte

Die videobasierte Instanziierung von Objekten ist wie das videobasierte Objekttracking auf die Objektklassen beschränkt, für die Modelle vorliegen. Wie im Abschnitt 9.5.1 erwähnt, ist aufgrund der gewählten Sensorkonfiguration eine Instanziierung von Objekten ohne vorhergehende Erfassung durch den Lidarsensor an sich nicht möglich. Es sind daher zwei Fälle zu unterscheiden:



1. Instanziierung einer “Video“-Objekthypothese eines bereits durch den Lidarsensor erfaßten und instanziierten Objekts
2. Instanziierung einer Objekthypothese eines noch nicht erfaßten Objekts

Im ersten Fall steht die Initialisierung der Bildattribute der Objektmerkmale im Mittelpunkt. Aufgrund des bereits angesammelten Wissens über das Objekt in dem bereits erfolgten Tracking ist der Suchraum der möglichen Objekte deutlich eingeschränkt. Im zwei-

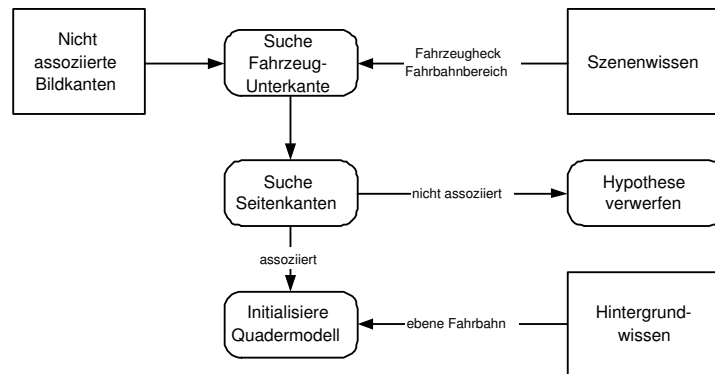


Abbildung 9.12: Vorgehen bei der videobasierten Instanziierung von Objekthypothesen

ten Fall hingegen kann allein der Szenenbereich als Hinweis auf das zu instanziiierende Objekt dienen. Für getrackte Objekte steht mehr Wissen zur Verfügung. Da in dieser Arbeit bzgl. des Videosensors nur die Fähigkeit zur Fahrzeugerkennung betrachtet wird, erfolgt in Abhängigkeit der Szene und eines im Bild gemessenen signifikanten Detektionsmerkmals (Unterkante des Fahrzeugs zur Fahrbahn) eine Instanziierung als seitlich, frontal oder schräg gesehenen Fahrzeugs. Ausgehend von dem signifikanten Merkmal werden weitere, benachbarte Merkmale gesucht (Seitenkanten). Die Skalierung und Lageinitialisierung erfolgt über die Annahme, daß sich die gefundene Unterkante auf der Fahrbahnebene befindet. Aus deren bekanntem Verlauf oder über eine Inversion der Abbildungsgleichungen unter der Annahme einer ebenen Fahrbahn, wird wie bei [Thomanek 96] eine erste Lagehypothese aufgestellt.

Bei der Instanziierung von Objekthypothesen wird stets hierarchisch vorgegangen, d. h. es wird stets zunächst eine Hypothese für eine Quaderform aufgestellt, die dann bei Bedarf durch Aufsetzen weiterer Hypothesen (z. B. Fahrzeug-Polyedermodell) verfeinert werden können. Das Vorgehen bei der videobasierten Objektinstanziierung ist in Bild 9.12 zusammenfassend dargestellt.

# Kapitel 10

## Szeneninterpretation

Unter dem Begriff Szeneninterpretation wird die Informationsgewinnung verstanden, welche aus der Nutzung von Hintergrundwissen über die Szene, die Domäne und die relationalen Beziehungen zwischen Objekten in der Szene zu neuen Schlußfolgerungen führt. Beispiele sind etwa die Annahme, daß sich das Eigenfahrzeug auf einer Fahrbahn befindet und sich im Fahrbahnbereich nur bestimmte Objektklassen aufhalten können, während sich andere Objektklassen wiederum außerhalb des Fahrbahnbereichs befinden. Oder aber, daß sich in der Domäne “Autobahn“ weitere Einschränkungen für die möglichen Objektklassen, ihre Bewegungsfreiheitsgrade (z. B. kein Gegenverkehr) und typische Geschwindigkeitsbereiche ergeben. Die Interpretation greift dabei auf Hintergrundwissen auf den verschiedensten Ebenen zurück, das sowohl statisch wie auch dynamisch sein kann ebenso wie es implizit oder explizit repräsentiert sein kann.

Da im Bereich der Szeneninterpretation eine große Vielfalt an Begriffen verwendet wird, und eine allgemeinverbindliche Definition der Begriffe nicht vorliegt, sollen hier zunächst die wesentlichen Begriffe im Zusammenhang mit der Szeneninterpretation definiert werden. Im Mittelpunkt steht bei dieser Definition die Unterscheidung zwischen den Begriffen “Szene“ und “Situation“.

- **Räumlich-zeitliche Szene**

Beschreibung der Gesamtheit aller Wahrnehmungen (Umfeldsensorik und Egosenso-  
rik) durch physikalische Größen. I. a. wird es sich dabei um eine Beschreibung im  
Zustandsraum handeln, ergänzt durch zusätzliche Attribute im Merkmalsraum der  
Sensoren.

- **Symbolische Szene**

Eine durch eine Klassifizierung aus der räumlich-zeitlichen Szene hervorgegangene  
Beschreibung anhand von symbolischen Bedeutungen (z. B. Fahrzeug, PKW, LKW,  
Verkehrszeichen, Leitpfosten, ...). Bei modellbasierten Verfahren zur Objekterken-  
nung ist diese häufig eng verwoben mit der räumlich-zeitlichen Beschreibung.

- **Situation**

Relationale Beschreibung der Bedeutungen und Rollen von Objekten in der Szene  
unter Verwendung von Verhaltensattributen.

- **Objekte**

Elementare Bestandteile sowohl von Szenen wie von Situationen. Sie enthalten Attribute, in welchen die Informationen auf den verschiedenen Ebenen gehalten werden, bzw. sind durch Assoziationen miteinander verbunden, welche die wechselseitigen Rollen in Situationen beinhalten. Auf der Verhaltensebene kann weiter unterschieden werden zwischen einfachen Objekten, die nicht über einen “eigenen Willen“ verfügen und Objekten, bei denen es sich eigentlich um Subjekte handelt, deren Handlungsabsichten jedoch nicht bekannt sind.

- **Subjekt**

Das Eigenfahrzeug wird als Subjekt bezeichnet, da es eigene Absichten verfolgt, die je nach Grad der Automatisierung mehr oder weniger vollständig bekannt sind. Bei Fahrerassistenzsystemen kann die Fahrerabsicht zum Teil durch eine Fahrerbeobachtung bestimmt werden.

Im Mittelpunkt der Betrachtungen steht hier die reine Szeneninterpretation sowohl zu einem Zeitpunkt als auch anhand zeitlicher Attribute. Die Bewertung der Situation und die damit verbundene Unterscheidung von Subjekt und Objekt sollen allein der Abgrenzung dienen und sind Gegenstand vielfältiger anderer Arbeiten.

In den nächsten Abschnitten wird zunächst die statische Struktur der Szeneninterpretation beschrieben. Da das Wissen über die Domäne, in welcher man sich gerade befindet, eine wichtige Rolle bei der Interpretation spielt, folgt dann die Skizzierung des Vorgehens, um die Domäne aus Szenendaten heraus zu ermitteln. Im Anschluß wird dann auf das “Referenz“-Szenario eingegangen, welches dem Systementwurf dieser Arbeit zugrundeliegt. Die Darstellung der zeitlichen Abläufe konkretisiert auf der Interpretationsebene das Vorgehen zur Szenenerkennung, wie es sich aus den Bausteinen - vornehmlich des 9. Kapitels zur Objekterkennung - ergibt. Daran anknüpfend werden die Methoden zur Hypothesenbewertung auf Szenenebene erläutert, ehe mit dem Aspekt der Aufmerksamkeitssteuerung geschlossen wird.

## 10.1 Wissensrepräsentation und Regelbasis

Das bei der Szenenrepräsentation benutzte Wissen ist im wesentlichen regelbasiert implizit und explizit in Form von Bedingungen, Parameterwertebereichen oder Prozeduren im System hinterlegt. Eine Wissenshinterlegung in allgemeinerer Form, z. B. in Form expliziter Datenbanken, würde den möglichen Rahmen im Hinblick auf die Echtzeitanforderungen des Systems sprengen und beim derzeitigen Stand der Szeneninterpretation über das Ziel hinausschiessen. Gleichwohl ist das Wissen in verschiedener Hinsicht strukturiert. Wie aus Bild 10.1 zu erkennen ist, gibt es einerseits eine hierarchische Gliederung der Wissensrepräsentationen:

- **Merkmalsebene:** Hier ist das Wissen über die Sensoren, die Meßverfahren und die gemessenen Merkmale und deren Unsicherheiten hinterlegt. Wissen über zeitliche Vorgänge ist auf dieser Ebene nicht enthalten.

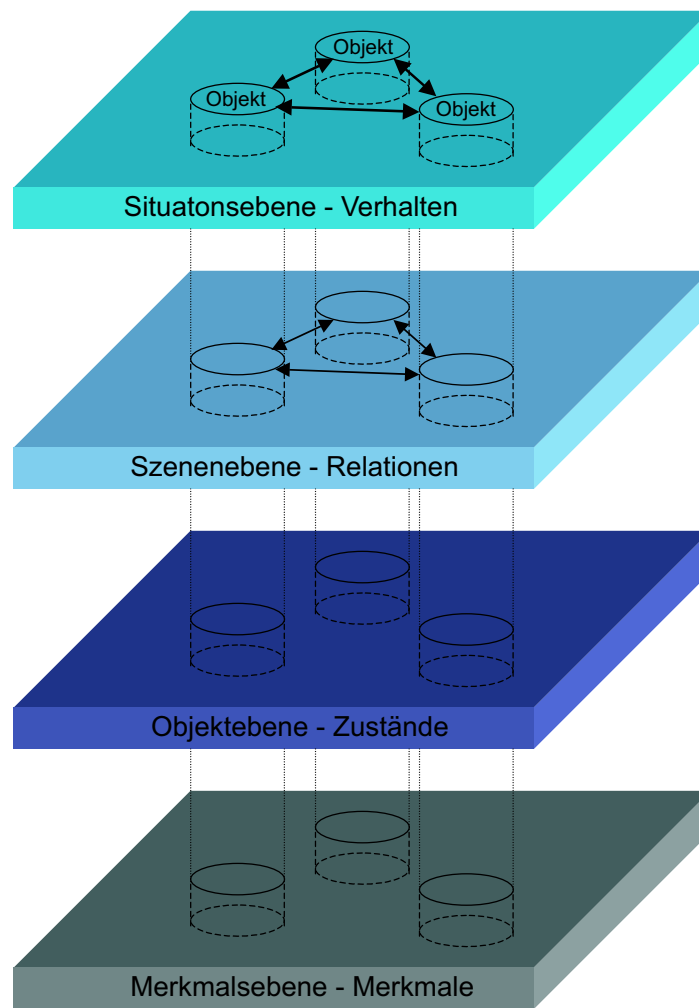


Abbildung 10.1: Repräsentation von Wissen auf mehreren Ebenen

- **Objektebene:** Die Objektebene beinhaltet das Wissen über die Objekte als Datenbasis der Bewegungs- und Formmodelle, die dem System bekannt sind. Hierzu zählen auch bestimmte Wertebereiche von Parametern, um aus einem Form- bzw. Bewegungsmodell verschiedene Objektklassen abzuleiten (z. B. PKW, LKW, ...). Die Verbindung mit der Merkmalsebene erfolgt über die in den Formmodellen enthaltenen Objektmerkmale, die es erlauben, anhand dedizierter Messungen gezielt nach bestimmten Objektklassen zu suchen, bzw. umgekehrt anhand eines gemessenen Merkmals mögliche Kandidaten für eine zu instanziiierende Objekthypothese zu ermitteln. Zusätzlich ist für alle Objektklassen zeitliches Wissen zur Beurteilung der Objekthypothese über der Zeit enthalten.
- **Szenenebene:** Auf der Szenenebene ist das Wissen repräsentiert, welches in Verbindung mit der Wechselwirkung der Objekte genutzt wird. Hierzu zählen die Konsistenz der Szene oder auch die Einschränkung von Objektklassen auf bestimmte Szenenbereiche (z. B. Fahrzeuge auf Straßen). Ganz wesentlich ist auf dieser Ebene auch das Prozeßwissen darüber, anhand welcher Erkennungsabläufe oder Erkennungspläne die Szeneninterpretation erfolgt.

- **Situationsebene:** Auf der höchsten, der Situationsebene werden die Verhaltensaspekte der Objekte und ihre gegenseitigen Rollen repräsentiert. Hierbei handelt es sich einerseits um die Erkennung von Objektmanövern als auch die Ableitung zukünftiger Handlungen von Objekten aus der Vergangenheit in Verbindung mit dem jeweiligen Situationskontext.

Wie in Bild 10.1 zu erkennen ist, werden Aspekte der verschiedenen Ebenen in Form von Attributen in den Objekten gehalten. Darüberhinaus gibt es aber auf jeder Ebene auch eigene Strukturen, welche von den Objekten unabhängige, übergeordnete Informationen halten.

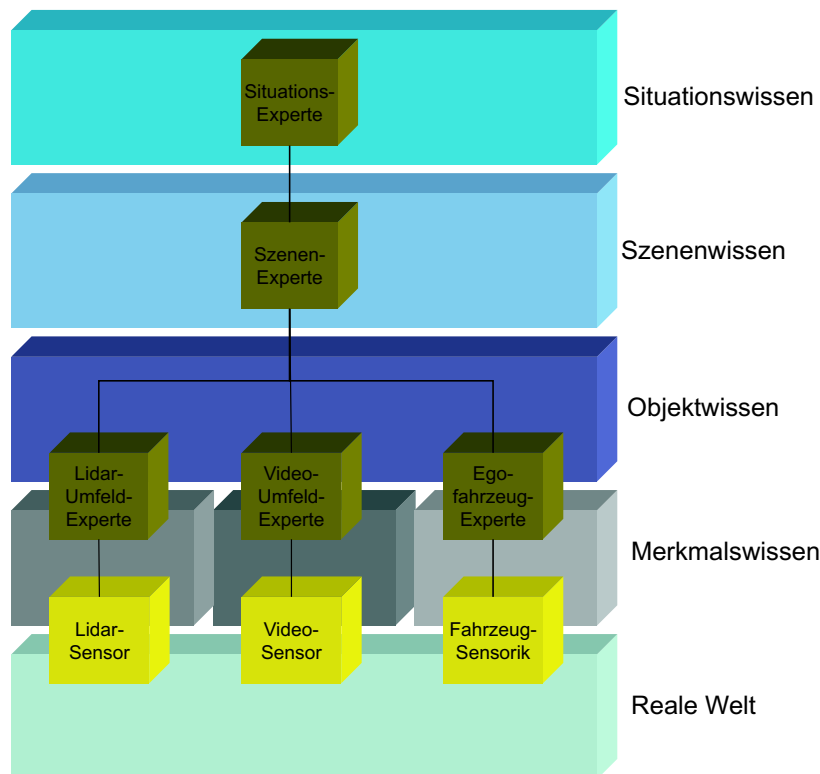


Abbildung 10.2: Expertennetzwerk und unterlagerte Wissens Ebenen

Einen weiteren Punkt bildet die in Bild 10.1 nicht enthaltene Unterscheidung in statisches und dynamisches Wissen:

- **Statisches Wissen** ist das Wissen, welches allgemein als Hintergrundwissen über die o. g. Aspekte im System hinterlegt ist. Es beinhaltet allgemeingültige Aussagen beispielsweise über Objektklassen und nicht über konkrete Objekte. Dieses Wissen ist in der überwiegenden Zahl der Systeme zur visuellen Szeneninterpretation von Experten fest im System hinterlegt. Prinzipiell kann man für das Hintergrundwissen auch eine Lernkomponente vorsehen und so vom statischen zum quasistatischen Hintergrundwissen übergehen. Diese Lernkomponente betrifft in den meisten Fällen jedoch nur Teilaspekte des Systems. In dem in dieser Arbeit betrachteten System ist das Hintergrundwissen rein statisch.

- Unter **Dynamischem Wissen** werden alle Aspekte des Erkennens einer Szene zur Laufzeit bezeichnet. Hierzu gehören folglich die Repräsentation der Szene mit den konkreten in der Szene enthaltenen Objekten und deren Attributen. Dieses Wissens ist ebenfalls hierarchisch in Form des Szenenbaums organisiert (Kapitel 6.3).

In Analogie zur hierarchischen Repräsentation des Wissens ist auch der eigentliche Interpretations-Prozeß hierarchisch gegliedert. Die in Abbildung 10.2 dargestellte Struktur basiert auf einem baumartig organisierten Netz von sogenannten “Experten“. Im wesentlichen gibt es zwei Kategorien von Experten: Die einen korrespondieren direkt mit einer Repräsentationsebene, während andere an der Schnittstelle zwischen zwei Ebenen angesiedelt sind. Die zweite Kategorie betrifft die als “Umfeldexperten“ bezeichneten Module, welche für jeden Sensor separat unter Rückgriff auf das allgemeine, sensorübergreifende Objektwissen und das sensorspezifische Merkmalswissen organisiert sind. Die Experten können jeder wiederum hierarchisch in Form weiterer Experten organisiert sein. So sind der “Video-Umfeldexperte“ entsprechend Abbildung 10.3 und der in 10.4 dargestellte “Lidar-Umfeldexperte“ in weitere Experten zur Erkennung bestimmter Objektklassen unterstrukturiert. Die in den Abbildungen 10.3 und 10.4 neben den Experten für die bekannten

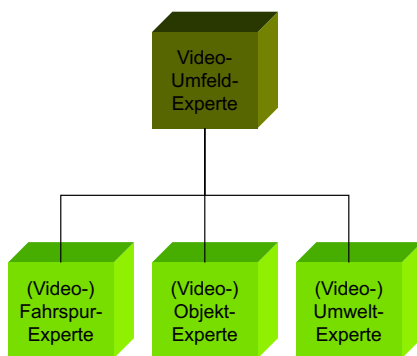


Abbildung 10.3: Video-Umfeldexperte

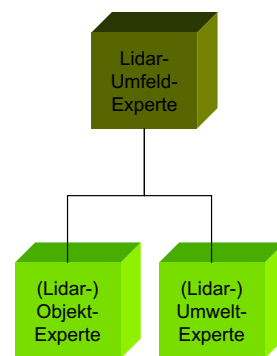


Abbildung 10.4: Lidar-Umfeldexperte

Aufgaben der Fahrspur- und Objekterkennung enthaltenen “Umweltexperten“ haben die Aufgabe, die Umgebungsbedingungen wie Tag - Nacht, Sonne - Regen - Schnee etc. aus den Sensordaten zu bestimmen.

## 10.2 Domäne und Einsatzzweck

Das in dieser Arbeit vorgestellte System ist allgemein auf die Erkennung von Straßenverkehrsszenen angelegt. Diese werden mittels fahrzeugfester Sensoren aus der Perspektive des Eigenfahrzeugs wahrgenommen. Als relevante Objekte für die Wahrnehmung werden in erster Linie die Fahrbahn sowie andere Fahrzeuge im Fahrbahnbereich angesehen. Da ferner die Sensoren zunächst den vor dem Eigenfahrzeug befindlichen Frontbereich erfassen, ist eine solche Sensorik auch primär auf die Aufgabe der Abstandsregelung und Spurhaltung in Szenen außerhalb von Ortschaften ausgerichtet. Eine solche Einteilung von Szenen- und Aufgabenbereichen wird mit dem Begriff “Domäne“ bezeichnet. Im genannten Bereich außerhalb von Ortschaften lassen sich mehrere Domänen unterscheiden. Die Kriterien zur

Abgrenzung zwischen den Domänen können aber je nach Ziel der Kategorisierung unterschiedlich sein. So wird eine Funktion zur Abstandsregelung möglicherweise die Domänen “Fließender Verkehr” und “Staufahrt” voneinander unterscheiden wollen, weil dies für die Algorithmen zur Situationsbewertung, Verhaltensentscheidung und Regelung von Bedeutung ist.

Für die Wahrnehmung, insbesondere die modellbasierte Wahrnehmung, sind im Gegensatz zu diesem Beispiel jedoch andere Domäneneinteilungen relevant, nämlich in erster Linie solche, die eine Einschränkung des Interpretationsraums durch Vorannahmen erlauben. So findet man üblicherweise auf Autobahnen getrennte Fahrbahnen für jede Fahrtrichtung und die Auftretenswahrscheinlichkeit bestimmter Objektklassen (z. B. Menschen, Radfahrer) ist in Normalsituationen gering. Wesentlich in diesem Zusammenhang ist der Begriff *Auftretenswahrscheinlichkeit*, der impliziert, daß alle Schlüsse, welche aus der Bestimmung der Domäne gezogen werden, stets unter dem Vorbehalt der Sicherheit stehen, mit der die Bestimmung der Domäne selbst erfolgt ist.

Im Rahmen dieser Arbeit wurde ein sehr einfaches Regelsystem entwickelt, mit dessen Hilfe es möglich ist, basierend auf Eigenfahrzeugdaten und ergänzt durch Umfelddaten eine Abschätzung der Domäne vorzunehmen. Die Ergebnisse der Domänenklassifikation gehen im wesentlichen bei der Instanziierung neuer Hypothesen in die Betrachtungen ein. So werden Hypothesen für andere Fahrzeuge auf Autobahnen nur in Wertebereichen für die Geschwindigkeit initialisiert, die dem Mitverkehr entsprechen, während in anderen Szenarien abhängig von der Fahrspurzuordnung auch Gegenverkehr zugelassen wird. In ähnlicher Weise wird in der Domäne Autobahn im Gegensatz zu den anderen Domänen kein Querverkehr auf gleicher Fahrbahnebene erwartet.

Das Feld der Domänenklassifikation aus Eigenfahrzeug- und Umfelddaten wird mit diesem Ansatz sowohl was die Klassifikation betrifft als auch was die daraus abgeleiteten Vorannahmen betrifft nur am Rande gestreift. Gerade im Hinblick auf die Sicherheit der Klassifikation wie der statistischen Gültigkeit der Vorannahmen werden noch ausführliche Untersuchungen folgen müssen.

Merkmal	Domäne		
	Autobahn	Landstraße	Ortschaft
Eigengeschwindigkeit	> 80 km/h	60 – 120 km/h	< 60 km/h
Fremdgeschwindigkeit	> 80 km/h	60 – 120 km/h	< 60 km/h
Eigenfahrspur	beliebig	rechte Spur	beliebig
Fremdfahrspur	beliebig	rechte Spur	beliebig
Gegenverkehr	nicht vorhanden	vorhanden	vorhanden
Fahrbahnkrümmung	sehr gering	beliebig	beliebig

Tabelle 10.1: Merkmale zur Domänenklassifikation

## 10.3 Referenzszenario

Die Systemauslegung in Bezug auf die Auswahl der Sensoren wurde unter dem Gesichtspunkt der Beherrschung eines typischen “Referenz“-Szenarios getroffen, welches im folgenden beschrieben wird. Abbildung 10.3 zeigt hierzu die typische Abfolge eines Szenarios im Längsverkehr, das sich aus insgesamt 7 Phasen zusammensetzt. Die Phasen sind gekenn-

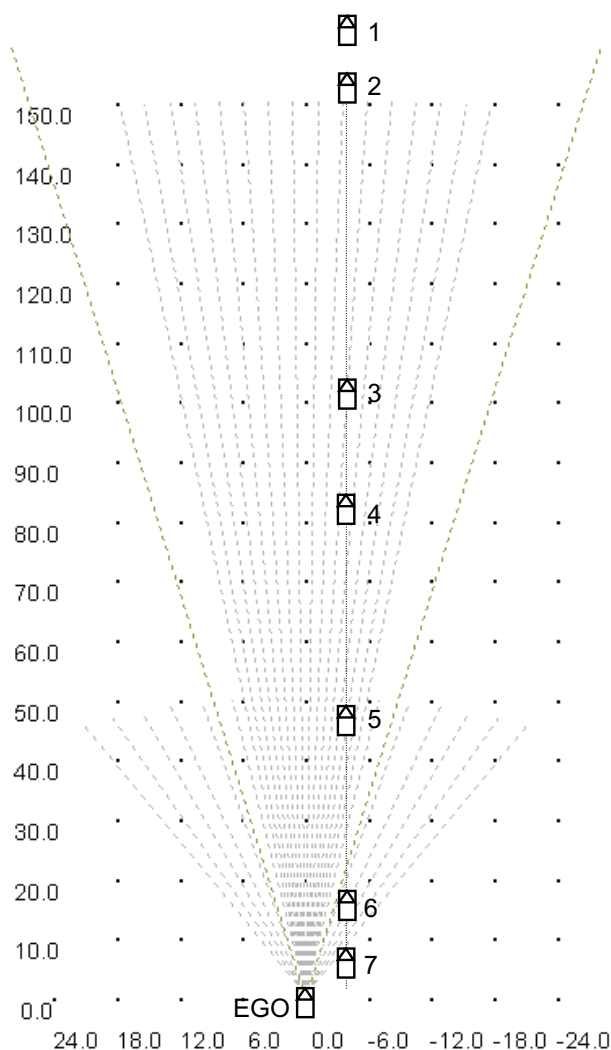


Abbildung 10.5: Darstellung eines Referenzszenarios in der Draufsicht

zeichnet durch die unterschiedlichen “verallgemeinerten“ Aspektbedingungen unter denen ein Objekt vom Eigenfahrzeug (EGO) und seinen Sensoren aus gesehen wird. Unter “verallgemeinerten“ Aspektbedingungen wird in diesem Zusammenhang jede Änderung der Szenenkonstellation verstanden, die zu einem geänderten Wahrnehmungs“aspekt“ für einen der Sensoren führt. D. h., wenn das Objekt erst von einem, dann von zwei Laserstrahlen in der Heckansicht gesehen wird, ist dies ebenso eine Änderung der Aspektbedingungen wie ein Wechsel von der Heck- in die Seitenansicht für den Videosensor. Das in Abb. 10.3 dargestellte Szenario könnte beispielsweise ein Überholszenario auf der Autobahn sein, oder eine



Vorbeifahrt an einem stehenden Hindernis. Dabei sind die Abmessungen des in der Skizze dargestellten Objekts so gewählt, daß sie ungefähr einem Mittelklasse-PKW entsprechen.

Die einzelnen Phasen des Szenarios teilen sich wie folgt auf:

1. Phase: Das Eigenfahrzeug nähert sich dem Objekt von hinten an, das Objekt befindet sich noch außerhalb beider Sensorerfassungsbereiche.
2. Phase: Das Objekt tritt bei ca. 150 m Entfernung in den zentralen Erfassungsbereich des Mehrstrahl-Lidars ein und wird dort zunächst von nur einem Laserstrahl erfaßt. Damit kann die Position des Objekts bestimmt werden, eine Schätzung der Abmessungen und der Winkellage ist nicht möglich.
3. Phase: Bei weiterer Annäherung wird das Objekt ab spätestens 100 m Distanz von 2 Lidarstrahlen erfaßt. Nun kann erstmals auch eine erste grobe Näherung der Abmessungen einer Dimension gegeben werden.
4. Phase: Bei Erreichen der 80 m Marke erreicht das Objekt eine Größe im Videobild, die auch in diesem eine Vermessung ermöglicht. Damit können jetzt sowohl Breite und Höhe als auch die Ablage mit größerer Genauigkeit sowie redundant bestimmt werden.
5. Phase: Bei ca. 40-50 m Entfernung ist die geringste Entfernung erreicht, bei der ein Fahrzeug in der Nebenspur sich gerade noch im fein aufgelösten zentralen Erfassungsbereich des Lidarsensors und im Erfassungsbereich der Videokamera befindet. An dieser Position ist damit i. a. die höchstmögliche Güte der Erfassung gegeben. Unterhalb von 30 m wird es dann nur noch von den peripheren Strahlen des Lidarsensors und der Videokamera gesehen. (Wenn das Fahrzeug sich nicht in der Nebenspur, sondern in der eigenen Fahrspur aufhielte, bliebe es natürlich bis kurz vor dem Eigenfahrzeug im zentralen Erfassungsbereich des Lidarsensors. Dieses ist ein wichtiger Aspekt hinsichtlich Sicherheitssystemen zur Kollisionsfolgenminderung bzw. Kollisionsvermeidung.)
6. Phase: Bei ungefähr 20 m verläßt das Objekt auch den Videoerfassungsbereich und wird dann nur noch von den peripheren Strahlen des Lidars erfaßt. Da es im Heckbereich meist nur noch von einem Strahl erfaßt wird, ist eine Vermessung der Breite dann nicht mehr möglich.
7. Phase: Das Objekt verläßt bei ungefähr 8 m auch den peripheren Lidarerfassungsbereich und tritt in den "toten" Winkel der Umfelderkennung ein.

Das gesamte Szenario kann man sich natürlich auch "rückwärtsgespielt" vorstellen: Dann tritt ein Objekt zunächst von hinten kommend in den peripheren Erfassungsbereich (vorzugsweise von links) des Lidarsensors ein und entfernt sich dann in der Folge vom Eigenfahrzeug.

In ähnlicher Weise wie dieses Längsverkehrsszenario hätte man auch ein Querverkehrsszenario darstellen können. Allerdings gibt es für diese Klasse von Szenarien deutlich mehr

individuelle Einschränkungen, da typische Querverkehrsszenarien häufig von Verdeckungen großer Teile der Szene durch Randbebauung oder andere Objekte in Kreuzungsszenarien geprägt sind und demzufolge Aspekte wie die weite Vorausschau hier weniger relevant sind.

Anhand des geschilderten Szenarios soll in der Hauptsache der wesentliche Aspekt der gewählten Sensoranordnung deutlich werden. Sowohl bei der Annäherung an entfernte Objekte wie für Objekte in der Vorbeifahrt wird eine Grob-zu-Fein-Strategie verfolgt. Bei Eintritt in den Erfassungsbereich kann zunächst nur die Aussage getroffen werden “Da ist Etwas !“ in Verbindung mit einer groben Positionsschätzung. Bei Annäherung an den zentralen Erfassungsbereich, welcher sich im wesentlichen mit dem Hauptbewegungsfreiraum des Eigenfahrzeugs überdeckt, wird diese Information zunehmend besser aufgelöst. Zusätzlich erfolgt im zentralen Winkelerfassungsbereich bis zu einer Entfernung von 80 m eine teilredundante Erfassung durch zwei Sensoren unterschiedlicher Meßprinzipien. Die Wahl der beiden Sensorprinzipien stützt diese Herangehensweise, da mit dem Lidarsensor der periphere Bereich und der Fernbereich gut abdeckt sind und die Entfernungsmeßgenauigkeit in allen Entfernungen identisch ist, während der Videosensor im wichtigen zentralen Erfassungsbereich nicht nur die Objekte redundant erfaßt und zum Teil klassifizieren kann, sondern darüberhinaus auch die Fahrspur als zentrales Objekt zur Szeneninterpretation bereitstellt.

Die Sensoranordnung läßt damit deutliche Parallelen zur Vorgehensweise der menschlichen Wahrnehmung (Kapitel 3) erkennen, die ebenfalls eine Grob-zu-Fein-Strategie verfolgt. Im Unterschied zum Menschen müssen aufgrund derzeitiger technischer Randbedingungen allerdings mehrere unterschiedliche Sensorprinzipien miteinander kombiniert werden, um diesen Effekt “seriennah“ zu erzielen. Erst der Einsatz eines Systems aus mehreren Videokameras unterschiedlicher Brennweite in Kombination mit einer aktiven Blickrichtungssteuerung vermag einen ähnlichen Ansatz basierend auf nur einem - dem Menschen analogen - Sensorprinzip zu verfolgen.

Die Tragfähigkeit der hier getroffenen Sensorauswahl unter den gegebenen Szenarien wird anhand der experimentellen Beispielszenarien in Kapitel 12 deutlich werden.

## 10.4 Zeitlicher Ablauf

Der in diesem Abschnitt dargestellte zeitliche Ablauf der Szeneninterpretation ist eng verwandt mit den in Abschnitt 9.1 enthaltenen Datenflüssen der Sensordatenfusion (Abbildungen 9.1 und 9.2). Im Unterschied zu diesen Darstellungen, in denen es um den grundsätzlichen Ablauf der Informationsfusion auf einer sensorunabhängigen Ebene ging, soll der Datenfluß hier konkreter für die beiden in dieser Arbeit eingesetzten Umfeldsensoren beschrieben werden. Im folgenden wird zunächst der Informationsfluß für die Szeneninterpretation mittels Videosensorik beschrieben; im Anschluß daran folgt dann die Darstellung für den Lidarsensor.

## Videosensor

Ausgehend von einer aktuellen Szenenbeschreibung, etwa in Form eines Szenenbaums, erfolgt, wie in Abbildung 10.6 dargestellt, bei Eintreffen neuer Sensorinformationen zunächst die Prädiktion der bereits instanziierten Objekte (Objekte-4D). Nach der Prädiktion werden diese ebenso wie die ebenfalls aus der Szene abgeleiteten Detektionsbereiche (ROI-2D) in den Merkmalsraum, in diesem Fall also die Bildebene transformiert. Es folgen daraus die „Tracking“-Merkmale für die Objekte respektive die „Detektions“-Merkmale für die „ROI's“. Als Ergebnis der sich anschließenden Vermessung ergeben sich gemessene „Merkmalprimitive“, die in einem Aggregationsschritt zu aggregierten, gemessenen „Tracking“-bzw. „Detektions“-Merkmale werden. Die aggregierten gemessenen „Tracking-Merkmale“ werden - soweit möglich - mit den erwarteten „Tracking“-Merkmale assoziiert, räumlich plausibilisiert und im „Innovations“-Schritt zur Aktualisierung der Objektzustände herangezogen. Die innovierten Objektzustände werden einem zeitlichen Plausibilitätstest unter-

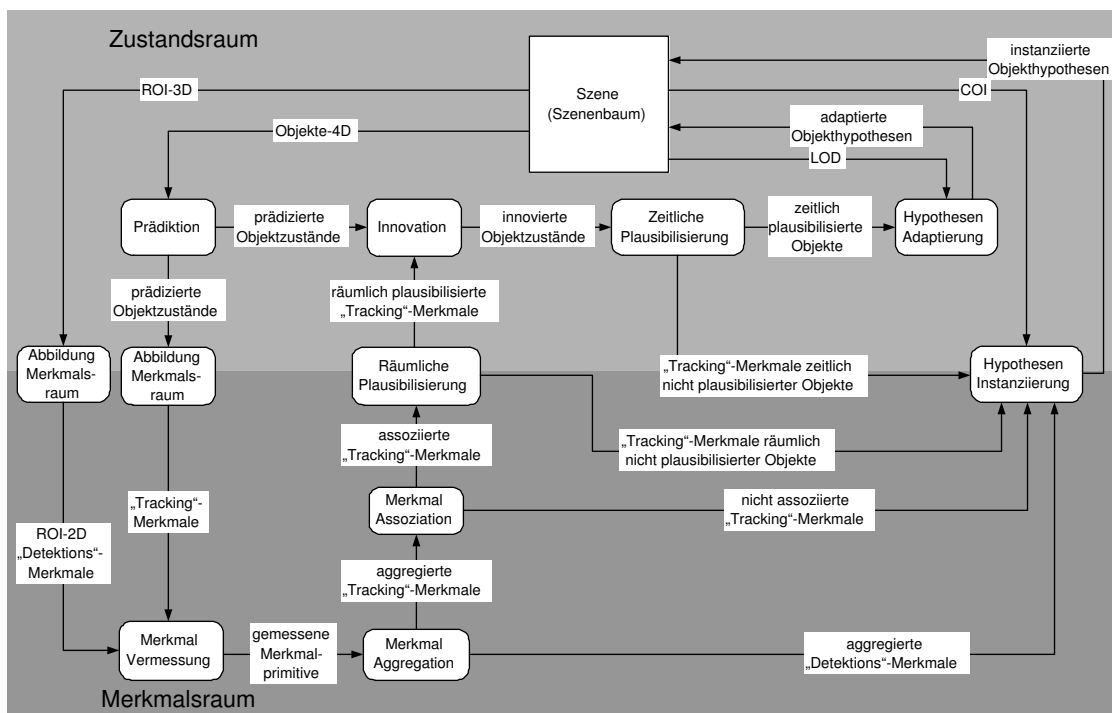


Abbildung 10.6: Ablaufdiagramm der Umfeldwahrnehmung mit Videosensor

zogen und im Rahmen der Hypothesen-Adaptierung im Rahmen ihrer Rolle in der Szene, ihrer aktuellen Wahrnehmungsgüte und eventueller Vorgaben entsprechend der Aufmerksamkeitssteuerung beispielsweise in bezug auf den gewünschten Level-Of-Detail (LOD) angepaßt. Wird im Rahmen der Plausibilisierungsschritte eine Objekthypothese verworfen, so werden die mit dieser Hypothese assoziierten Meß-Merkmale wieder verfügbar und gehen zusammen mit den nicht assoziierten Meß-Merkmalen sowie den aggregierten, gemessenen „Detektions“-Merkmale in die Instanziierung neuer Objekthypothesen ein. Dieser Instanziierungsschritt der modellgetriebenen Bildverarbeitung wird gesteuert durch die in der gegebenen Szene oder bestimmten Szenenbereichen erwarteten Objektklassen „Classes-Of-Interest“ (COI). So werden in Autobahnsszenarien andere Objekte erwartet als in der

Stadt, ebenso wie im Fahrbahnbereich Objekthypothesen, welche einen Verkehrsteilnehmer darstellen gegenüber Hypothesen, die Randobjekte repräsentieren, bevorzugt werden.

## Lidarsensor

Der wesentlichste Unterschied zum Datenfluß bei der Videosensorik liegt bzgl. des Lidarsensors in der nicht erwartungsgesteuerten Vermessung. Der Lidarsensor kann zwar, wie in Abbildung 10.7 zu sehen, durch die Auswahl der auszuwertenden Strahlen sowie die Vorgabe einer Maximalentfernung gesteuert werden, die einzelnen Merkmale, die er liefert, sind jedoch stets Entfernungspunkte. Daher entfallen hier auch sämtliche Verarbeitungswege,

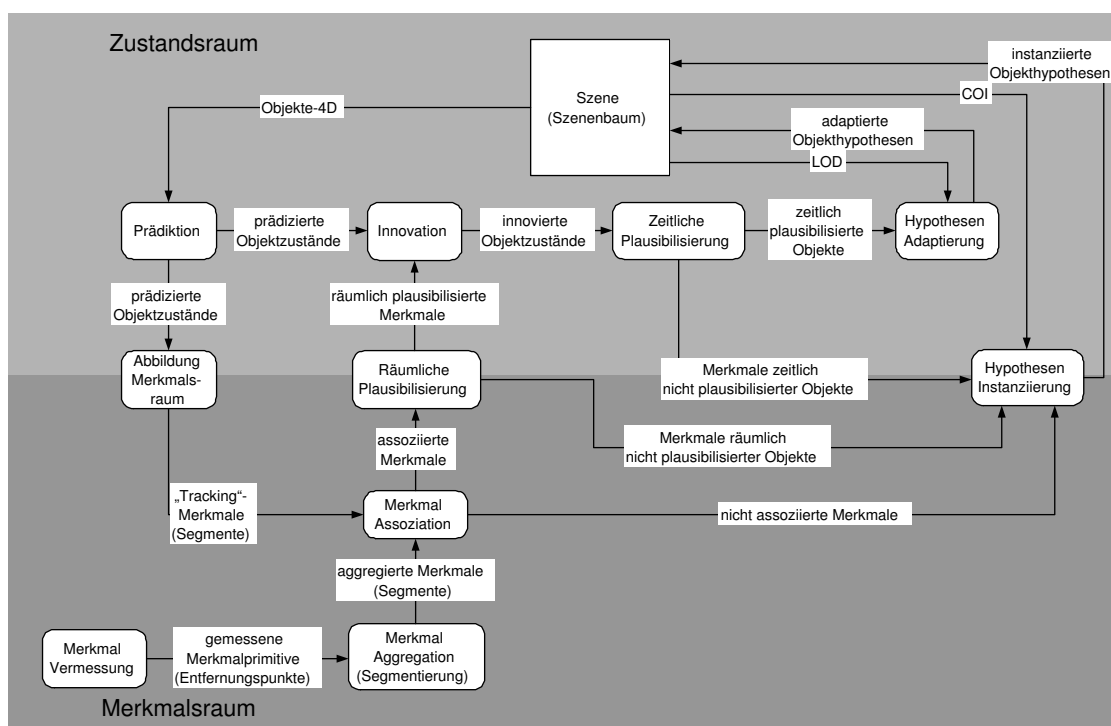


Abbildung 10.7: Ablaufdiagramm der Umfeldwahrnehmung mit Lidarsensor

welche die sog. „Detektions“-Merkmale beim Videosensor betreffen. Stattdessen erfolgt zunächst eine rein datengetriebene Segmentierung der Entfernungsdaten. Anschließend werden die so erhaltenen Segmente mit den aus der Prädiktion hervorgegangenen erwarteten Segmenten assoziiert. Mit den so zugeordneten Merkmalen wird im weiteren Verlauf wie beim Videosensor erläutert verfahren. Die nicht assoziierten Segmente dienen sodann zur Instanziierung neuer Hypothesen in Analogie zum Videosensor. Die innovierten Objektzustände werden einem zeitlichen Plausibilitätstest unterzogen und im Rahmen der Hypothesen-Adaptierung, ihrer Rolle in der Szene, ihrer aktuellen Wahrnehmungsgüte und eventueller Vorgaben entsprechend der Aufmerksamkeitssteuerung - beispielsweise in bezug auf den gewünschten Level-Of-Detail (LOD) - angepaßt. Wird im Rahmen der Plausibilisierungsschritte eine Objekthypothese verworfen, so werden die mit dieser Hypothese assoziierten Meß-Merkmale wieder verfügbar und gehen zusammen mit den nicht assoziierten

Meß-Merkmalen in die Instanziierung neuer Objekthypothesen ein. Dieser Instanziierungsschritt greift auf das gleiche Hintergrundwissen über die Szene und die Domäne zu, wie dies beim Videosensor der Fall ist.

## 10.5 Hypothesenbewertung

Die Bewertung einer Objekthypothese erfolgt stets in mehreren Stufen:

1. Räumliche Plausibilisierung auf der Merkmalsebene:

Hier wird die räumliche Konsistenz einer Merkmalgruppierung (z. B. Kantenmuster bei der Bildverarbeitung, Segmente beim Mehrstrahl-Lidar) sowie der Abstand der Gruppierung oder einzelner Meßwerte von der Erwartung getestet. Bei Unterschreiten eines gegebenen Gütemaßes wird die Gruppierung verworfen. In einem gegebenen Meßzyklus erfolgt somit keine Aktualisierung eines Objekts anhand neuer Meßdaten, stattdessen wird die Prädiktion fortgeschrieben.

2. Zeitliche Plausibilisierung auf Objektebene (im Zustandsraum):

Jede Objekthypothese wird nicht aufgrund einer einmaligen Messung als gültig akzeptiert bzw. bei Nichtvorliegen eines Meßwerts (z. B. aufgrund fehlender räumlicher Plausibilität) sofort verworfen. Vielmehr erfolgt eine zeitliche Beurteilung der Güte einer Hypothese über mehrere Meßzyklen in Kombination mit absoluten Zeitfenstern.

Die Gesichtspunkte zur räumlichen Plausibilisierung waren schon Gegenstand der Betrachtungen in den Abschnitten zur Assoziation in Kapitel 9. An dieser Stelle soll der übergeordnete Aspekt der zeitlichen Plausibilisierung im Mittelpunkt stehen. Auch bei dieser zeitlichen Plausibilisierung wird unterschieden zwischen der Fahrspur auf der einen und Objekten auf der anderen Seite.

Anmerkung: Die Festlegung einer Strategie zur Plausibilisierung ist im wesentlichen von den Funktionen abhängig, die auf der Umfeldwahrnehmung aufsetzen. Eine Sicherheitsfunktion wie die Automatische Notbremse (ANB) wird andere Anforderungen an die Plausibilität von Hypothesen haben als eine Komfortfunktion wie ACC. In diesen Fällen dient der Leistungsumfang der Funktion als Beurteilungskriterium zur Festlegung der Plausibilisierungsstrategie. Da hier keine konkrete Anwendung betrachtet wird, wurden die im folgenden aufgeführten Strategien empirisch auf der Basis ausgewählter Testdatensätze bestimmt.

### Zeitliche Plausibilisierung der Fahrspur

Die zeitliche Plausibilisierung der Fahrspur erfolgt in einem mehrstufigen Verfahren. Ausgehend von den räumlichen Gütemaßen der linken und rechten Fahrspurbegrenzung wird

in einem ersten Schritt ein zeitliches Plausibilitätsmaß separat für beide Begrenzungslinien bestimmt. Diese müssen für sich genommen bestimmte Kriterien zur zeitlichen Plausibilisierung erfüllen. In einem zweiten Schritt wird eine Gesamtplausibilität für die Fahrspurhypothese aus den beiden zuvor bestimmten Werten der beiden Begrenzungslinien ermittelt.

Weiter wird bei der Bestimmung der Plausibilitäten unterschieden zwischen einer Initialisierungsphase nach der Detektion und dem Aufsetzen der Fahrspurhypothese, sowie der anschließenden Trackingphase. In der Initialisierungsphase werden erhöhte Anforderungen an die Plausibilität gestellt, um die Hypothese kritisch zu testen. Bei einer einmal verifizierten Hypothese wird in der Trackingphase dann auch ein zwischenzeitlich weitergehendes Absinken der Plausibilität toleriert, um kurzzeitige Meßstörungen ausgleichen zu können.

Im einzelnen gestaltet sich die Plausibilisierung der Fahrspurbegrenzungen wie folgt:

- Initialisierungs-Phase:

$$\begin{array}{llll} 0.5 & \leq & \frac{N_{ASO}}{N_{MSR}} & \leq 1.0 \Rightarrow Pl_{L/RM}(t) = Pl_{L/RM}(t-1) + 0.05 \\ 0 & \leq & \frac{N_{ASO}}{N_{MSR}} & < 0.5 \Rightarrow Pl_{L/RM}(t) = Pl_{L/RM}(t-1) - 0.05 \end{array}$$

- Tracking-Phase:

$$\begin{array}{llll} 0.5 & \leq & \frac{N_{ASO}}{N_{MSR}} & \leq 1.0 \Rightarrow Pl_{L/RM}(t) = Pl_{L/RM}(t-1) + 0.05 \\ 0.25 & \leq & \frac{N_{ASO}}{N_{MSR}} & < 0.5 \Rightarrow Pl_{L/RM}(t) = Pl_{L/RM}(t-1) \\ 0 & \leq & \frac{N_{ASO}}{N_{MSR}} & < 0.25 \Rightarrow Pl_{L/RM}(t) = Pl_{L/RM}(t-1) - 0.05 \end{array}$$

mit

$N_{ASO}$	Anzahl zugeordneter Merkmale der linken bzw. rechten Fahrspurbegrenzung
$N_{MSR}$	Anzahl gemessener Merkmale der linken bzw. rechten Fahrspurbegrenzung
$Pl_{L/RM}$	Plausibilität der linken bzw. rechten Fahrspurbegrenzung.

Im Anschluß daran wird das Gesamtplausibilitätsmaß für die Fahrspur gebildet:

$$Pl_{FS} = 1/2 (Pl_{LM} + Pl_{RM})$$

mit

$Pl_{FS}$  Plausibilität der Fahrspur aus linker und rechter Begrenzung

Die Bewertung der Hypothesen erfolgt anhand eines Satzes an Regeln, welcher zum teilweisen oder vollständigen Verwerfen der Hypothese führen kann:

1. Innerhalb der ersten 5 Meßzyklen müssen beide Begrenzungslinien ohne Unterbrechung mit einer räumlichen Plausibilität von 0.5 bestätigt worden sein, damit die Hypothese nicht verworfen wird.

2. Wenn die Plausibilität  $Pl_{L/RM}$  einer Begrenzungslinie auf 0 sinkt, wird die Hypothese für diese Begrenzungslinie zurückgesetzt, d. h. sämtliche damit verbundenen Merkmalsattribute werden in ihren Initialzustand versetzt. Zusätzlich werden alle nur bei Vorhandensein beider Begrenzungslinien beobachtbaren Zustandsgrößen neu initialisiert. Dies sind insbesondere die Breite  $b$  sowie der Nickwinkel  $\theta$  und deren zugehörige Systemkovarianzen.
3. Wenn zusätzlich zum vorgenannten Kriterium oder auch unabhängig davon die Gesamtplausibilität  $Pl_{FS}$  der Fahrspur den Wert von 0.25 unterschreitet, wird die Gesamthypothese neu initialisiert, d. h. die Attribute beider Begrenzungslinien werden zurückgesetzt, ebenso wie sämtliche Zustandsgrößen der Fahrspur und des relativen Eigenzustands und deren zugehörige Systemkovarianzen.

## Zeitliche Plausibilisierung von Objekten

Im Gegensatz zur Fahrspur führt die räumliche Plausibilisierung bei Objekten nur zu einer Akzeptanz der Merkmalgruppierung als Ganzem oder zum Verwerfen der Gesamtgruppierung. Für die zeitliche Plausibilisierung bildet dieses Vorgehen demzufolge eine rein digitale Entscheidung als Eingangsgröße. Ebenso wie im Fall der Fahrspur wird die zeitliche Plausibilität durch ein abgestuftes Verfahren aus einer Initialisierungs-Phase und einer Tracking-Phase bestimmt.

- Initialisierungs-Phase:

Diese umfaßt die ersten 3 Meßzyklen, nachdem die Objekthypothese gebildet worden ist. Es zählen nur Meßzyklen von Sensoren, in denen das Objekt auch sichtbar und vermeßbar ist. Typischerweise detektiert der Lidarsensor die Objekte, die sich bei ihrer Detektion überwiegend noch nicht im Erfassungsbereich des Videosensors befinden. In diesem Fall zählen also nur die Lidar-Meßzyklen. In dieser Phase wird eine Hypothese verworfen, sobald sie in einem Meßzyklus nicht verifiziert werden konnte, sei es durch nicht erfolgreiche Assoziation oder durch nicht mögliche Erfassung.

- Tracking-Phase:

In dieser Phase wird die Hypothese verworfen, wenn das Objekt innerhalb eines Zeitfensters von 1 s im Mittel in weniger als der Hälfte der Messungen erfolgreich assoziiert werden konnte:

$$\frac{N_{ASO} [t \leq 1 \text{ s}]}{N_{MSR} [t \leq 1 \text{ s}]} < 0.5$$

mit

$N_{ASO}$	Anzahl erfolgreicher Assoziationen
$N_{MSR}$	Anzahl Messungen

## 10.6 Aufmerksamkeitssteuerung

Eine Aufmerksamkeitssteuerung ist immer dann erforderlich, wenn durch eine Begrenzung der verfügbaren Ressourcen nicht alle für das Verhalten relevanten Bereiche bzw. Objekte mit gleich hoher Güte kontinuierlich wahrgenommen werden können. Eine solche Beschränkung kann im wesentlichen zwei Ursachen haben: Beschränkte Rechenleistung auf der einen Seite und den beschränkten Erfassungsbereich der Sensoren auf der anderen Seite. Daraus resultierend gibt es auch zwei unterschiedliche Klassen von Maßnahmen, die sich durch eine Aufmerksamkeitssteuerung beeinflussen lassen. Im folgenden wird zunächst eine Reihe von Maßnahmen, die bei beschränkter Rechenleistung zum Einsatz gelangen können, beschrieben. Im Anschluß daran folgt dann eine kurze Betrachtung von Möglichkeiten bezüglich der Sensorerfassungsbereiche, ehe auf einige mögliche Strategien zur Steuerung der Aufmerksamkeit eingegangen wird.

### 10.6.1 Aufmerksamkeitssteuerung bei beschränkter Rechenleistung

Die Beschränkung der Rechenleistung stellte in der Vergangenheit gerade im Hinblick auf die Bildfolgenverarbeitung eine wesentliche Randbedingung zur Verarbeitung in Echtzeit dar. Heutzutage ist diese Beschränkung insofern gelockert, als die verfügbare Rechenleistung - auch für den Einsatz im Automobilbereich - mittlerweile ausreichend für den Einsatz einfacher Erkennungsansätze ist. Bei der Vielfalt der möglichen Verfahren und dem sicher zunehmenden Bedarf an Wahrnehmungsfähigkeiten wird es aber auf absehbare Zeit stets ein Problem bleiben, alle Ansprüche gleichermaßen zu befriedigen. In Analogie zum biologischen Sehen ließe sich auch argumentieren, daß auch der Mensch sein Umfeld sehr abgestuft aufgelöst - vom fovealen Zentrum zur Peripherie hin - wahrnimmt. Allerdings ist eine der Zielsetzungen gerade im Bereich der Fahrerassistenz die Unterstützung des Menschen in kritischen Situationen, woraus sich die Forderung ableiten läßt, das maschinelle System möge "mehr" sehen als der Mensch.

Möglichkeiten, die erforderliche Rechenleistung zu beschränken, bestehen darin:

#### 1. **Beschränkung des auszuwertenden Erfassungsbereichs (Region-Of-Interest ROI):**

Die Region-Of-Interest (ROI) wird bezogen auf den Videosensor i. a. ein Bildbereich sein, der entweder direkt in 2D gegeben oder aus einem in 3D gegebenen Bereich durch Transformation erhalten wird. Ebenso können aber auch nur Objekte in einem gewissen Entfernungsbereich verfolgt werden. Letzteres Vorgehen ist vor dem Hintergrund des Verlustes der Tiefeninformation durch die perspektivische Abbildung zumindest bei Einsatz monokularer Bildverarbeitungsansätze aber mit Vorsicht zu genießen. Im Hinblick auf den Mehrstrahlaser sind beide Varianten denkbar; es lassen sich sowohl nur bestimmte Strahlen bzw. Strahlbereiche als auch jeder einzelne Strahl nur bis zu einer gewissen Entfernung auswerten. Beide Möglichkeiten führen zu einer Verkürzung des Meßzyklus und sind damit zum Einsatz in Notsituationen gut geeignet.



Eine einfache Strategie zur Priorisierung relevanter Objekte besteht darin, bei der Erfassung vom Nahbereich zum Fernbereich vorzugehen und bei Überschreiten eines definierten Rechenfensters die Verarbeitung abubrechen. Bei diesem Ansatz wird implizit davon ausgegangen, daß nahe Objekte relevanter sind als weiter entfernte Objekte.

## 2. Beschränkung auf bestimmte Objekte respektive Objektklassen:

Hierunter läßt sich sowohl eine Priorisierung in der Verarbeitung bereits erkannter, getrackter Objekte verstehen, als auch eine Beschränkung in der Detektion auf bestimmte, relevante Objektklassen, z. B. auf Fahrzeuge, bewegte Objekte oder aber Objekte ab einer bestimmten Größe. Ebenso lassen sich hierunter in Erweiterung des vorangehenden Punktes räumliche Beschränkungen fassen, welche durch Relationen zwischen Objekten bedingt sind (Objekte im Fahrbahnbereich, Fahrzeuge in der linken Spur, etc.). Im Prinzip lassen sich zur Eingrenzung der Objekte nahezu alle verfügbaren Attribute nutzen.

## 3. Beschränkung der Erkennungsgüte und des Detaillierungsgrades (Level-Of-Detail LOD):

Die Erkennungsgüte kann in mehrfacher Hinsicht beschränkt werden. Eine offensichtliche Möglichkeit besteht darin, Objekte nicht in jedem Zyklus oder nicht mit jedem Sensor zu tracken. So ist es beispielsweise denkbar, für jedes Objekt - eventuell in Abhängigkeit einer von der Anwendung festgelegten Priorität - eine gewünschte Erkennungsgüte zu definieren. Ein Tracking des Objektes erfolgt immer dann, wenn die Erkennungsgüte in die Nähe dieser definierten Schwelle kommt. Ähnlich verhält es sich mit dem Level-Of-Detail (LOD): Ob ein Objekt als einfacher Quader, als Polyeder-Fahrzeugmodell oder gar unter Verwendung eines Formmodells mit gekrümmten Konturen getrackt wird, ist ebenfalls aufgabenbezogen steuerbar.

### 10.6.2 Aufmerksamkeitssteuerung bei beschränktem Erfassungsbereich

Im Gegensatz zu den vorgenannten Varianten zur Aufmerksamkeitssteuerung, die im wesentlichen durch eine reine Softwaresteuerung zu realisieren sind, bedingen die in diesem Abschnitt genannten Maßnahmen zusätzliche Unterstützung durch die Verhaltensebene und Aktorik, da die Beschränkung der Sensorerfassungsbereiche sich nur durch eine Relativbewegung des Sensors gegenüber dem Umfeld - also eine Verlagerung des Erfassungsbereichs umgehen läßt. Eine ausführliche Darstellung der Thematik Aufmerksamkeitssteuerung findet sich in [Pellkofer 03]. An dieser Stelle sollen die bestehenden Möglichkeiten nur der Vollständigkeit halber kurz angerissen werden:

#### 1. Lagerung des Sensors auf einer beweglichen Plattform:

Diese Möglichkeit kam für Videosensoren im Rahmen der Arbeiten am ISF bereits sehr früh zum Einsatz [Mysliwetz 90]. Handelte es sich dabei anfangs um eine reine Lösung zur Nachführung des Erfassungsbereichs beispielsweise bei Kurvenfahrt,

so wurde dies durch die Konstruktion immer leistungsfähigerer Plattformen dahingehend ausgedehnt, daß auch sakkadische Blickzuwendungen durchgeführt werden konnten [Schiehlen 95]. Neben dem mit entsprechenden Kosten sowie notwendigem Bauraum verbundenen Einsatz solcher Plattformen muß auch eine entsprechende Positioniergenauigkeit gewährleistet sein, um die Güte der Messung sicherzustellen.

Trotz der mit einer solchen Lösung verbundenen Probleme, stellt sie gerade im Hinblick auf die weite Vorausschau mittels Videosensoren in Kombination mit dem sehr eingeschränkten Erfassungsbereich bei hohen Brennweiten und der dabei auch erforderlichen Nickstabilisierung, einen längerfristig interessanten Ansatz dar. Derzeit ist an einen Einsatz im Rahmen von Fahrerassistenzsystemen allerdings nicht zu denken, auch wenn die Idee in letzter Zeit in Verbindung mit der Beherrschung innerstädtischer Szenarien (Ampelblick) in der industriellen Forschung wieder aufgegriffen wurde [Gehrig, Franke 03]

## 2. Änderung der Aspektbedingungen durch Bewegung des Eigenfahrzeugs:

Bei festverbauter Sensorik oder auch in Ergänzung der vorgenannten Lösung, kann der Erfassungsbereich natürlich auch stets durch Eigenbewegung und eine daraus resultierende Positionierung gegenüber dem zu erfassenden Bereich bzw. Objekt erfolgen. Diese Möglichkeit besteht natürlich nur beim autonomen Fahren, wo das System über die Bewegung des Eigenfahrzeugs bestimmt. Aber selbst dann sind natürlich die aktuellen Randbedingungen zu beachten, beispielsweise bei Straßenfahrt die StVO. Am ehesten ist eine solche Form der Szenenexploration noch bei autonomen Fahrten abseits befestigter Wege einzusetzen. Für den Einsatz im Bereich des assistierten Fahrens sind derzeit keine Möglichkeiten zu sehen.

### 10.6.3 Strategien zur Aufmerksamkeitssteuerung

Im Bereich der Strategien zur Aufmerksamkeitssteuerung lassen sich drei typische Verfahren unterscheiden:

1. **Statische Strategien** zur Aufmerksamkeitssteuerung sind unabhängig von den aktuellen Anforderungen der Funktionen und verfolgen fest vorgegebene Ziele. Beispiele für statische Strategien sind oben bereits genannt worden:
  - Räumlich nächste Objekte zuerst
  - Zeitlich nächste Objekte zuerst (Time-To-Collision(TTC)-Ansatz)
  - Priorisierung von Objektklassen
2. **Dynamische Strategien** zur Aufmerksamkeitssteuerung sind situationsabhängig und müssen damit durch die Funktion(en) vorgegeben werden, da nur diese über die Relevanz bestimmter Objekte und Informationen entscheiden können. Bei Systemen, bei denen mehrere Funktionen auf eine gemeinsame Sensorik / Sensorplattform zugreifen bedarf es einer Priorisierung der Anforderungen. Diese Priorisierung kann ebenfalls statisch oder dynamisch angelegt sein.

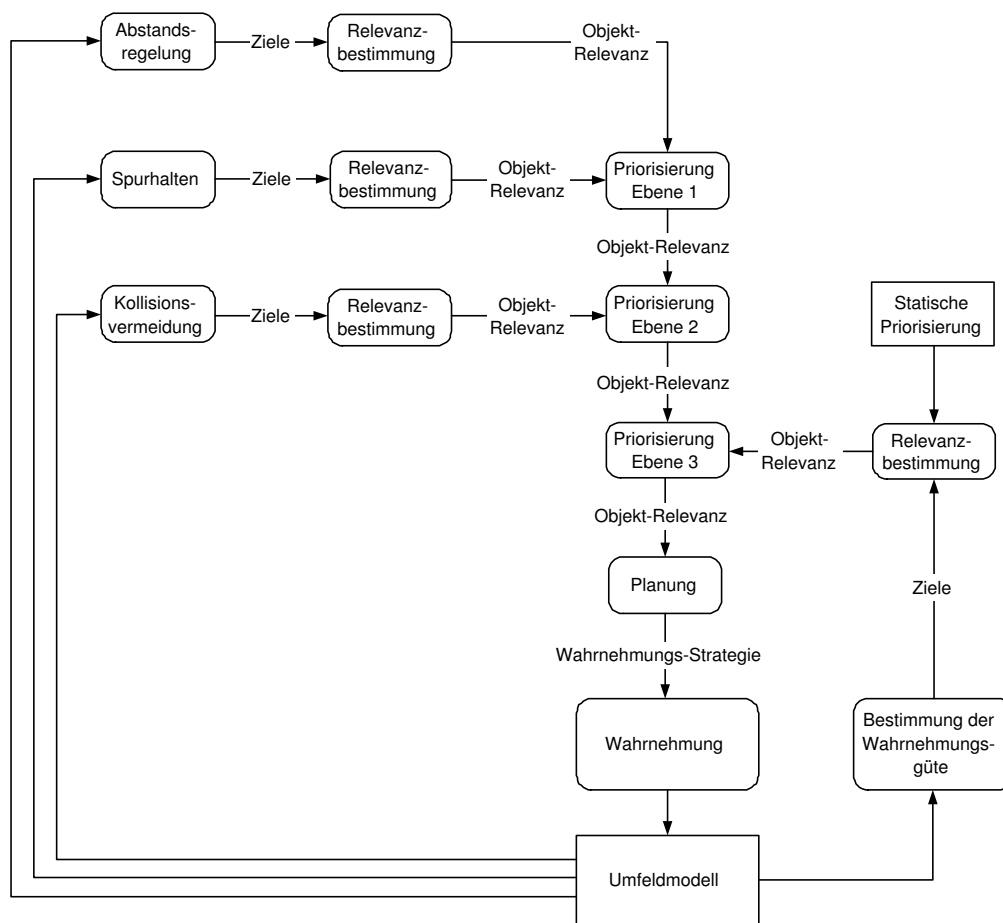


Abbildung 10.8: Beispiel einer Aufmerksamkeitssteuerung bei mehreren, unterschiedlich priorisierten Funktionen

3. **Hybride Strategien** zur Aufmerksamkeitssteuerung vereinen beide vorgenannten Ansätze, d. h. es werden grundsätzlich definierte statische Strategien verfolgt, die bei Bedarf durch dynamische Strategien ergänzt oder vielmehr übersteuert werden können.

In Abbildung 10.8 ist ein Beispiel eines auf eine hybride Strategie ausgerichteten Aufmerksamkeitssteuerungs-“Experten“ dargestellt. Grundsätzlich werden die vordefinierten statischen (Basis-) Strategien genutzt, welche durch eine kaskadierte Abfolge priorisierter dynamischer Strategien ergänzt werden. Die Priorisierung bzw. Konfliktlösung der dynamischen Anforderungen ist hier als korrespondierend zur Relevanz des Ziels der Funktion im Sinne einer “Überlebensstrategie“ angeordnet. Wichtig ist, auch an dieser Stelle zu betonen, daß das Ergebnis der Aufmerksamkeitssteuerung keine “ad hoc“-Strategien sein dürfen, die von Zyklus zu Zyklus erstellt und verworfen werden. Es muß sich dabei vielmehr um Pläne mit entsprechendem Planhorizont handeln.

## Teil III

# Realisierung und Experimentelle Ergebnisse



# Kapitel 11

## Architektur des Wahrnehmungssystems

In diesem Kapitel steht die Architektur und die Implementation des Wahrnehmungssystems im Mittelpunkt. Angefangen bei einer kurzen Rekapitulation der historischen Entwicklung am ISF, welche vom Autor in wesentlichen Teilen mitgeprägt wurde, wird zunächst die Versuchshardware, Fahrzeuge wie Rechnerhardware, beschrieben. Den größten Teil macht dann die Darstellung der Software aus, angefangen bei der Wahl von Betriebssystem und Programmiersprache über die Prozeßstruktur hin zur Diskussion der essentiellen Klassen des Systems und eines beispielhaften Ablaufs eines Meßzyklus.

### 11.1 Historische Entwicklung

Wie in der Einleitung bereits erwähnt, hat die Hard- und Softwareentwicklung zum dynamischen maschinellen Sehen an der UniBwM eine lange Tradition, die bis in die frühen 80er Jahre zurückreicht [Meissner 82], [Graefe 84]. Angefangen bei den ersten Experimenten mit dem Stab-Wagen-System über das Satellitenandockmanöver hin zu den Versuchen im Straßen- und Luftfahrzeugbereich, werden sowohl von der Hardware- wie von der Software-seite sehr hohe Anforderungen gestellt, um Echtzeitfähigkeit zu erzielen. Im Bereich der Hardware wurden in den Jahren 1980-1992 verschiedene Multiprozessorsysteme auf Intel-Basis mit analogem Videobus entwickelt, welche auch die ersten autonomen Fahrten mit dem VaMoRs ermöglichten. Softwareseitig wurde das Ganze flankiert von hochoptimierten Algorithmen und deren effizienter Umsetzung mittels verschiedenster Programmiersprachen. Mit Beginn der 90er Jahre ergab sich dann die Möglichkeit, von den voluminösen Rechnern mit sehr hohem Leistungsbedarf auf ein seit Ende der 80er Jahre aufgekommenes Multiprozessor- Parallelrechnerkonzept aus Transputern umzusteigen. Diese waren als Off-The-Shelf(OTS)-Hardware nicht nur kostenmäßig deutlich günstiger, sondern waren durch das Link-Konzept wesentlich flexibler zu vernetzen und beanspruchten weniger Bauraum bei höherer Rechenleistung und geringerem Leistungsbedarf. Da bei Transputern kein Softwarebetriebssystem nötig war - dieses war vielmehr in Hardware implementiert - waren sie auch den Anforderungen der Echtzeitfähigkeit ohne weiteres gewachsen.

Mit dem Umstieg auf Transputer einher ging auch eine komplette Neuimplementation der Software nach dem zu Beginn der 90er Jahre maßgeblichen Stand der Technik in "C" bzw. in besonders zu optimierenden Programmteilen in "OCCAM" - der "Assembler"-Sprache der Transputer. Die Konfiguration des ersten Transputersystems [von Holt 91] mit dem an der UniBwM im Jahr 1991 erstmals autonom gefahren wurde, ist in Bild 11.1 dargestellt. Der Bildeinzug von einer Kamera erfolgte über den Framegrabberknoten. Dieser verteilte auf Anforderung Bildausschnitte an die drei mit ihm verbundenen Transputer, auf welchen die Algorithmen zur Kantenextraktion abliefen. Die eigentliche Spurerkennung inklusive des Kalman-Filters liefen auf einem 32-Bit Transputer; ein weiterer diente der Fahrzeugführung sowie der Abwicklung von I/O-Aufgaben zum Fahrzeug und zum Benutzer. Die gesamte Rechenleistung für das autonome Straßenfahren - ohne Erkennung anderer Fahrzeuge - paßte damit schon zu Beginn der 90er-Jahre in eine Box der Größe eines Schuhkartons !

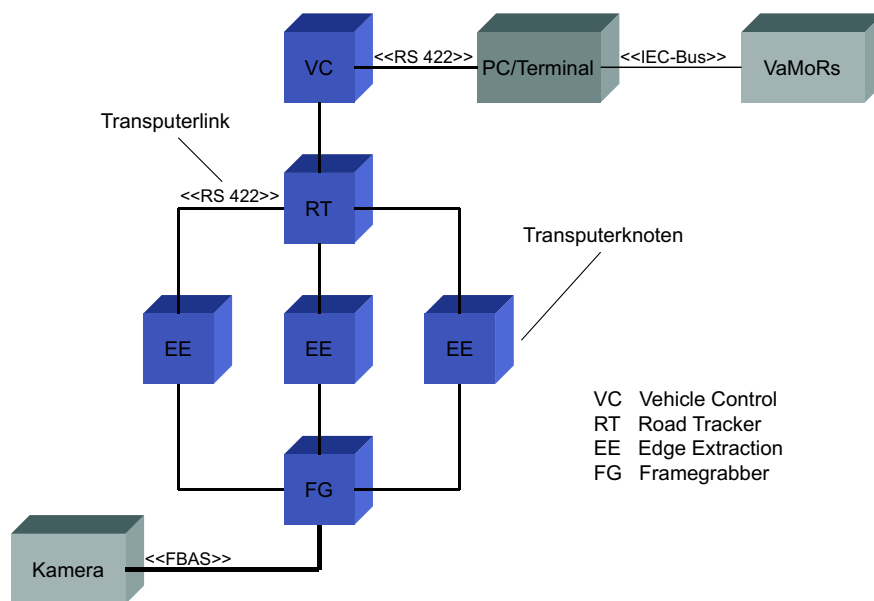


Abbildung 11.1: Struktur des ersten Transputersystems zum autonomen Fahren am ISF

In der Folge wurde dieses System kontinuierlich weiter ausgebaut. Anfangs wurde zur Objekterkennung noch das Intel-Mehrprozessorsystem eingesetzt und zu diesem Zweck über einen Transputerlink mit der Fahrspurerkennung gekoppelt; später wurde auch diese Fähigkeit auf Transputer portiert. In seiner höchsten Ausbaustufe bestand das System zur bordautonomen Fahrzeugführung letztlich aus über 30 Transputern [Dickmanns et. al. 94].

Aber so gut das Transputerkonzept auch war, gelang es der englischen Herstellerfirma INMOS nicht, die angekündigte Nachfolge-Prozessor-Generation zeitgerecht am Markt zu platzieren. Anderen proprietären Entwicklungen wie etwa Verbundsystemen aus Transputern (zur Kommunikation) und PowerPC-Prozessoren (zur Berechnung) war nur ein kurzes Leben beschieden. Die Entwicklung wurde damals vielmehr von der sich immer weiter ausbreitenden Welle der PC's überrollt, die mit der Zeit so leistungsfähig und dabei so konkurrenzlos günstig geworden waren, daß sie auch für einen Einsatz im Bereich echtzeitfähiger Bildverarbeitung geeignet erschienen. Auch in diesem Bereich wurden am ISF Maßstäbe gesetzt als es 1996 gelang, die gesamte Software zum autonomen Straßenfah-

ren - wiederum zunächst ohne Objekterkennung - auf einem Pentium-90-PC ablaufen zu lassen [Gregor et. al. 97]. Das System wurde dann auf der Basis der Arbeiten von D. Dickmanns [D. Dickmanns 97] zu einem modularen Konzept “EMS-Track“ [von Holt, Baten 98] weiterentwickelt, welches basierend auf einem in C++ implementierten Rahmensystem eine baumartige Szenenrepräsentation beinhaltet, an die in weitem Umfang konfigurierbare Module zur Erkennung, Vermessung und Schätzung von Objekten angelagert werden konnten. Die Erfahrungen mit diesem System liegen auch dem im Rahmen dieser Arbeit entwickelten Systemansatz zugrunde.

## 11.2 Hardware

### 11.2.1 Versuchsfahrzeuge

Als Versuchsfahrzeug diente am ISF der dortige 5t-Kastenwagen VaMoRs (Bild 11.2). Auf diesem Fahrzeug wurde das o. g. Transputersystem erstmalig eingesetzt; mit diesem Fahrzeug wurde 1996 auch erstmals rein PC-basiert autonom gefahren. Für die im Umfang dieser Arbeit erfolgten Untersuchungen bildete das VaMoRs die erste Sensorplattform, mit der die Testdatensätze für die weiteren Untersuchungen im Labor eingefahren wurden. Die Umfeldwahrnehmung im VaMoRs war allerdings rein videobasiert mit einer in 2-Achsen beweglichen Kameraplattform sowie bis zu 4 Kameras.

Zur Untersuchung der hier vorgestellten Sensoranordnung aus Lidar- und Videosensor wurde daher auf eine andere Versuchsplattform übergegangen, einen AUDI A8 der 2. Generation (D2), der in Abb. 11.3 zu sehen ist. Dieser war mit den in Kapitel 4 vorgestellten Sensoren ausgerüstet sowie der Möglichkeit einer Sensordatenaufzeichnung ausgestattet. Mit diesem Fahrzeug wurden die Testsequenzen aufgezeichnet, um dann im Labor weiterbearbeitet zu werden. Da eine Einbindung des hier vorgestellten Wahrnehmungssystems in eine Fahrfunktion, sei es nun das autonome Fahren oder das assistierte Fahren, an dieser Stelle nicht betrachtet wird, fanden keine Fahrversuche im engeren Sinne statt.



Abbildung 11.2: Versuchsfahrzeug VaMoRs des ISF



Abbildung 11.3: Versuchsfahrzeug AUDI A8 der AUDI AG



### 11.2.2 Rechnerhardware

Die Realisierung des Systems zur Umfeldwahrnehmung erfolgte auf gängiger Standard-PC Hardware. Für die Umsetzung im Fahrzeug stand ein 1 GHz schneller PENTIUM III - Rechner zur Verfügung, bei dem es sich um einen fahrzeuggerecht in ein anderes Gehäuse umgebauten Standard-Rechner handelte. Dieser wurde erweitert um die notwendigen Schnittstellenkarten zum Einlesen des Videobildes mit einer Analog-Framegrabber-Karte der Firma MATROX sowie um zwei 2-kanalige CAN-Bus Steckkarten der Firma VECTOR. Die Konfiguration des Systems ist in Abb. 11.4 enthalten. Mit dieser Anordnung war sowohl

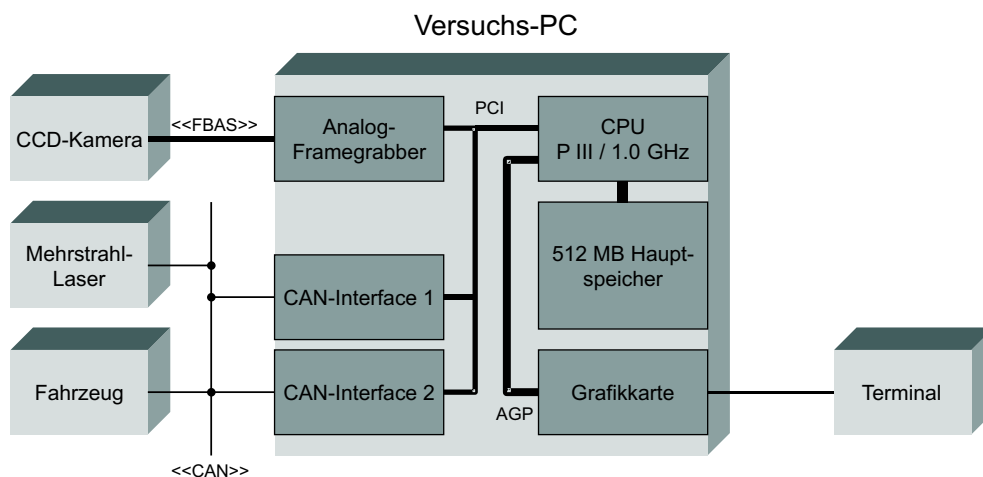


Abbildung 11.4: Hardware-Anordnung des Versuchssystems

der Online-Betrieb im Fahrzeug wie die Aufzeichnung von Testdaten möglich. Der überwiegende Teil der Entwicklung konnte so auf aufgezeichneten Daten mittels eines Standard-Notebooks mit einem PENTIUM IV - 1.8 GHz erfolgen.

## 11.3 Software

### 11.3.1 Betriebssystem und Programmiersprache

Die Software zur Umfelderkennung entstand als nicht verteiltes, "peripher"-parallelisiertes System. "Peripher"-parallelisiert meint in diesem Zusammenhang, daß nur periphere Module aufgrund der Systemrandbedingungen parallelisiert wurden (CAN-Bus-I/O, Bild-I/O, Benutzeroberfläche), während das Kernmodul zur Wahrnehmung als einzelner Prozeß realisiert ist. Es bestehen zwar mehrere Möglichkeiten zur Parallelisierung auch des Kernmoduls, die aber aus Gründen der Performance nicht notwendig waren. Somit bestand kein Anlaß, das System in dieser Hinsicht komplexer als notwendig zu gestalten. Die Implementation erfolgte in der Programmiersprache C++ unter dem Betriebssystem WINDOWS 2000. Mit der Wahl von WINDOWS 2000 kann zwar eine tatsächliche Echtzeitfähigkeit nicht gewährleistet werden, für die Zykluszeiten der Sensoren von 40 bzw. 60 ms ergeben

sich aber keine Probleme, sofern der Rechner nur moderat ausgelastet wird. Erst unter einer Rechenlast von ca. 75 % und mehr kann auch von “annähernder“ Echtzeitfähigkeit unter WINDOWS 2000 nicht mehr die Rede sein.

### 11.3.2 Prozeßstruktur

In Abbildung 11.5 ist das Verteilungsdiagramm in UML-Notation zu sehen. Der PC beinhaltet die vier Komponenten *Wahrnehmung*, *Datenbank*, *Graphical User Interface(GUI)-Server* und *I/O-Server*. Über den CAN-Bus sind das Eigenfahrzeug an sich sowie der Lidarsensor an den PC angebunden. Der Anschluß der CCD-Kamera erfolgt über eine herkömmliche analoge FBAS-Leitung. Das Terminal, also Bildschirm und Tastatur des PC, sind über die PC-üblichen Verbindungen angeschlossen. In Verfeinerung der in Abbildung 11.5 enthaltenen Komponenten ist die Prozeßstruktur auf dem PC in Bild 11.6 dargestellt. Jede Komponente entspricht einem parallel laufenden Prozeß auf dem PC und enthält wiederum mehrere Subkomponenten. Dies sind im Fall der Datenbank z. B. Datenfelder für die Sensoreingangsdaten, die Szenendaten, die Situation beschreibende Daten und - hier als Beispiel für eine Unzahl systeminterner Daten zur Ablaufsteuerung - die Ereignisverwaltung und Ereignistabelle. Die Wahrnehmungskomponente als eigentlicher Kern des Systems besteht wie zu erkennen ist aus vier “Experten“ genannten Einheiten. Diese korrespondieren weitestgehend mit den in Abbildung 10.2 in Kapitel 10.1 dargestellten “Experten“-Modulen. Es handelt sich um Module, die in Form von Softwareagenten für eine bestimmte Aufgabe konzipiert sind, für welche sie dann eben im Sinne des Gesamtsystems als “Experten“ erscheinen. Die Kommunikation zwischen den einzelnen Komponenten erfolgt stets zentral und asynchron über die Datenfelder und Ereignisverwaltung in der Datenbank. Dadurch lassen sich eine weitreichende Entkoppelung der Module und eindeutige Modulschnittstellen erreichen.

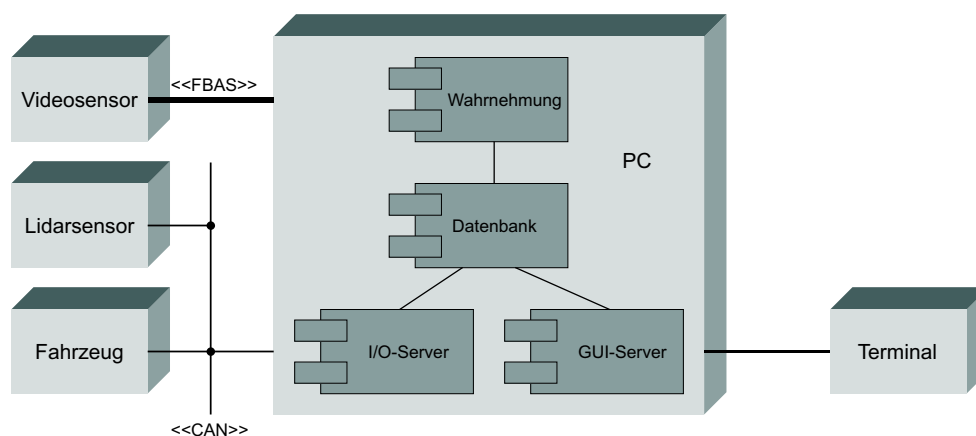


Abbildung 11.5: Verteilungsdiagramm mit Komponenten

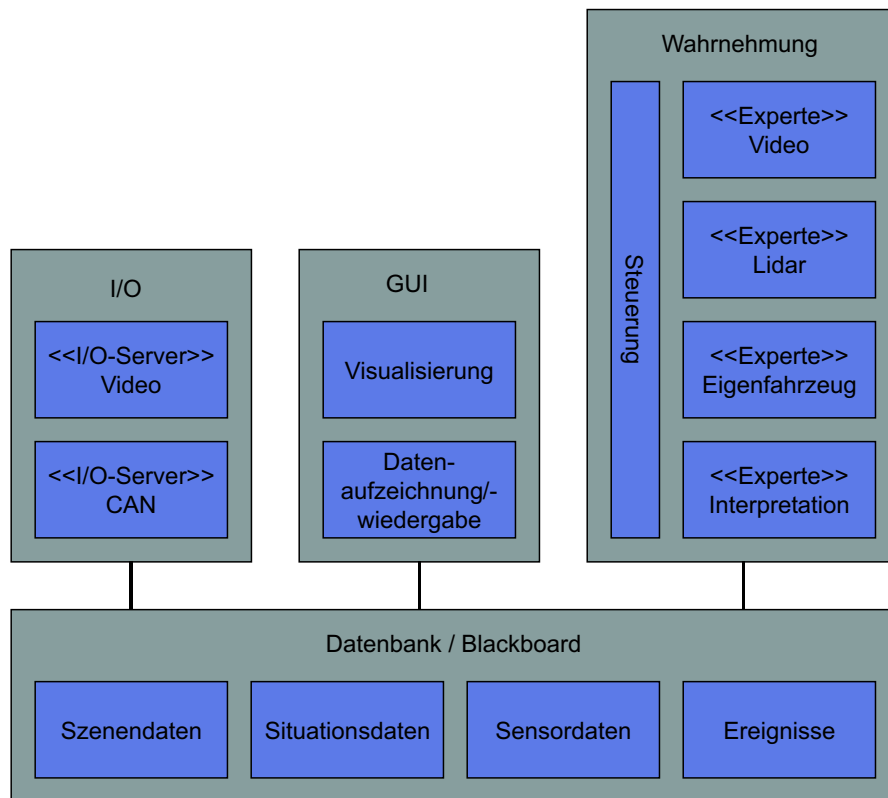


Abbildung 11.6: Prozeßstruktur zur Umfeldwahrnehmung

### 11.3.3 Essentielle Klassen

Die statische Struktur der essentiellen Klassen der Wahrnehmungskomponente ist in Abb. 11.7 wiedergegeben. In ihr sind nur die wesentlichsten Klassen auf einer hohen Abstraktionsstufe enthalten, wobei jegliche Klassen zur Ablaufsteuerung und Kontrolle ebenfalls nicht dargestellt sind. Die beiden Hauptklassen sind einerseits *CSensor* als Basisklasse für alle weiteren Sensoren und andererseits *CUmfeldModell*. *CUmfeldModell* beinhaltet zum einen - wie der Name vermuten läßt - die eigentliche Szenenrepräsentation der Fahrspur *CFahrspur*, der Objekte *CObjekt* und des Eigenfahrzeugs *CEgo*, zum anderen stellt es auch die sensorunabhängigen Methoden zur Wahrnehmung, d. h. die zeitliche Prädiktion *CUmfeldModell::UDT*, die Innovation *CUmfeldModell::UDM* und die zeitliche Plausibilisierung *CUmfeldModell::UDE* bereit. Die Elemente der Szenenrepräsentation *CEgo*, *CObjekt* und *CFahrspur* bestehen ihrerseits jeweils aus einem oder mehreren Bewegungsmodellen *CBewegungsModell* und einem oder auch mehreren Formmodellen *CFormModell*, welches die vermeßbaren Merkmale des Objekts umfaßt. Die Klasse *CUmfeldModell* ist assoziiert mit einem oder mehreren Sensoren, die von der Klasse *CSensor* abgeleitet sind: *CLidarSensor*, *CVideoSensor* oder *CEgoSensor*. Jeder Sensor verfügt über die sensorspezifischen Methoden der Transformation vom Zustandsraum in den Merkmalsraum *CSensor::TSF*, der Meßdatenassoziation *CSensor::ASO*, der Hypotheseninstanziierung *CSensor::INS* und einer eventuellen Meßdaten“vorverarbeitung“ *CSensor::MPP*. Im Schritt Meßdaten“vorverarbeitung“ ist beim Videosensor etwas irreführend die eigentliche, erwartungsbasierte Bildvermessung

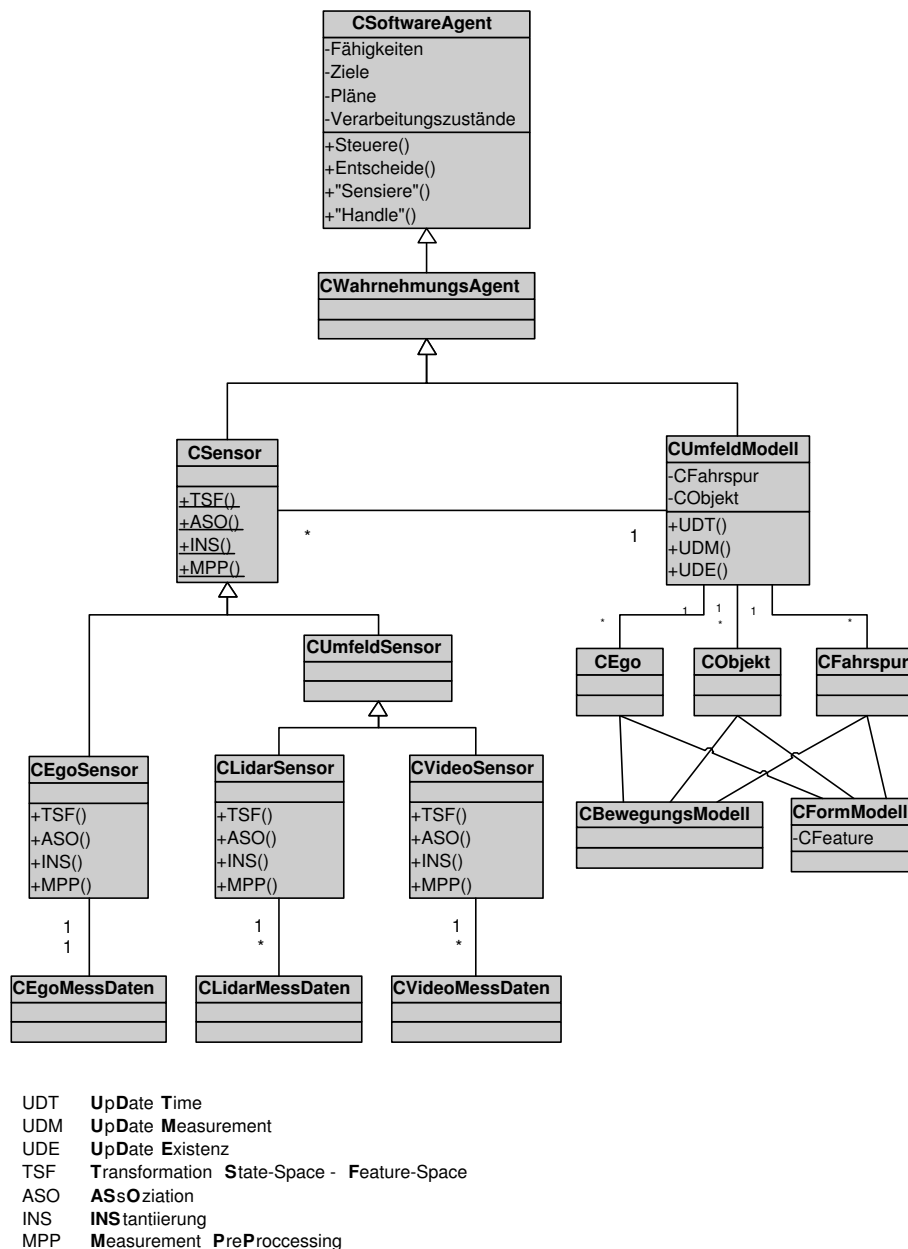


Abbildung 11.7: Hierarchie essentieller Klassen der Wahrnehmungskomponente

angesiedelt, beim Lidarsensor hingegen die Segmentierung. Die vorgenannten Methoden sind in der Basisklasse *CSensor* als abstrakte Methoden definiert, so daß jede daraus abgeleitete Sensorklasse diese Methoden zwangsweise implementieren muß.

Darüberhinaus kann man erkennen, daß sowohl die Klasse *CSensor* wie auch *CUmfeldModell* aus der Klasse *CWahrnehmungsAgent* abgeleitet ist, die ihrerseits wiederum aus der Basisklasse *CSoftwareAgent* abgeleitet ist. Diese beiden Klassen definieren Basisfunktionalitäten, wie sie für Softwareagenten typisch sind.

Charakteristisch für einen Agenten sind die eigenständige zielorientierte Bearbeitung von Aufgaben. Hierzu verfügen Softwareagenten über eine Gedächtniskomponente mit den dar-

in gespeicherten Zielen respektive abzuarbeitenden Aufträgen sowie einer Repräsentation der eigenen Fähigkeiten. Anhand der vorhandenen Fähigkeiten werden Pläne zur Zielerreichung generiert, deren Abarbeitung sich über mehrere Verarbeitungsschritte ausdehnen kann und darum die Zwischenspeicherung des jeweiligen Verarbeitungszustands erfordert.

Die KI-Literatur räumt der Behandlung von Softwareagenten breiten Raum ein; einen Überblick gibt z.B. [Burkhard 00]. Es sei allerdings darauf hingewiesen, daß die im Bereich der KI verwendeten Begriffe zur Beschreibung von Softwareagenten gerade im Umfeld der Robotik dazu angelegt sind, Verwirrung zu stiften, wenn von Sensoren, Aktoren, Umweltmodell, Fähigkeiten etc. gesprochen wird. Diese - auch in der Robotik verwendeten - Begriffe sind jedoch im Umfeld der Softwareagenten gänzlich anders besetzt: Das sogenannte "Umweltmodell" eines Softwareagenten für Flugbuchungen etwa wird aus Flugplänen und Reiseplänen bestehen und hat nichts mit der Umwelt im Sinne des in dieser Arbeit und allgemein in der Robotik benutzten Begriffs der Umwelt oder des Umfelds zu tun. Erschwert wird die Begrifflichkeit allerdings dadurch, daß es auch im Bereich der Robotik KI-Ansätze gibt, die auf dem Einsatz von Softwareagenten beruhen und bei denen dann tatsächlich ein Bezug zu den - in der Robotik üblichen - Begriffen gegeben ist.

Agenten lassen sich darüberhinaus in vielfältiger Weise beispielsweise hierarchisch miteinander verknüpfen um zu interagieren oder zu delegieren. Den Einsatz eines Multi-Agenten-Ansatzes im Bereich Fahrerassistenz zeigt [Görzig 03]. Dort wird ebenfalls ein Überblick über das Thema speziell im Hinblick auf Fahrerassistenzsysteme gegeben.

### 11.3.4 Beispielhafte Abläufe

Der Ablauf der Wahrnehmung und die Interaktion soll an einem beispielhaften Ablauf kurz umrissen werden.

In der Abb. 11.8 ist ein Sequenzdiagramm für die Einbringung von Meßdaten des Lidarsensors zu erkennen. Nach dem Eintreffen der Meßdaten im Kontrolltask *Wahrnehmung* ermittelt diese den Meßzeitpunkt und ruft mit diesem als Parameter die Prädiktionsroutine des *Umfeldmodells* *UmfeldModell::UDT* auf, welches alle Objekte des Umfeldmodells auf den Zeitpunkt der Messung prädiziert. Daran anschließend werden die sensorspezifischen Methoden zur Transformation der prädizierten Merkmale in den Merkmalsraum des Lidarsensors *LidarSensor::TSF*, die Meßwertvorverarbeitung *LidarSensor::MPP* zur Segmentierung und die Merkmals-Assoziation *Lidarsensor::ASO* aufgerufen. Mit dem Ergebnis der Assoziation können nun durch die sensorübergreifende Methode *UmfeldModell::UDM* die Objektzustände aktualisiert werden, ehe - wiederum sensorspezifisch - die Instanziierung neuer Objekthypothesen in *LidarSensor::INS* erfolgt. Den Abschluß der Einbringung der neuen Lidarmeßdaten bildet dann die Prüfung der zeitlichen Konsistenz des Umfeldmodells *UmfeldModell::UDE*, in dessen Rahmen zeitlich inkonsistente Hypothesen verworfen werden.

In der danebenstehenden Abbildung 11.9 ist der analoge Ablauf für die Einbringung neuer Videodaten aufgezeigt. Die Abläufe auf dieser Ebene entsprechen einander; anstelle der sensorspezifischen Methoden des Lidarsensors werden in diesem Fall natürlich die korre-

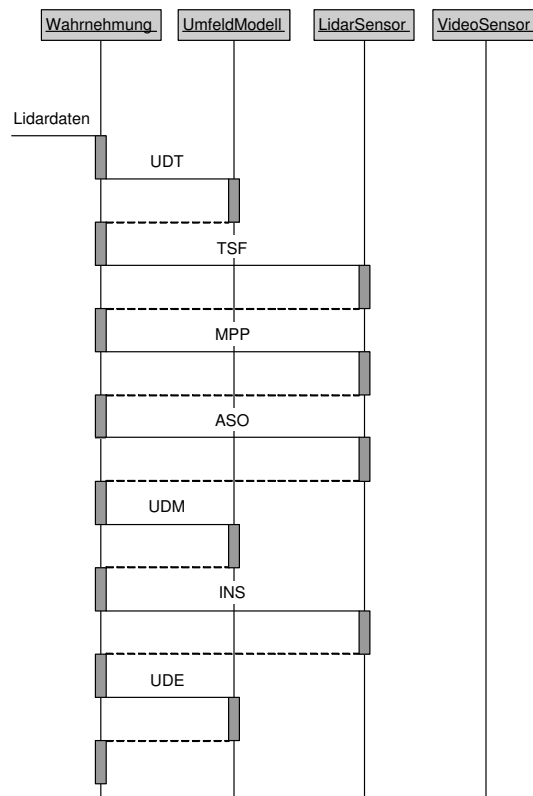


Abbildung 11.8: Sequenzdiagramm für die Einbringung von Lidarmessdaten

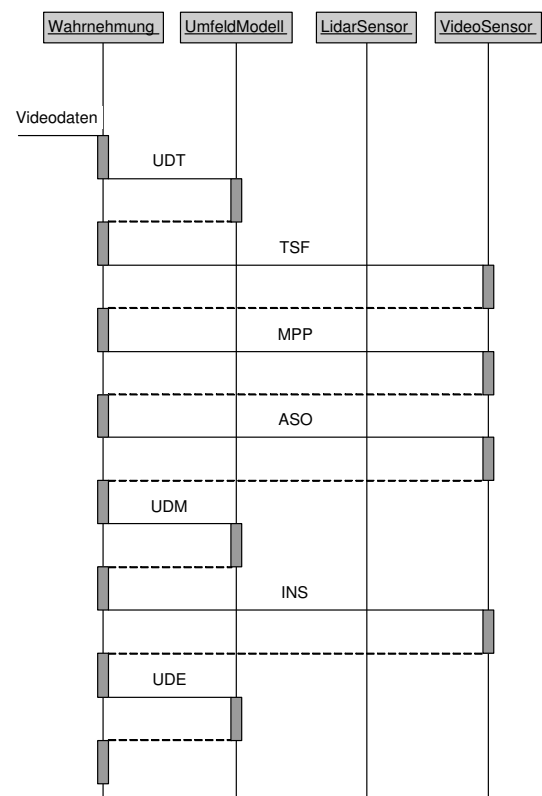


Abbildung 11.9: Sequenzdiagramm für die Einbringung von Videomessdaten

spondierenden Methoden des Videosensors (z. B. Bildvermessung anstelle Segmentierung) aufgerufen. Daß jeder Sensor über die entsprechenden Methoden im Sinne einer einheitlichen Schnittstelle verfügt, ist durch die in Abb. 11.7 gezeigte Vererbungshierarchie über die Klasse *CSensor* und die dort definierten abstrakten Methodenaufrufe sichergestellt.



# Kapitel 12

## Experimentelle Ergebnisse

In diesem Abschnitt werden beispielhaft einige Ergebnisse gezeigt und diskutiert, die mit den in dieser Arbeit vorgestellten Umfelderkennungsverfahren erzielt wurden. Hierbei haben zwei Aspekte besondere Beachtung verdient. Zum einen werden an dieser Stelle nur Ergebnisse dargestellt, die mit dem Multisensorsystem aus Lidar und Video erzielt wurden und die zunächst als reine Wahrnehmungskomponente ohne nachgeschaltete Anwendung im Fahrzeug “getestet“ wurden. Ein wirklicher “Test“ der Wahrnehmungskomponente ist aber erst in einem geschlossenen Kreis mit einer Anwendung möglich, welche die Ziele vorgibt und damit die Leistung beurteilbar macht. Somit sind die hier gezeigten Ergebnisse in erster Linie qualitativer Natur, denn den Schwerpunkt der Arbeit bildet das Aufzeigen der Perspektive gekoppelter multisensorieller Umfelderkennung mit integrierter Interpretationskomponente. Damit sind wir auch bereits beim zweiten Punkt angelangt, denn auch die in dieser Hinsicht wesentlichen Aspekte sind eher struktureller Natur und daher bereits in den vorangegangenen Kapiteln abgehandelt. Auch sie lassen sich durch einige exemplarische Messungen nur qualitativ stützen. Gleichwohl vermitteln die dargestellten Ergebnisse einen Einblick in die Möglichkeiten und in die Grenzen, welche der hier gewählten konkreten Sensoranordnung zu eigen sind.

Im ersten Abschnitt wird zunächst kurz die Fahrspurerkennung anhand eines Beispielszenarios auf der Autobahn dargestellt. Aufbauend auf diesem Szenario wird in den Folgekapiteln die Objekterkennung, sowohl rein-lidarbasiert wie auch lidar- und videobasiert anhand verschiedener Szenarien wie auch anhand unterschiedlicher Objekte demonstriert. Abschließend erfolgt eine vergleichende Diskussion der Ergebnisse verbunden mit einer Wertung des Systemansatzes.

In sämtlichen Szenarien wurde die Eigenbewegung des Versuchsträgers aus der Filterung der von der Fahrzeug-Sensorik gelieferten Meßwerte für die Radgeschwindigkeiten bzw. die Gierrate geschätzt. Da es sich bei diesen Sensoren um herkömmlichen Sensoren aus der Serienproduktion handelt, ist die Güte der Meßwerte entsprechend gering. In der Folge leidet auch die Güte der geschätzten Zustandsgrößen von Objekten, da die Güte der Eigenzustandsschätzung mittelbar oder unmittelbar in deren Schätzung Eingang findet.



## 12.1 Fahrspurerkennung

Als Beispielszenario für die Leistungsfähigkeit der Fahrspurerkennung dient eine Fahrscene auf der A 39 von Wolfsburg in Richtung des Autobahnkreuzes Wolfsburg/Königslutter. Der Ablauf des ca. 90 Sekunden dauernden Fahrtausschnittes ist beispielhaft als Bildsequenz in den Bildern 12.1 und 12.2 wiedergegeben. Über den einzelnen Bildern ist zur späteren Referenzierung jeweils eine Bildnummer sowie der Aufnahmezeitpunkt angegeben.

### 12.1.1 Szenarienbeschreibung

Im folgenden wird eine kurze verbale Beschreibung des Szenarios gegeben:

- Zum Aufnahmestart befindet sich das Versuchsfahrzeug auf der rechten Spur der zweispurigen Autobahn in einer leichten Linkskurve.
- Zum Zeitpunkt  $t_1$  wird ein Spurwechseln nach links eingeleitet, der zur Zeit  $t_2$  abgeschlossen ist.
- Es wird zum in der linken Spur vorausfahrenden Fahrzeug aufgeschlossen und dabei ein wenig in der Spur gependelt ( $t_3$ ).
- Als das Folgefahrzeug auf die rechte Spur wechselt, wird diesem weiterhin gefolgt und ebenfalls ein Spurwechsel nach rechts durchgeführt ( $t_4 - t_5$ ).
- Zum Zeitpunkt  $t_6$  ist das bisherige Folgefahrzeug aufgrund eines Kleintransporters in der rechten Spur zum Überholen ausgesichert. Das Versuchsfahrzeug folgt nun dem Kleintransporter.
- Während der Folgefahrt hinter dem Kleintransporter werden zweimal größere Aufbaubewegungen durch Betätigen der Bremse herbeigeführt ( $t_7$ ).
- Zur Zeit  $t_8$  setzt das Versuchsfahrzeug zum Wechsel auf die Verzögerungsspur an, um auszufahren.
- Das Versuchsfahrzeug befindet sich auf der Verzögerungsspur ( $t_9$ ) und fährt in die Ausfahrt.
- In der Ausfahrt erfolgt ein Vorzeichenwechsel der Krümmung und eine Aufweitung der Fahrspur ( $t_{10}$ ).
- Am Ende der Ausfahrt ( $t_{11}$ ) befindet sich eine Einmündung in T-Form in eine Bundesstraße, auf die nach links eingebogen wird; ( $t_{12}$ ) zeigt einen Schnappschuß nach dem Einbiegen.

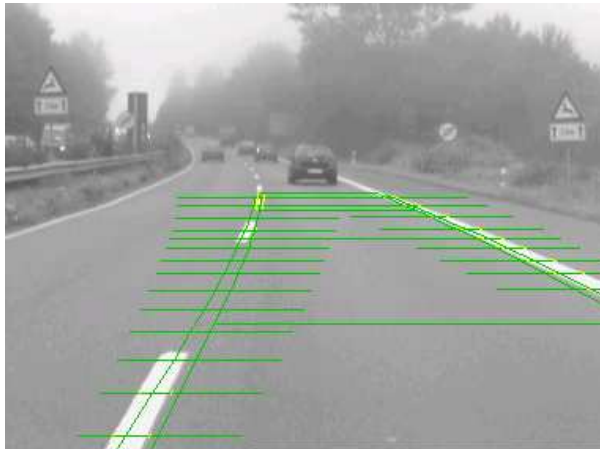
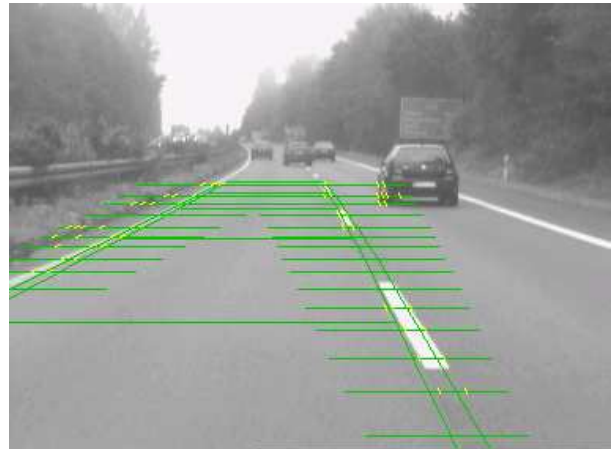
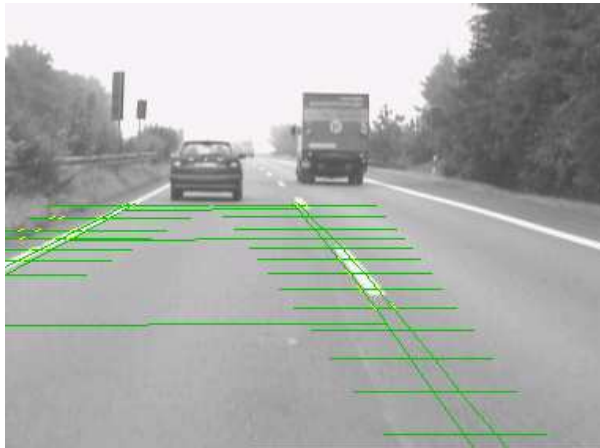
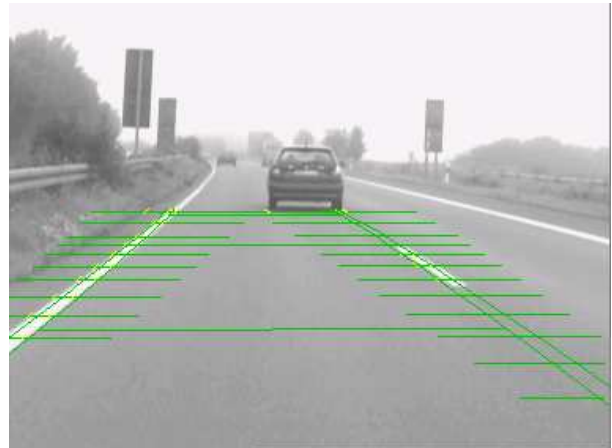
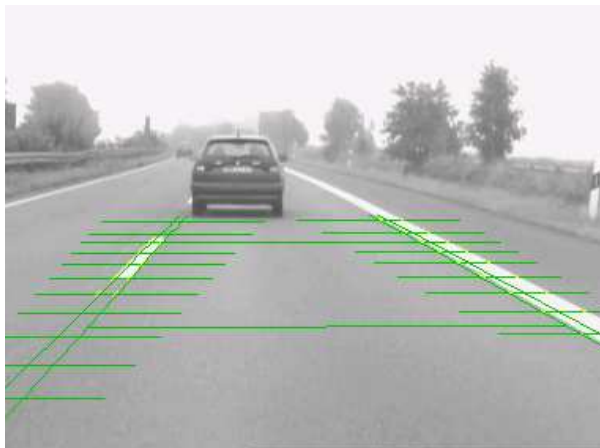
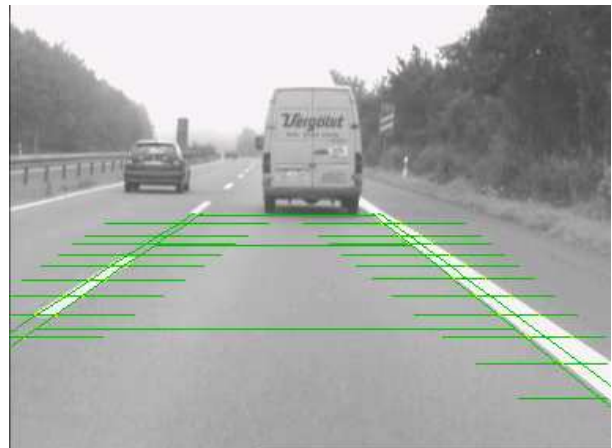
Bild 1 ( $t_1 = 5.22$  s)Bild 2 ( $t_2 = 7.94$  s)Bild 3 ( $t_3 = 19.29$  s)Bild 4 ( $t_4 = 25.18$  s)Bild 5 ( $t_5 = 27.79$  s)Bild 6 ( $t_6 = 41.17$  s)

Abbildung 12.1: Beipielsszenario auf der A 39 bei Wolfsburg, Bild 1-6

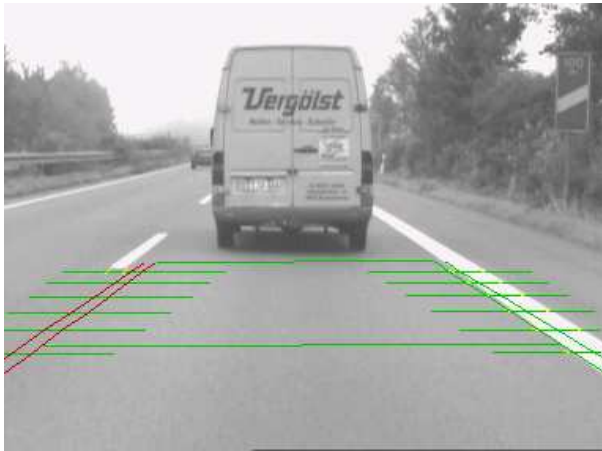
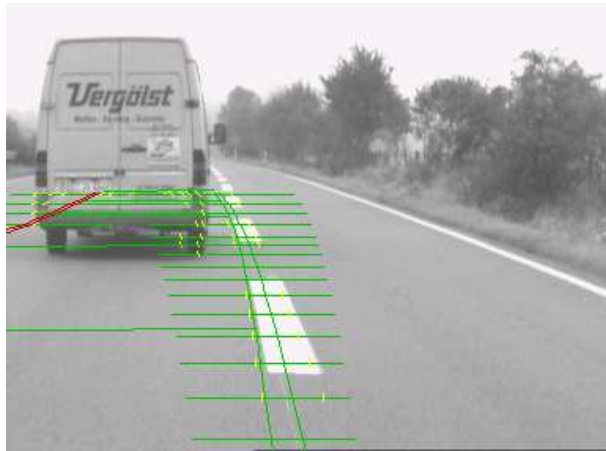
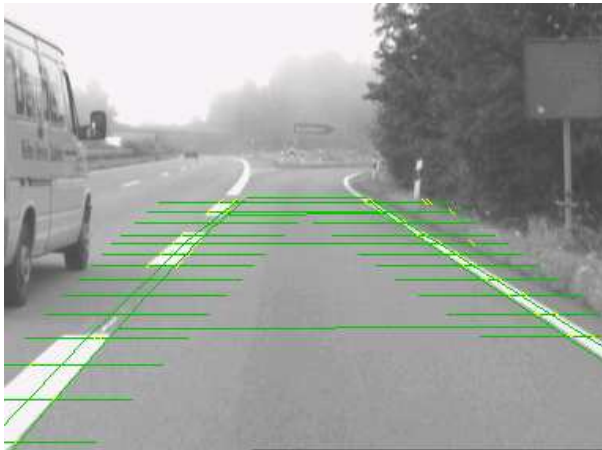
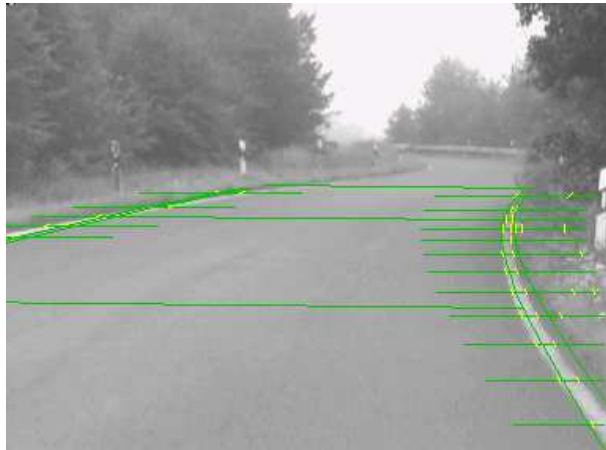
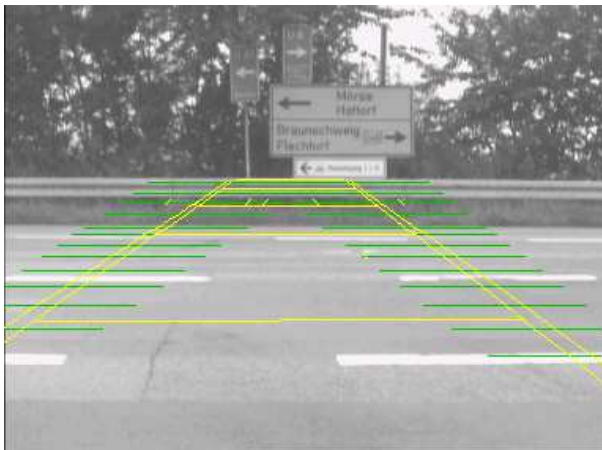
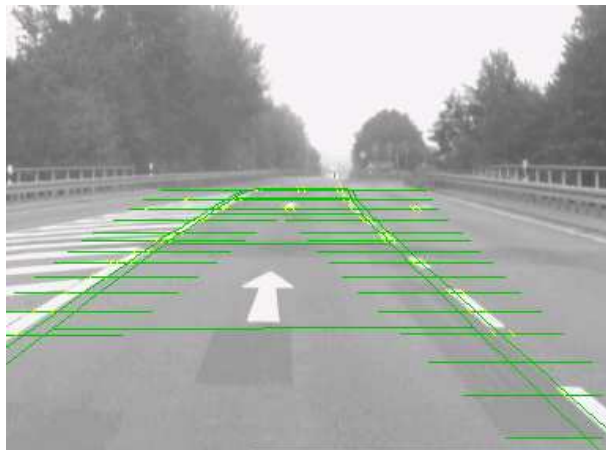
Bild 7 ( $t_7 = 46.28$  s)Bild 8 ( $t_8 = 50.48$  s)Bild 9 ( $t_9 = 56.14$  s)Bild 10 ( $t_{10} = 62.10$  s)Bild 11 ( $t_{11} = 78.08$  s)Bild 12 ( $t_{12} = 86.98$  s)

Abbildung 12.2: Beipielsszenario auf der A 39 bei Wolfsburg, Bild 7-12

### 12.1.2 Ergebnisse der Fahrspurschätzung

Das Ergebnis der Zustandsschätzung der fahrspurbeschreibenden Größen ist in Abbildung 12.3 enthalten. Die Darstellung enthält die wesentlichsten Zustandsgrößen Ablage  $y$ , Nickwinkel  $\theta$ , Gierwinkel  $\psi$ , horizontale Krümmung  $c_{0hm}$  sowie horizontale Krümmungsänderung  $c_{1hm}$  und Fahrspurbreite  $b_0$ .

Anhand der Ablage lassen sich zunächst die vorgenommenen Spurwechsel erkennen, immer dann wenn ein Sprung von einer großen positiven Ablage zu einer großen negativen Ablage erfolgt und umgekehrt. So erfolgt der erste Fahrstreifenwechsel nach ca. 7-8 s (zwischen Bild 1 und Bild 2 des Szenarios) von der rechten auf die linke Fahrspur. Der zweite Fahrstreifenwechsel in umgekehrter Richtung erfolgt nach ca. 26-27 s zurück auf die rechte Spur (zwischen Bild 4 und Bild 5). In der Folge wird zwischen Bild 8 und Bild 9 ein erneuter Spurwechsel nach rechts - auf die Ausfahrt - durchgeführt.

Aus dem darunter dargestellten Verlauf des Nickwinkels ist zum einen ein stärkeres Bremsmanöver hinter dem in Bild 6 zu erkennenden Transporter nach ungefähr 45 s zu erkennen. Der zweite Einbruch des Nickwinkels ist größtenteils eine Fehlschätzung, die daraus resultiert, daß kurz vor der Ausfahrt - wie in Bild 8 zu sehen - sehr nah an den vorausfahrenden Transporter aufgefahren wurde. Beim anschließenden Fahrspurwechsel auf die Ausfahrt wurde schlagartig die Vorausschau wieder größer, gleichzeitig kam eine Rechtskrümmung in Verbindung mit einer starken Aufweitung der Fahrspur und ein leichter vertikaler Anstieg der Fahrspur dazu. Die dabei auftretenden Abweichungen führten in der Folge zu der dargestellten Fehlschätzung des Nickwinkels.

Aus den Graphen, welche die gemittelte horizontale Krümmung, deren Krümmungsänderung, die Breite wie auch den Gierwinkel wiedergeben, läßt sich insgesamt die ungefähr 60 s dauernde Fahrt auf der gleichmäßig breiten Autobahn mit nur geringer Krümmung nachverfolgen. Einzig der Gierwinkel weist größere Schwankungen auf, wobei zu ergänzen ist, daß neben den Spurwechseln auch in der linken Spur Schlangenlinien gefahren wurden (um die Auswahl des direkten Zielobjekts zu stören - Bild 3). Die Ausfahrt ist vor allem im Verlauf der mittleren Krümmung, insbesondere aber am Vorzeichenwechsel der mittleren Krümmungsänderung gut zu verfolgen. Ebenso deutlich ist die Abweichung der Spurbreite der Ausfahrt - ca. 4.5 - 5 m - von der Spurbreite der Autobahn zu erkennen.

In der Ausfahrt erfolgte aufgrund der höheren Gierrate, der geringeren Eigengeschwindigkeit sowie der höheren geschätzten Krümmungswerte eine Adaption der Dynamik des Krümmungsschätzers sowie - damit gekoppelt - eine Reduzierung der Vorausschauweite. Erst durch diese Maßnahmen war eine Fahrspurerkennung in der Ausfahrt überhaupt möglich. Umgekehrt führt erst die herabgesetzte Dynamik auf der Autobahn zu einem stabilen, von Störungen durch Objekte freien Tracking des Spurverlaufs bei höheren Geschwindigkeiten.

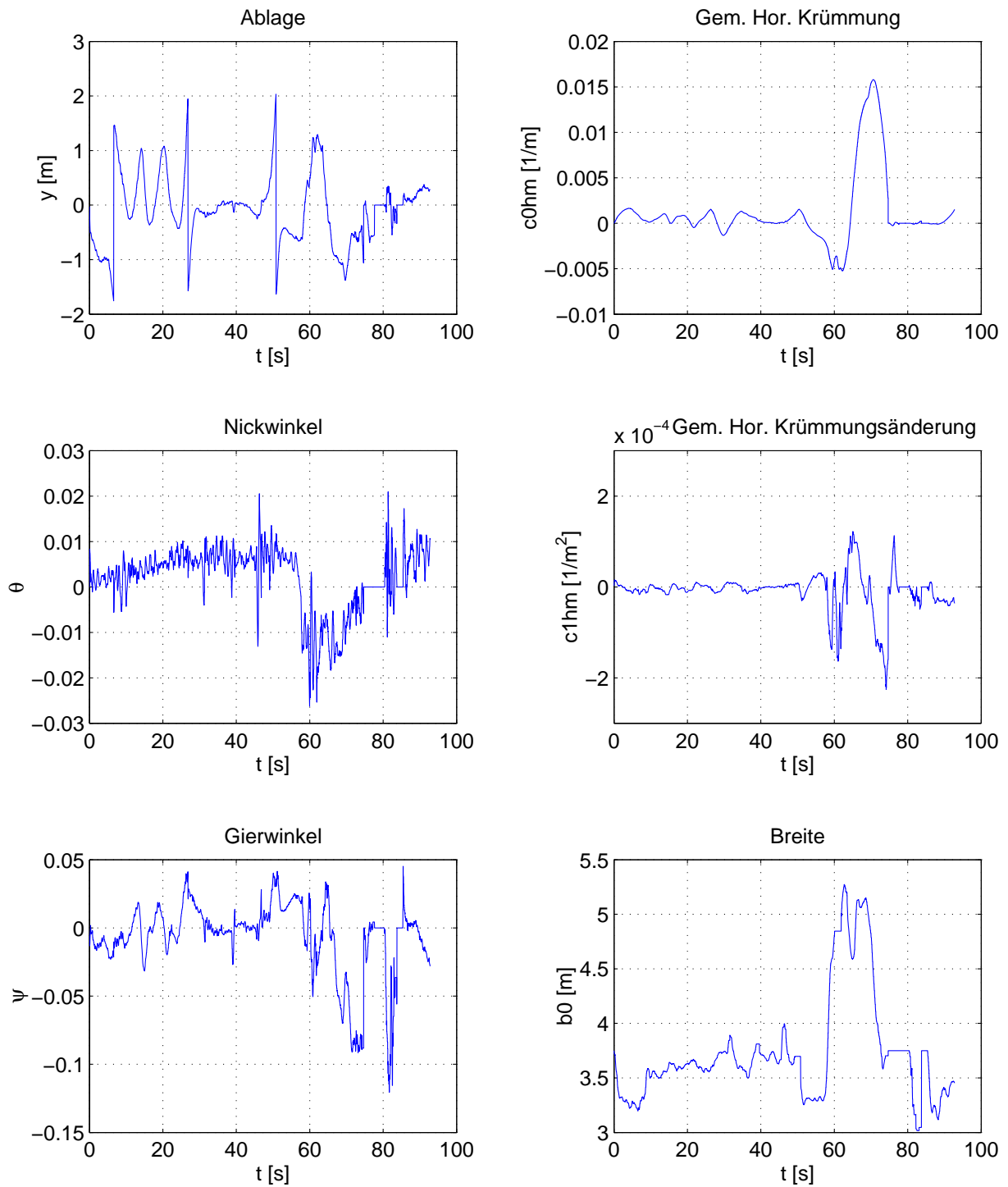


Abbildung 12.3: Zustandsgrößen der Fahrspur im A 39 Szenario

## 12.2 Objekterkennung

Die Objekterkennung soll sowohl rein lidarbasiert als auch lidar- und videobasiert demonstriert werden. Bezüglich der videobasierten Vermessung kam lediglich ein einfaches Quadermodell zum Einsatz, das im wesentlichen auf die Erkennung der Heckansicht von Fremdfahrzeugen optimiert wurde. Im folgenden werden zwei Beispiele gezeigt, für die Erkennung eines PKW sowie eines Kleintransporters. Das Augenmerk in den Betrachtungen sollte auf die stufenweise verfeinerte Erkennung der Objekte gerichtet werden. Dies gilt einerseits für die Annäherung an die Objekte, wobei diese zunächst nur mit einem Strahl des Lidarsensors detektiert werden, dann von mehreren Strahlen erfaßt werden ehe sie schließlich in den Vermessungsbereich des Videosensors eintreten. Je näher die Objekte dem Eigenfahrzeug kommen, desto relevanter sind sie i. a. und desto genauer werden sie in der Folge auch vermessen.

### 12.2.1 Erkennung eines PKW

Die erste Messung gibt eine Folgefahrt hinter einem PKW wieder, welche in dem bereits bei der Fahrspurerkennung geschilderten Autobahnszenario (Abb. 12.1 und 12.2) erfolgte. Die Abbildungen 12.4 und 12.5 stellen die initiale Detektion des PKW mit der Objekt Nummer 82 im Videobild und in der Draufsicht dar. Das Fahrzeug wird vom Lidarsensor in ca. 120 m Entfernung detektiert, bei 80 m Entfernung setzt dann die Vermessung durch den Videosensor ein. Wie den Bildern 12.1 und 12.2 des Fahrspurszenarios zu entnehmen ist, wechselt der PKW in Bild 2 gerade in etwa 100 m Entfernung auf die eigene Fahrspur, fährt dann bis zu Bild 5 als Zielfahrzeug vor dem eigenen Fahrzeug, um dann in Bild 6 auf die linke Spur zu wechseln. Der PKW wird über ca. 40 Sekunden verfolgt, angefangen bei Sekunde 5 bis hin zu Sekunde 45.

Abbildung 12.8 gibt die geschätzten Zustandsgrößen des PKW bei alleiniger Vermessung durch den Lidarsensor wieder. Zunächst ist der geschätzte longitudinale Abstand sowie die longitudinale Geschwindigkeit aufgetragen. Deutlich zu erkennen ist der Annäherungsvorgang an das Fahrzeug und die Folgefahrt in ungefähr konstantem Abstand, ehe das Fahrzeug am Ende auf die linke Spur wechselt, während das Eigenfahrzeug verzögert und ausfährt. Die Longitudinalgeschwindigkeit wird mit 0 m/s initialisiert, um sich dann in wenigen Zyklen auf Werte zwischen 20 und 30 m/s einzuschwingen. Zugrundegelegt wurde hierbei ein Modell mit konstanter Geschwindigkeit. Der laterale Abstand ebenso wie die Gierrate spiegeln im wesentlichen die Fahrspurwechsel des eigenen wie des Fremdfahrzeugs wider. Die Breitenschätzung liefert einen initialen Wert von 1.8-1.9 m, welcher bei direkter dichter Folgefahrt um die reale Breite von ca. 1.7 m schwankt.

Bei schlechteren Aspektbedingungen, d. h. wenn das Fremdfahrzeug sich in größerer Distanz oder aber im Außenbereich des Lidarerfassungsbereichs befindet, schwankt die Schätzung deutlich stärker. Dies wird durch die dann größere horizontale Divergenz der Lidarstrahlen bedingt. Die Höhe des Objekts ist anhand des Lidarsensors nicht zu bestimmen und wird aufgrund der Annahme, daß der Lidarsensor eine gewisse Montagehöhe am Eigenfahrzeug besitzt und darüberhinaus zur Erfassung eines Objekts dieses eine gewisse Mindestgröße

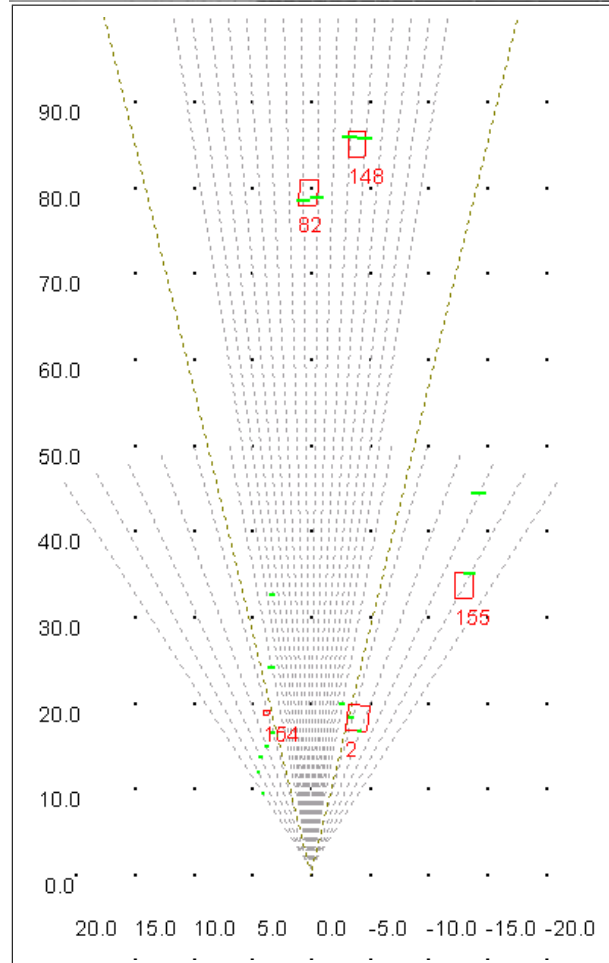
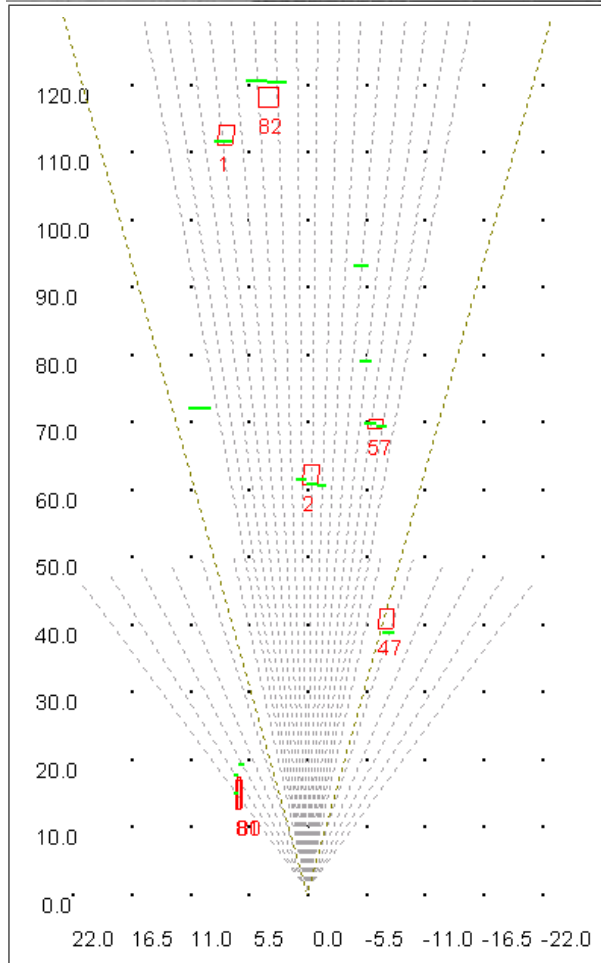


Abbildung 12.4: Erstmalige Detektion des PKW (82) durch den Lidarsensor

Abbildung 12.5: Erstmalige Detektion des PKW (82) durch den Videosensor

hat, zu 1 m initialisiert und bleibt konstant auf diesem Wert. Ein Problem der alleinigen Erfassung durch den Lidarsensor ist in Abb. 12.6 zu erkennen: Das Objekt gerät an den Rand des zentralen Lidar-Erfassungsbereichs und die vom Sensor gelieferten Entfernungspunkte werden mit dem im peripheren Bereich gemessenen Entfernungspunkt der Leitplanke zu einem Segment zusammengefasst. In der Folge schwingt sowohl die Ablagen- als auch die



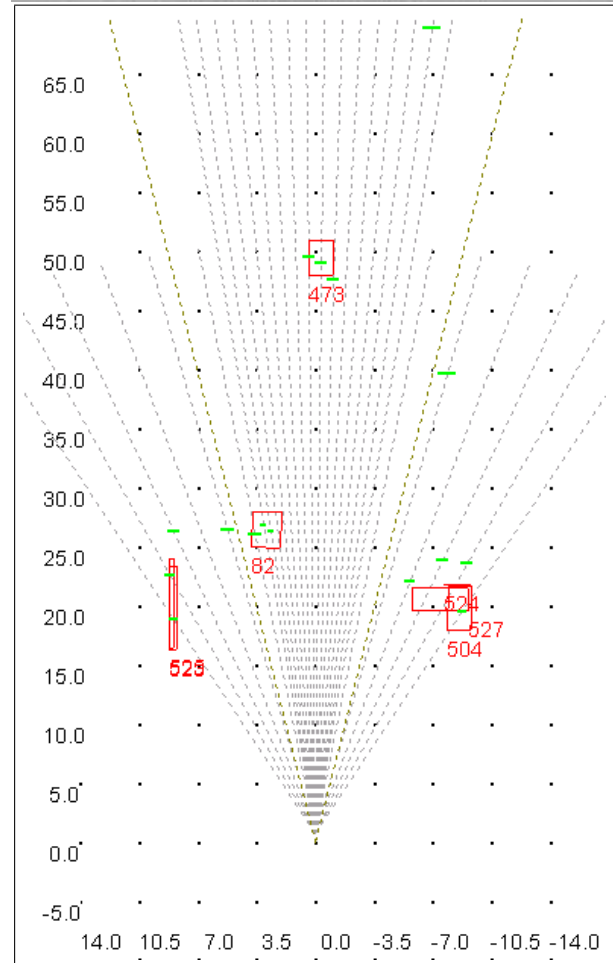
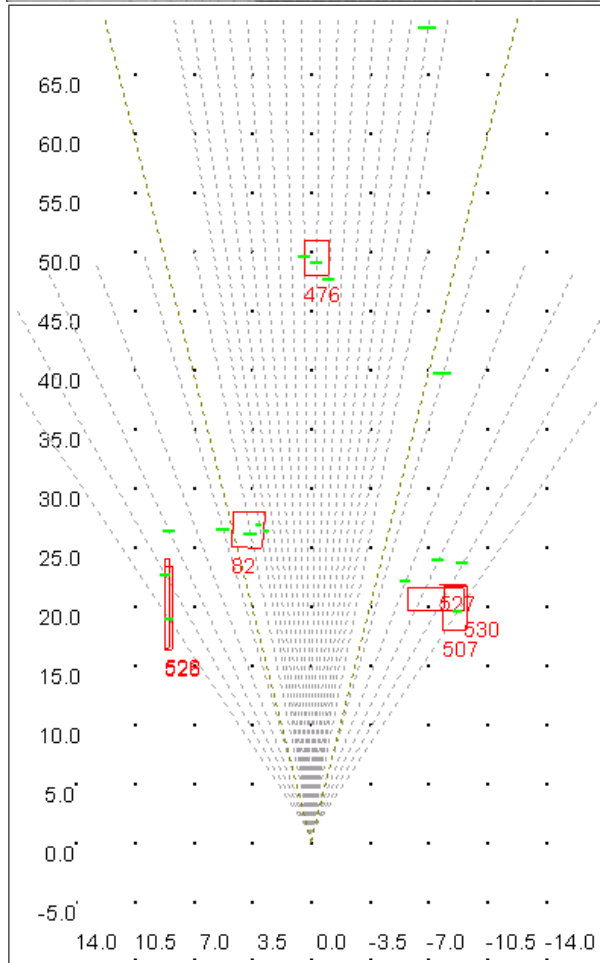
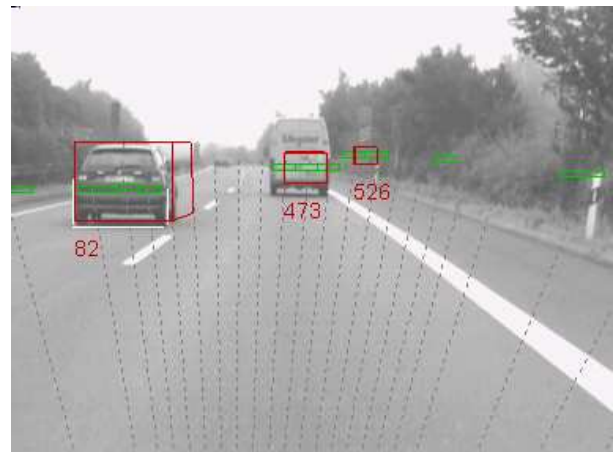
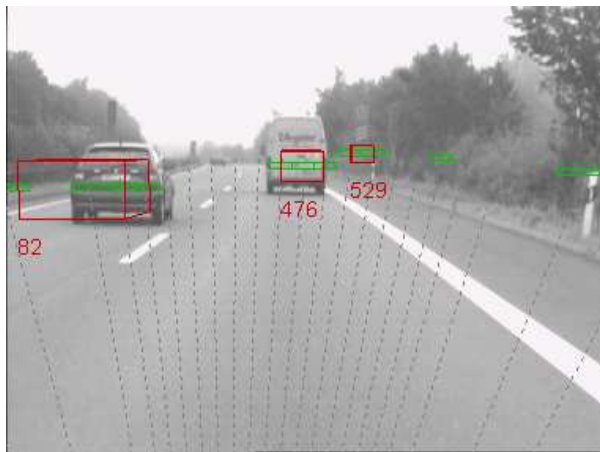


Abbildung 12.6: PKW mit “Fehl“-Segmen-  
tierung bei reiner Lidarvermessung

Abbildung 12.7: PKW mit Stabilisierung der  
Schätzung durch den Videosensor

Breitenschätzung gegen Ende des Erfassungszeitraums.

Abbildung 12.7 zeigt dieselbe Szene bei gleichzeitiger Vermessung mit dem Videosensor. Wie deutlich zu erkennen ist, führt die Vermessung durch den Videosensor zu einer Stabilisierung der Ablagen- und Breitenschätzung auch in dieser Phase des Szenarios. Dies



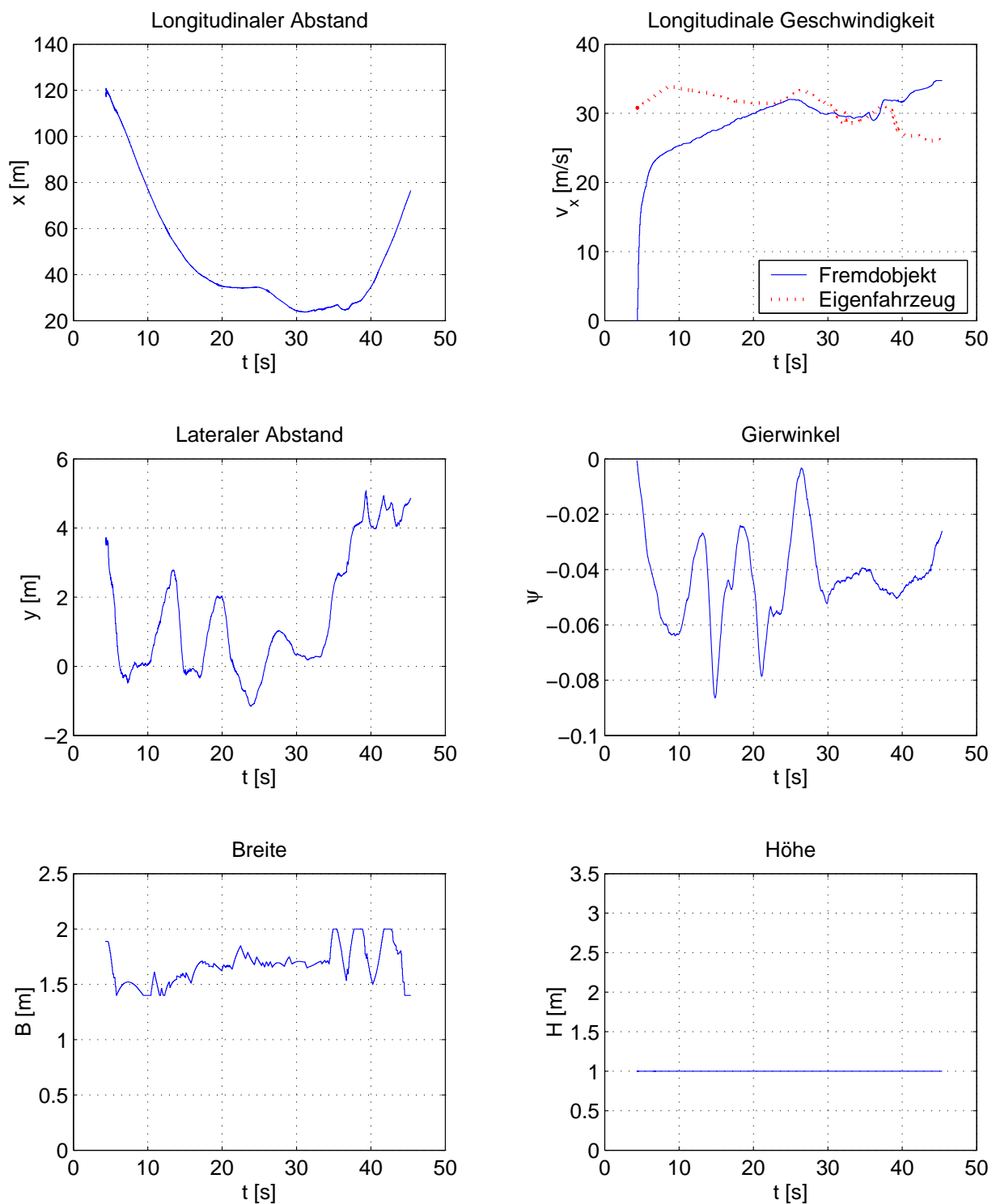


Abbildung 12.8: Zustandsgrößenverlauf des PKW bei reiner Lidarvermessung

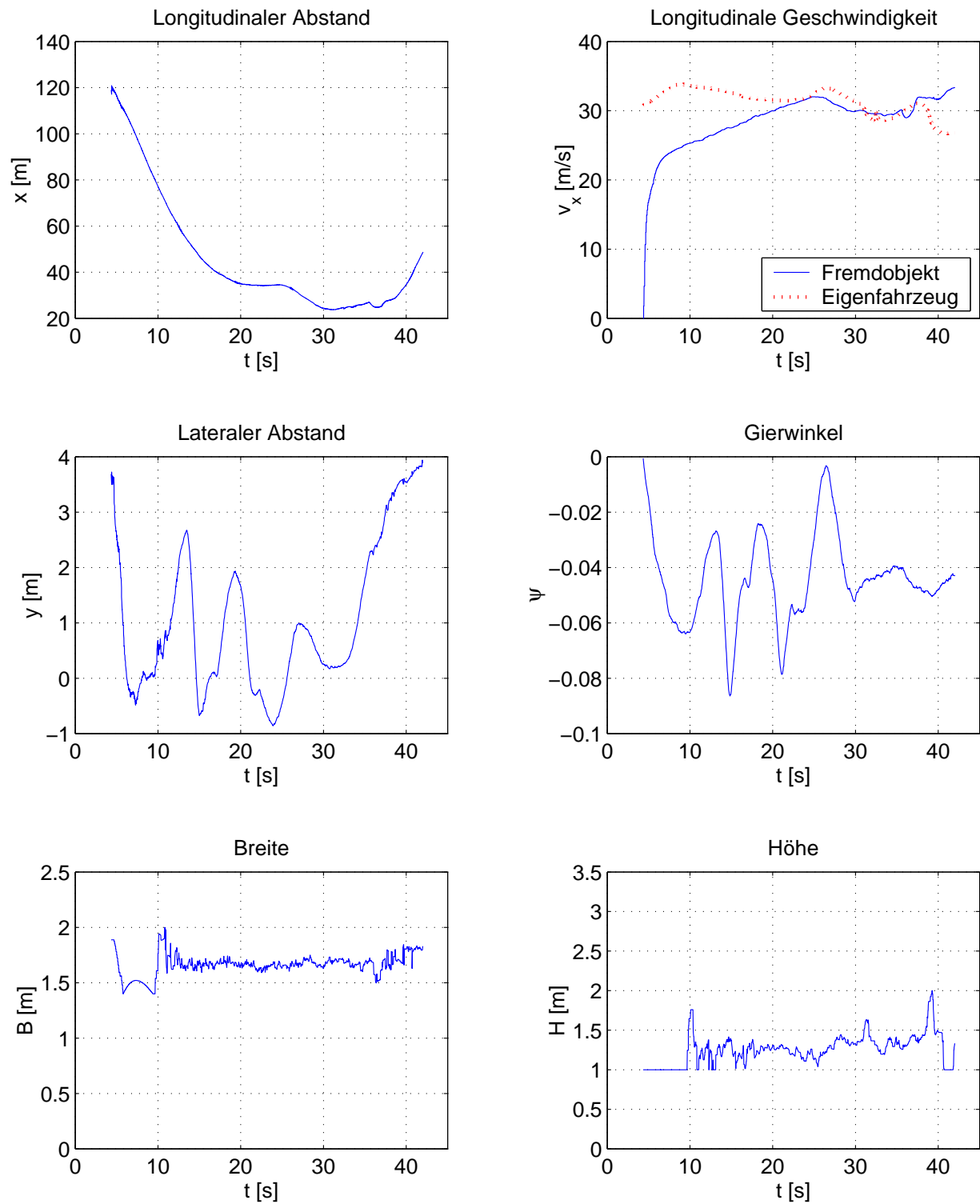


Abbildung 12.9: Zustandsgrößenverlauf des PKW bei Lidar- und Videovermessung

ist auch dem in Abbildung 12.9 zu sehenden Verlauf der Schätzgrößen bei gleichzeitiger Erfassung mittels Lidar- und Videosensorik zu entnehmen. Insgesamt kann man an den Zustandsgrößen, welche auch auf der Basis von Videomeßdaten geschätzt werden, glattere Kurvenverläufe erkennen. Sowohl Longitudinalposition wie auch Longitudinalgeschwindigkeit sind hingegen identisch, da beide aufgrund der schlechten Konditionierung nicht aus Videodaten geschätzt werden - gleiches gilt für die Gierwinkelschätzung des Objekts. Erstmals erfaßbar ist auch die Objekthöhe, wobei diese aufgrund vieler Störkanten und daraus resultierender Fehlzuordnungen noch sehr störungsbehaftet ist und stark schwankt. Gegen Ende der Messung bei  $t = 40$  s wird das Merkmal "Oberkante" kurzfristig sogar verloren, so daß dort der untere Grenzwert des Gültigkeitsintervalls für den Formparameter "Höhe" angenommen wird. Als nachteilig für die Schätzung der Objekthöhe hat sich im Laufe der Untersuchungen auch die Schätzung der Objekthöhe allein aus der oberen Objektkante herausgestellt. Dadurch ist die Schätzung der Höhe bei Nickwinkelschwankungen stark gestört, da sich die Lage der Objektoberkante im Videobild sowohl in Abhängigkeit der Objekthöhe als auch des eigenen Nickwinkels gleichartig ändert. Dieser Einfluß kann durch Schätzung der Objekthöhe aus der Differenz von Objektoberkante und Objektunterkante eliminiert werden, da sich bei einer Nickbewegung beide Kanten im Bild gleichartig verschieben, die Lagedifferenz beider aber unbeeinflusst bleibt.

### 12.2.2 Erkennung eines Kleintransporters

Die zweite Szene stammt aus demselben Autobahn-Szenario wie die PKW-Erkennung; in diesem Fall ist das zu erkennende Fahrzeug ein Kleintransporter. Es handelt sich hierbei um das vorausfahrende Fahrzeug hinter dem - wie in Abschnitt 12.1.2 schon erwähnt - kurz nacheinander zwei abrupte Verzögerungen durchgeführt wurden, die mithin die Zustandschätzung erschweren. Auch in diesem Fallbeispiel sollen wiederum die rein lidarbasierte Erfassung mit der kombinierten lidar- und videobasierten Erkennung gegenübergestellt werden. Zwei beispielhafte Szenen des Erkennungsvorgangs sind in den Abb. 12.10 für die lidarbasierte und 12.11 für die kombinierte Erfassung enthalten. Der Transporter ist in der Bildmitte deutlich zu erkennen; er hat im Bild 12.10 die Objekt Nummer 477 und im Bild 12.11 die Nummer 474, bedingt durch die unterschiedlichen Strategien zur Hypothesenverifikation bzw. Hypothesenverwerfung. Den wesentlichen ins Auge springenden Unterschied macht in erster Linie die durch den Videosensor mögliche Vermessung der Objekthöhe aus, während die Breite auch durch den Lidarsensor scheinbar gut zu erfassen ist. In Bild 12.12 sind zunächst wieder die Schätzergebnisse der rein lidarbasierten Erkennung dargestellt. Man erkennt am longitudinalen Abstand wiederum den Annäherungsvorgang und an der dazugehörigen Längsgeschwindigkeit, welche mit der Eigengeschwindigkeit initialisiert wurde, das Einschwingen auf eine Geschwindigkeit von ca. 24 m/s. Die beiden Nickvorgänge des Eigenfahrzeugs sind als kleine Spitzen bei ungefähr 38-39 s bzw. 46 s zu erkennen. Lateral Abstand und Gierwinkel geben wie gehabt im wesentlichen das eigene Spurwechsel- und Querführungsverhalten wieder. Von den Formparametern ist mittels Lidarsensor nur die Breite zu schätzen, die anfangs mit 1.4 m als PKW initialisiert wird und in der Folge auf ca. 2 m ansteigt.

Vergleicht man hierzu den in Abbildung 12.13 gezeigten Verlauf der Zustandsgrößen bei

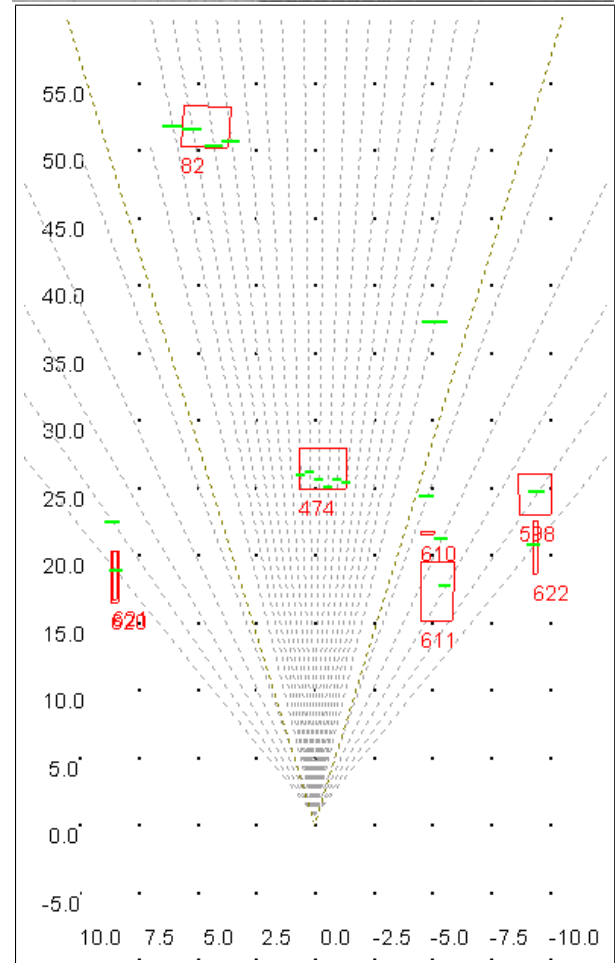
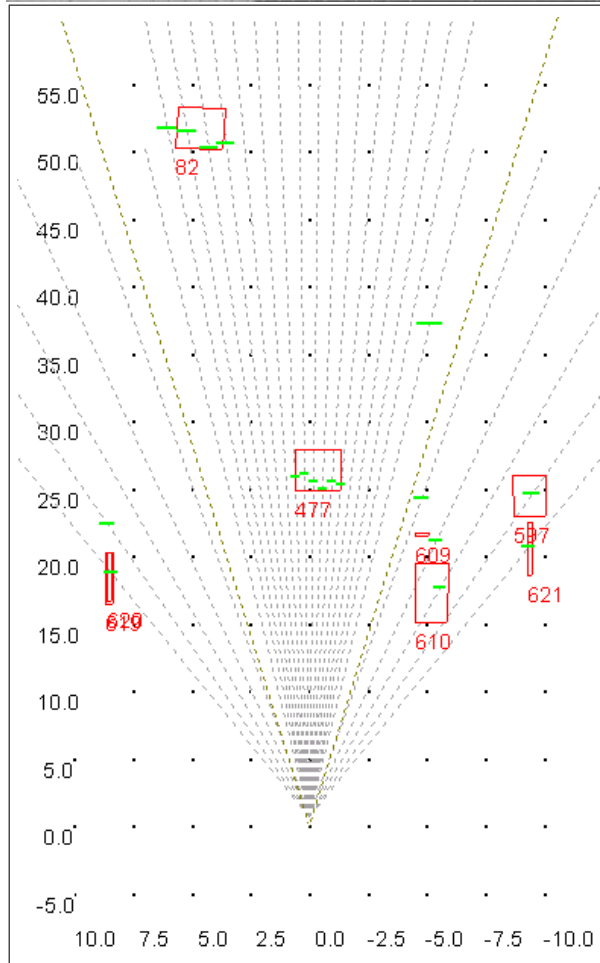
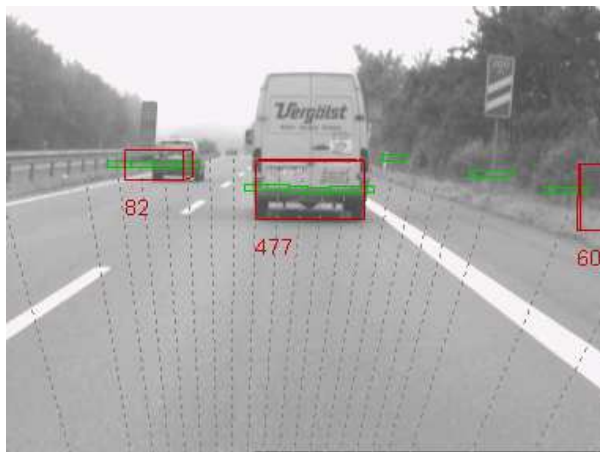


Abbildung 12.10: Erfassung des Transporters durch den Lidarsensor

Abbildung 12.11: Erfassung des Transporters durch Lidar- und Videosensor

kombinierter Lidar- und Videovermessung, so ist die qualitative Änderung der Größen für die Longitudinalbewegung gering, da der Videosensor auch hier keinen Beitrag liefert. Der laterale Abstand hingegen erhält einen deutlich geglätteteren Verlauf. Die Breite springt gleich nach Erfassung durch den Videosensor auf 2 m und bleibt an dieser Grenze. Da dieser Wert die Obergrenze für den Formparameter "Breite" in dieser Klasse darstellt, kann

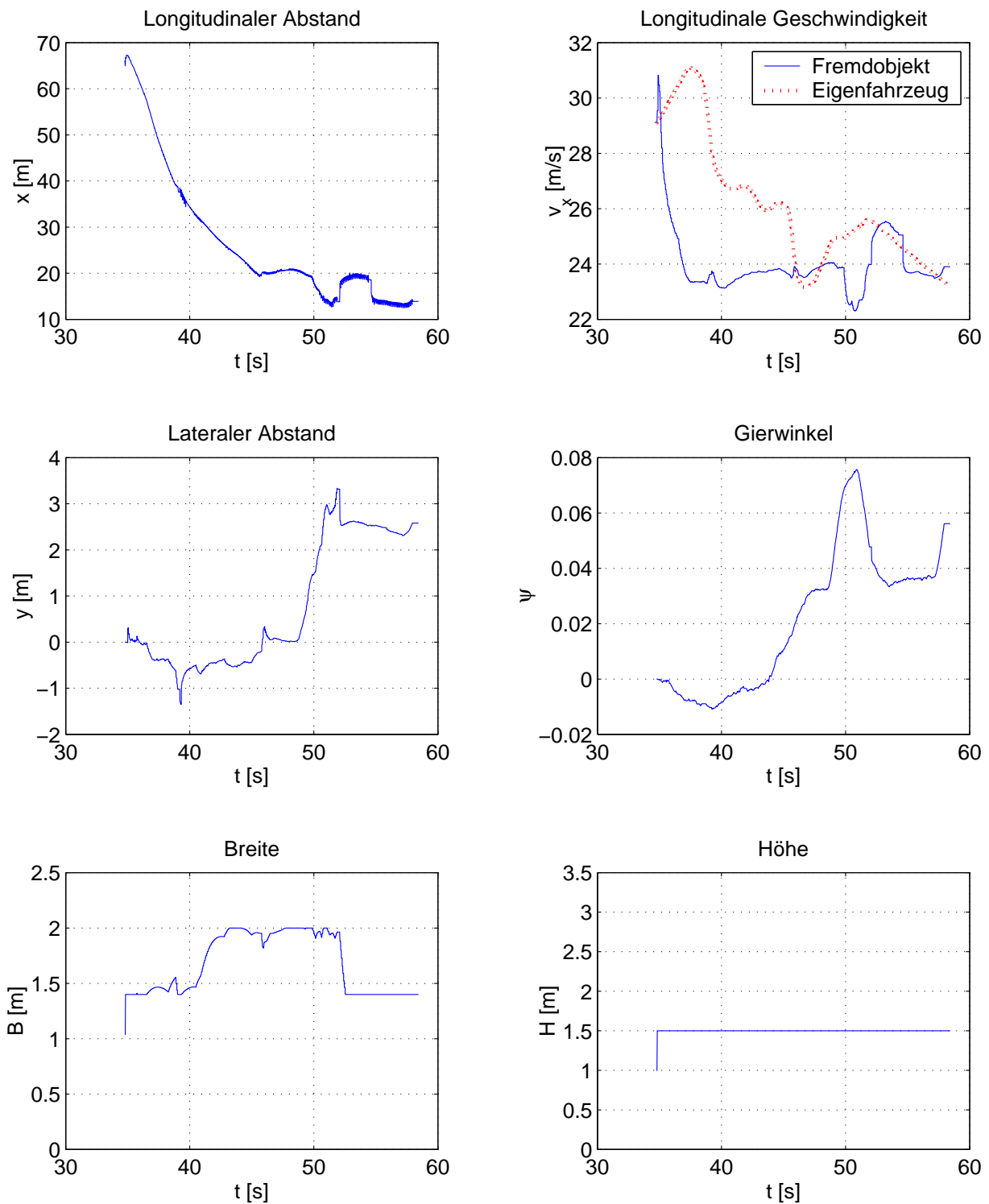


Abbildung 12.12: Zustandsgrößenverlauf des Transporters bei reiner Lidarvermessung

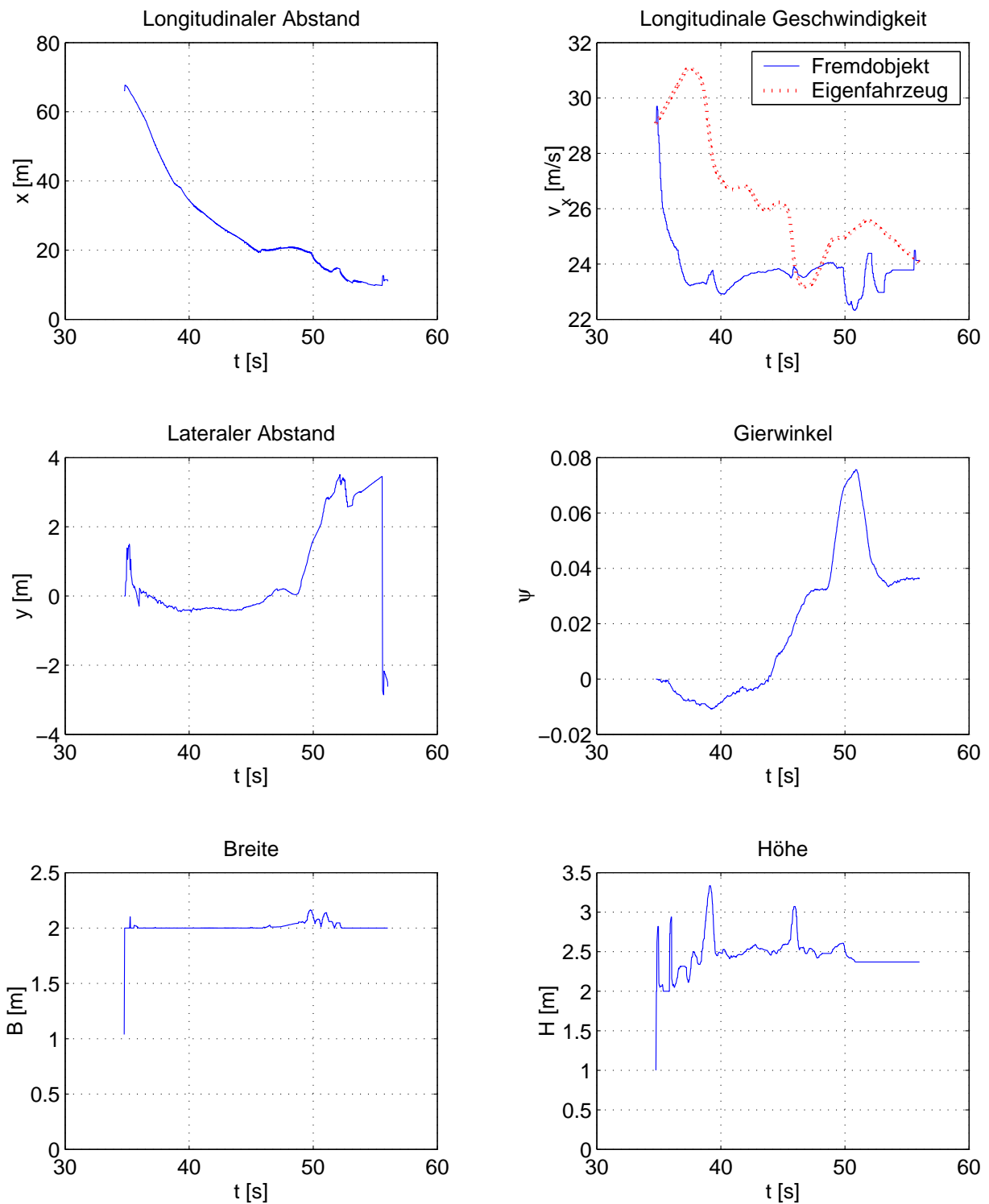


Abbildung 12.13: Zustandsgrößenverlauf des Transporters bei Lidar- und Videovermessung

daraus geschlossen werden, daß dieser Grenzwert offenbar zu klein gewählt worden ist. Die Höhe benötigt im Gegensatz zur Breite eine längere Einschwingzeit. Dies resultiert aus den Aspektbedingungen, die bei zunächst größerem Abstand zum Objekt dazu führen, daß der Hintergrund (Bäume) auch horizontale Kanten oberhalb der Fahrzeugoberkante liefert, die zu einer Fehlschätzung der Höhe führen. Bei dann geringer werdendem Abstand zum Objekt ändern sich die Aspektbedingungen, und der "störende" Hintergrund wird durch den hohen Transporter verdeckt, so daß sich die Schätzung für die Höhe bei 2.5 m einschwingt. Desweiteren sind im Schätzgrößenverlauf für die Fahrzeughöhe die beiden während der Fahrt erfolgten starken Bremsverzögerungen zu erkennen: Durch die hohe Dynamik der Bremsvorgänge konnte die Nickwinkelschätzung auf Basis der Fahrspurmarkierungen nicht mit der gleichen Dynamik folgen; in der Folge wird die aus der Fahrzeugoberkante geschätzte Fahrzeughöhe auf bis zu 3-3.5 m ansteigend geschätzt. Durch eine explizite Schätzung aus der Differenz von Oberkante zu Unterkante ließe sich das Schätzverhalten auch in diesem Fall verbessern.

### 12.2.3 Erkennung von Querverkehr

Eine Kategorie von Szenarien, welche sowohl im Bereich des autonomen Fahrens wie der Fahrerassistenzsysteme derzeit noch kaum betrachtet wird, sind komplexere Szenarien mit Querverkehr wie er auf Kreuzungen innerorts wie ausserorts auftreten kann. Gerade im Bereich der Fahrerassistenz gewinnt die Beherrschung solcher Szenarien im Bereich der Sicherheits- und Warnsysteme zur Kollisionsvermeidung aber an Bedeutung.

Die Fähigkeit, Querverkehr zu erkennen, beinhaltet dabei im wesentlichen auch die Einbringung von Domänenwissen, also des Wissens darüber, welche Objektbewegungen im aktuellen Szenario möglich und welche Wertebereiche für die Objektbewegungen am wahrscheinlichsten sind. Am Beispiel eines städtischen Kreuzungsszenarios soll die Fähigkeit der Beherrschung auch solcher Szenarien demonstriert werden. Zum Einsatz kommt wie im Fall der Verkehrschilderkennung wiederum nur der Lidarsensor, da auch für die sichere Detektion von Querverkehr mittels Bildverarbeitung im Rahmen dieser Arbeit keine Modelle entwickelt wurden. Der Grund liegt in der unter seitlichen Aspektbedingungen noch größeren Variabilität der Objektform, welche die Verwendung gänzlich anderer Merkmalskategorien nahelegt und damit einen ganz eigenen Arbeitspunkt neben der Heck- oder Frontalerkennung darstellt. Das hier betrachtete Querverkehrsszenario ist in den Bildern 12.14 und 12.15 wiedergegeben. In Bild 12.14 ist ein querender PKW von links zu erkennen - im Videobild ist bislang nur die Front zu sehen - der sich in Bild 12.15, welches 0.3 s später aufgenommen wurde, frontal vor dem Eigenfahrzeug befindet. Abb. 12.16 zeigt die zugehörigen Verläufe der geschätzten Zustandsgrößen. In erster Linie von Interesse ist in diesem Fall die Schätzung der Longitudinalgeschwindigkeit. Diese bezieht sich stets auf die Längsachse des Objekts, nicht des Eigenfahrzeugs. Daß die Instanziierung der Hypothese überhaupt in dieser Weise erfolgte, ist zwei Ursachen zu verdanken: Zum einen dem Wissen über die Domäne - abgeleitet u. a. aus dem eigenen Geschwindigkeitsprofil - sowie der Tatsache, daß das Objekt von links in den Erfassungsbereich eintritt und die Ausdehnung des aus den Lidarrohdaten gebildeten Segments einen definierten Wert überschreitet (bezogen auf das Ego-Koordinatensystem). Daraufhin wird der Objektgierwinkel zu  $90^\circ$  initialisiert

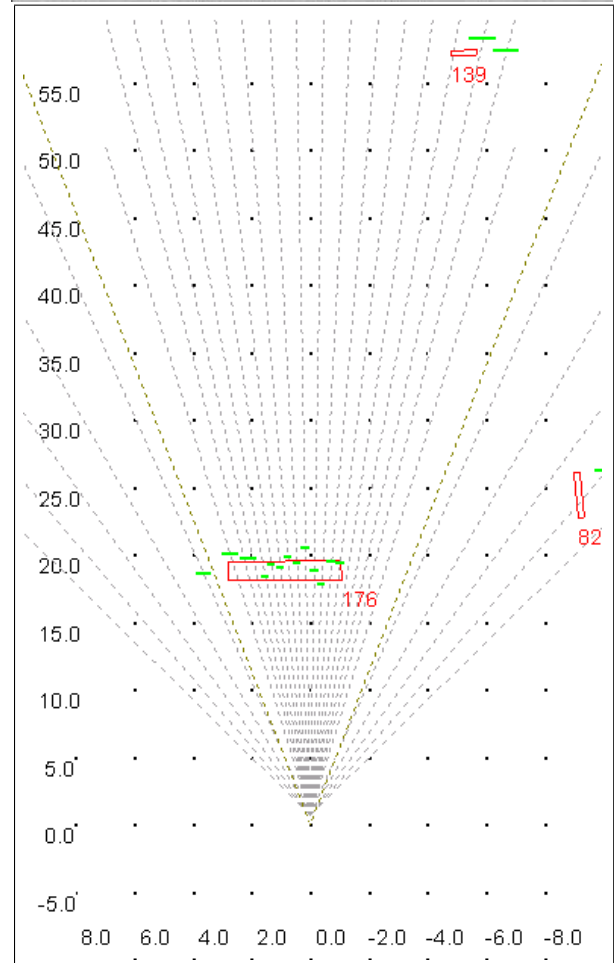
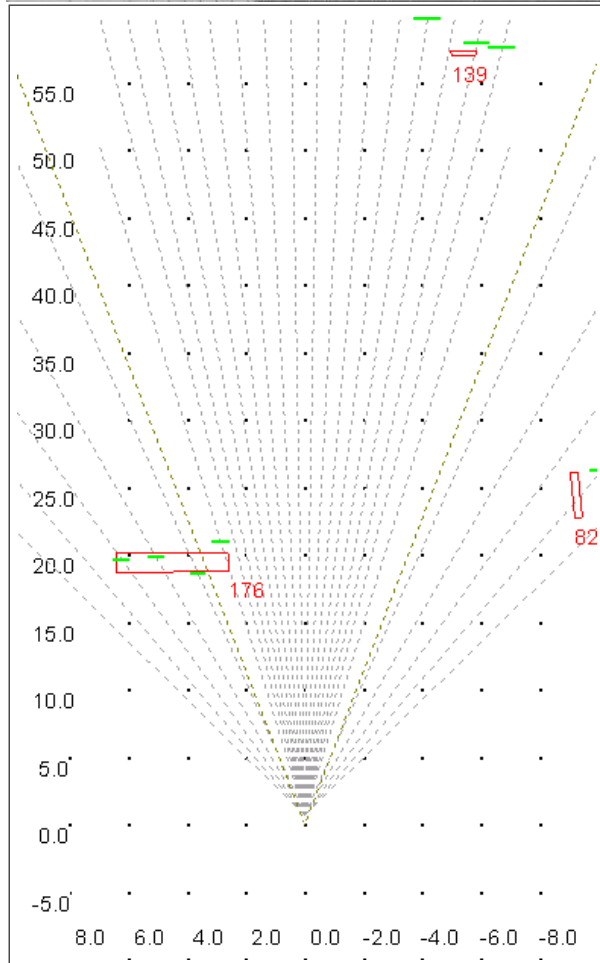
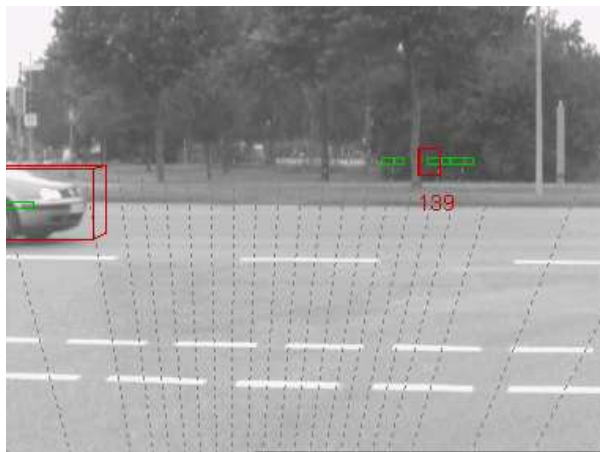


Abbildung 12.14: Querender PKW kurz nach dem Eintritt in den Sichtbereich

Abbildung 12.15: Querender PKW frontal vor dem Eigenfahrzeug

und die Longitudinalgeschwindigkeit mit  $+10$  m/s initialisiert, da das Objekt von links in den Erfassungsbereich einfährt und wir aus der Zykluszeit der Lidarmessungen abschätzen können, in welchem Zeitfenster es in den Erfassungsbereich des Sensors gelangt ist. Wie zu erkennen ist, schwingt sich die Longitudinalgeschwindigkeit auf ungefähr  $+10.7$  m/s ein. Als weitere Zustandsgröße ist die Schätzung der Länge - anstelle der Breite - interessant,



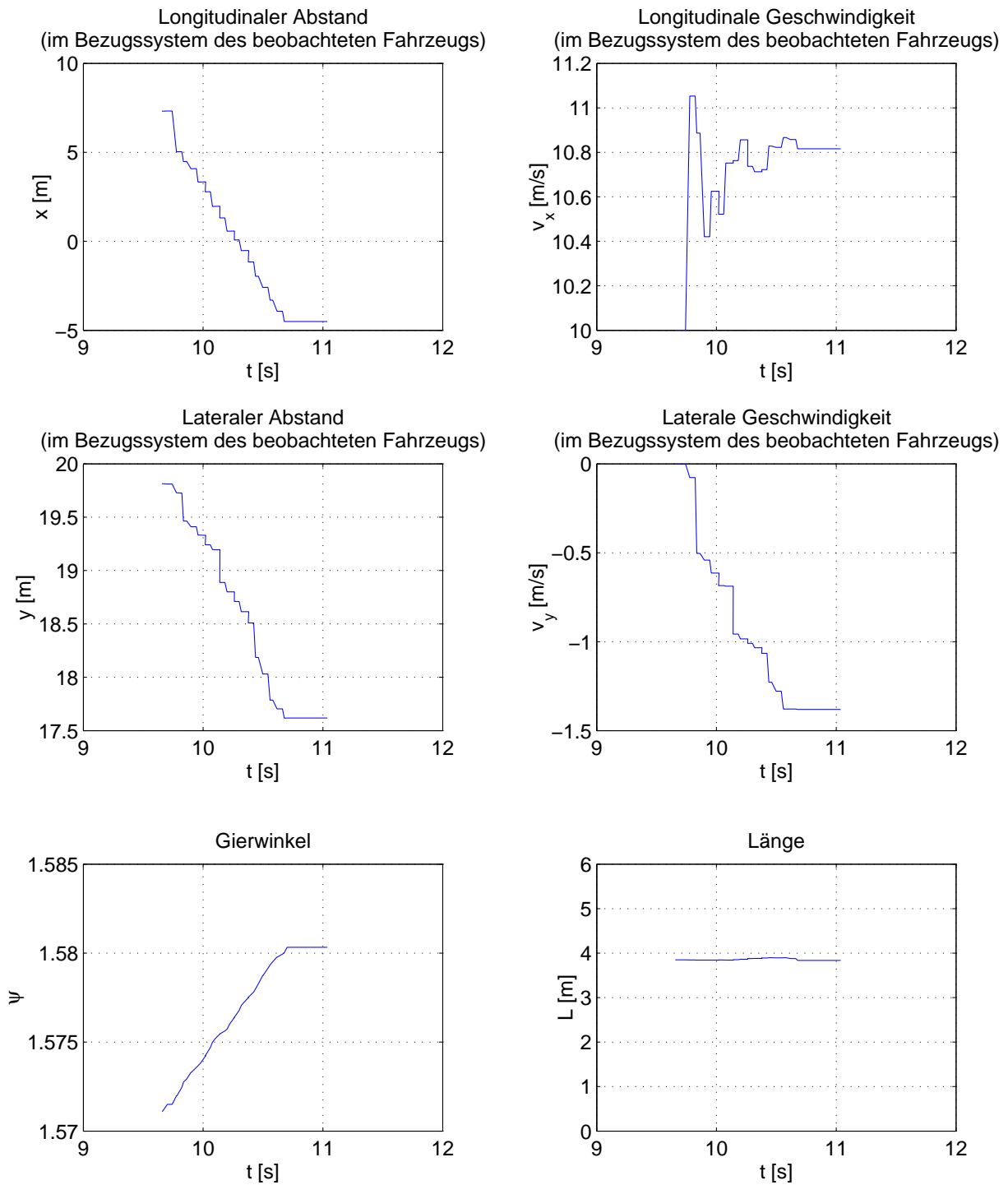


Abbildung 12.16: Zustandsgrößenverlauf eines querenden PKW

die relativ konstant etwas unterhalb von 4.0 m liegt. Der Verlauf der weiteren Zustandsgrößen ist wenig überraschend, bis auf die abnehmende laterale Position und in Verbindung damit einer leicht negativen Lateralgeschwindigkeit. Die Ursache hierfür liegt zum einen darin begründet, daß die Kreuzung keinen vollkommen rechtwinkligen Verlauf besitzt, zum anderen aber auch am Übergang vom zentralen in den peripheren Erfassungsbereich. Verbunden mit diesem Übergang kommt es aufgrund der unterschiedlichen Empfindlichkeiten der Strahlen in beiden Bereichen zu kleinen Differenzen in den gemessenen Entfernungen, die sich über eine Verschiebung der gebildeten Segmente letztlich in der Schätzung niederschlagen.

#### 12.2.4 Vergleichende Diskussion der Ergebnisse

Die in den vorangegangenen Abschnitten gezeigten Ergebnisse haben die Kombination aus Videosensor und Lidarsensor in verschiedenen Fahrumgebungen und in bezug auf unterschiedliche Objektkategorien beleuchtet.

Als wesentliche Schwäche der Sensoren läßt sich bzgl. des Lidarsensors zunächst die geringe Entfernungsgenauigkeit sowie die in der Entfernung und in den peripheren Bereichen prinzipbedingt geringe Winkelauflösung nennen. Für den Einsatz im Rahmen von ACC-Applikationen - für welche diese Art Sensor eigentlich konzipiert ist - sind diese Einschränkungen nicht entscheidend, es sei denn man bewegt sich vom klassischen ACC-Geschwindigkeitsbereich in den Niedergeschwindigkeitsbereich und hin zu Stausituationen. Eine weitere Schwäche, die auch für ACC-Applikationen relevant ist, ist die Empfindlichkeit hinsichtlich Nickbewegungen des Eigenfahrzeugs aufgrund der nur in einer horizontalen Ebene liegenden Strahlen und deren geringer Divergenz. Durch eine Kombination mit inertialen Sensordaten ließe sich diese Schwachstelle aber beheben - solange der Sensor die relevanten Objekte noch erfassen kann.

Als Schwäche des monokularen Videosensors in Verbindung mit den hier betrachteten Auswerteverfahren zählt ganz klar die Objekterkennung. Sofern diese - wie in dieser Arbeit - modellbasiert angelegt ist, werden in der Regel nur solche Objekte erkannt, die sich in Form eines Quaders beschreiben lassen. Aber auch für diese repräsentierten Objekte gelten Einschränkungen, insbesondere die longitudinale Entfernungs- und Geschwindigkeitsschätzung betreffend. Weitere kritische Punkte betreffen die Detektion von Objekten, die Initialisierung von Objekthypothesen sowie die Reichweite und den Erfassungsbereich des Sensors. Letztere stellen - bedingt durch den Einsatz nur einer Kamera - einen Kompromiß dar: Zur Erfassung von nahen Objekten sowie der Fahrspur ist ein großer Erfassungswinkel erstrebenswert; um auch weit entfernte Objekte noch mit ausreichender Auflösung im Bild zu erhalten bedarf es hingegen einer großen Brennweite mit dementsprechend geringerem Öffnungswinkel. Darüberhinaus besteht natürlich die Möglichkeit durch Hinzunahme weiterer Auswerteverfahren und Merkmale, wie beispielsweise flächenbasierter Informationen (Helligkeitsverläufe, Texturen, Farbe, ...), die Objekterkennung insgesamt robuster zu gestalten.

Betrachtet man die Messungen unter diesen Gesichtspunkten, so wird zunächst einmal deutlich, daß allein die Hinzunahme einer Videokamera zur Fahrspurerkennung für eine

lidarbasierte Objekterkennung einen großen Fortschritt bedeutet. Die Fahrspurinformation ermöglicht eine weitergehende Szeneninterpretation, z. B. das Aufsetzen von Hypothesen in Abhängigkeit des Szenenbereichs (Fahrbahn - nicht Fahrbahn). (Ganz davon abgesehen, daß die eigentlichen auf der Wahrnehmung aufsetzenden Funktionen durch die Fahrspurinformation einen deutlichen Informationsgewinn haben.) Umgekehrt kann die videobasierte Fahrspurerkennung auch von den Lidarmeißwerten profitieren, wenn der Lidarsensor zur Fahrbahninfrastruktur zählende Objekte wie die Begrenzungspfosten oder Leitplanken detektiert. Zwei Beispiele für die Detektion von Leitplanken sind in Abb. 12.17 dargestellt, in der eine Szene aus dem bekannten Autobahnszenario zu sehen ist. In der Abfahrt hat die videobasierte Fahrspurerkennung aufgrund der hohen Krümmungen, der starken Spuraufweitung, der ungünstigen Aspektbedingungen und der daraus resultierenden hohen Anzahl an Störmerkmalen große Schwierigkeiten. Wie in der Abbildung zu erkennen ist, ist die an solchen Stellen oft vorhandene Leitplanke vom Lidarsensor sehr gut vermeßbar und kann somit zur Stützung der Fahrspurhypothese herangezogen werden.

Bezüglich der Objekterkennung lassen die Meßergebnisse eine leichte Verbesserung der Schätzung von Breite und Ablage erkennen, die sich (Abschnitt 12.2.1 insbesondere bei der Erfassung von Nebenspurobjekten bemerkbar machen kann. Die lediglich graduelle Verbesserung ist auf der einen Seite darauf zurückzuführen, daß die Auflösung im zentralen Erfassungsbereich des Lidarsensors innerhalb der Erfassungsreichweite des Videosensors ebenfalls relativ hoch ist, zum anderen ermöglicht das verwendete Quadermodell keine so gute Erfassung, wie dies detailliertere Modelle (Fahrzeug-Polyedermodell) gestatten - auf Kosten einer höheren Rechenleistung.

Ein wesentlicher Punkt, der hier in Ergänzung zu den Ergebnissen erwähnt werden sollte, besteht in der Klassifikation der Objekte, die mit der modellbasierten Bildverarbeitung immanent verbunden ist. Sie gestattet auch dann Aussagen über eine Objektklasse zu tätigen, wenn andere Sensoren Probleme haben, z. B. im Fall stehender Objekte. Aufgrund der nicht vorhandenen Objektgeschwindigkeit haben Sensoren mit eingeschränkten Klassifikationsmöglichkeiten wie Lidar und insbesondere Radar in diesen Situationen Schwierigkeiten, die relevanten Fahrzeuge und sonstigen nicht funktionsrelevanten Objekte voneinander zu diskriminieren, was zu Funktionseinschränkungen führt (ACC).

Der wesentlichste Punkt, der für eine Kombination beider Sensoriken spricht, liegt jedoch in der Komplementarität der Sensoreigenschaften begründet. Der Lidarsensor hat seine Stärken im Bereich der Detektion und Verfolgung beliebiger Objekte und er besitzt - in der gegebenen Sensoranordnung - eine größere Reichweite als die Videokamera. Die Videokamera liefert die für die Szenendeutung wichtige Fahrspurinformation und kann durch Tracking und Klassifikation von Objekten im Nahbereich, die für gewöhnlich eine große Relevanz besitzen, einen Beitrag leisten. Damit gehen beide Sensoren eine Symbiose ein, die sich vornehmlich auf der Ebene der Szeneninterpretation auswirkt, auf der sich Möglichkeiten ergeben, die sicher noch weit über das in dieser Arbeit Dargestellte hinausgehen.

Den Kern des hier vorgestellten Ansatzes bildet allerdings die integrale Repräsentation und Interpretation beider hier verwendeten Sensoren. Bei üblichen modularen Ansätzen, die gerade im Bereich Fahrerassistenz verbreitet sind, interpretiert und repräsentiert jeder

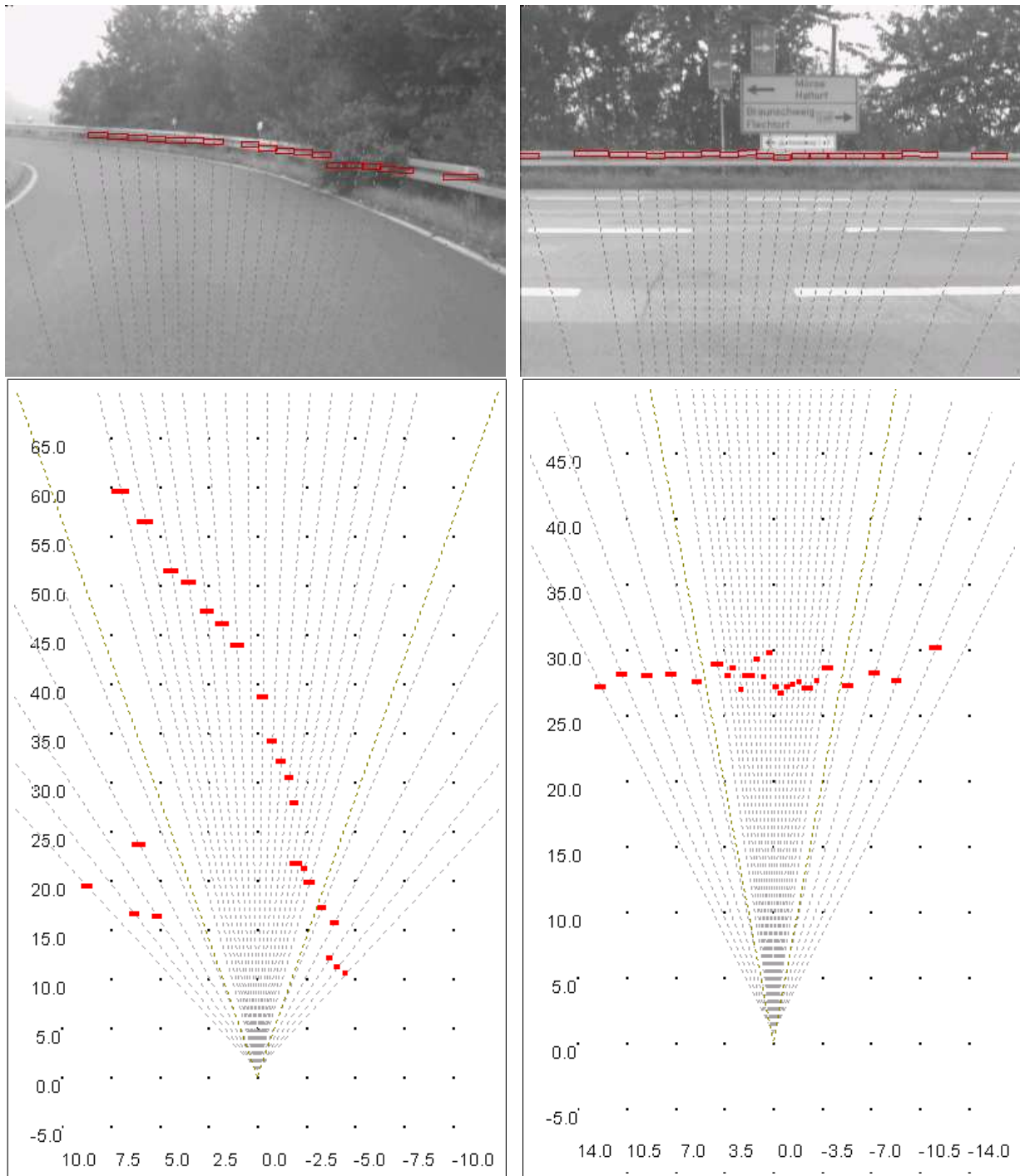


Abbildung 12.17: Vermessung der Fahrbahninfrastruktur mit dem Lidarsensor am Beispiel von Leitplanken

Sensor isoliert für sich die Szene. Die Ergebnisse werden dann nachgelagert in einem separaten Fusionsschritt nochmals gefiltert und interpretiert. Im Gegensatz dazu ist in dieser Arbeit nur **eine** für **alle** Sensoren gültige Interpretationsschicht vorhanden, die Daten werden nur **einmal** gefiltert und es bestehen weitreichende Möglichkeiten der **Kopplung** von Sensoren und Sensorinformationen auf **verschiedenen** Ebenen.

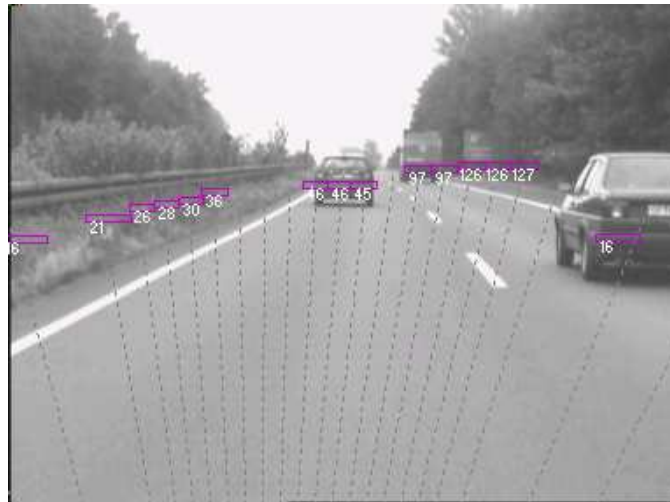


Abbildung 12.18: Lidarrohdaten mit den zugehörigen Entfernungswerten als Basis der Segmentierung des Videobildes

Der Ansatz in dieser Arbeit besteht in der Informationsfusion der Sensoren auf der 4D-Objektebene: Bei Eintreffen neuer Daten eines der Sensoren werden diese separat entsprechend des rekursiven Ansatzes der Szeneninterpretation in das interne Szenenmodell integriert. Eine direkte Kopplung der Sensordaten ist hier nicht betrachtet worden. Eine Möglichkeit hierzu besteht darin, die Meßdaten des Lidarsensors direkt als Unterstützung in der Segmentierung des Videobildes zu nutzen. Die Lidarmeßwerte schneiden eine “Scan“-Ebene“ aus der Bildebene heraus, für die sie an den Strahlen Entfernungswerte für Flächenausschnitte liefern. Die Größe dieser Flächenausschnitte ist nicht genau bekannt, läßt sich aber aus der gemessenen Entfernung und der bekannten horizontalen wie vertikalen Divergenz des Laserstrahls abschätzen. In Abb. 12.18 ist eine Fahrszene dargestellt, bei der unterhalb der eingetragenen Entfernungsrrohdaten des Lidars der zugehörige Entfernungswert (in [m]) eingetragen ist. Vorstellbar ist ein Verfahren, bei dem auf der Basis der Lidarmeßdaten eine erwartungsbasierte Segmentierung der Lidar-Meßebeene durchgeführt wird. Anhand zweier benachbarter Lidarmeßwerte kann eine räumliche Verbindung (3D) in der “Scan“-Ebene berechnet und anschließend ins Videobild projiziert werden. Im Videobild kann dann nach Strukturen gesucht werden, die entweder parallel oder orthogonal zu dieser projizierten “Scan“-Ebene verlaufen. Parallele Strukturen stützen die Hypothese eines Segmentes aus beiden Entfernungspunkten, während orthogonale Strukturen daraufhindeuten, daß eine Segmentgrenze zwischen beiden Lidarmeßwerten liegt. Alternativ hierzu ließen sich anstelle der Lidarrohdaten auch die daraus im Vorfeld ermittelten Lidarsegmente verwenden, oder aber eine gleichzeitige Segmentierung der Lidar- wie der Videodaten durchführen.

# Kapitel 13

## Zusammenfassung und Ausblick

In dieser Arbeit wurde ein Ansatz vorgestellt, welcher eine enge Kopplung zweier Sensoren unterschiedlicher Sensortechnologien zur Erfassung der Fahrumgebung zum Gegenstand hat. Bei den Sensoren handelte es sich zum einen um eine monochrome Videokamera und zum anderen um einen 28-kanaligen Mehrstrahllaser. Im Gegensatz zu anderen Ansätzen, bei denen die Sensordatenauswertung isoliert voneinander - gestützt auf unterschiedliche, verteilte Wissensbasen - erfolgt, wurden in dieser Arbeit die Sensorrohdaten durch ein integrales System zur multisensoriellen Fahrumgebungserfassung interpretiert. Der Vorteil dieses Verfahrens ist einerseits in der zentralen Wissensbasis zur Interpretation zu sehen, andererseits in den Möglichkeiten einer engeren und flexibleren Kopplung der Sensoren. So wurden hier beispielsweise die durch den Videosensor gelieferten Fahrspurdaten dazu genutzt, die Interpretation der Objekterkennung durch den Lidarsensor intelligent zu steuern. Eine weitere Möglichkeit, die im Rahmen dieser Arbeit noch nicht ausgeschöpft wurde, bildet z. B. die direkte Nutzung der Entfernungsdaten des Lidarsensors zur robusteren Segmentierung des Videobildes. Jenseits dieser Vorteile einer engeren Kopplung beider Sensoren und der integralen Szeneninterpretation konnte anhand der experimentellen Beispiele gezeigt werden, daß der Verbund beider Sensoren aufgrund deren komplementärer Stärken eine sinnvolle Kombination darstellt: Der Mehrstrahllaser kann mit hoher Zuverlässigkeit auch in der Entfernung und im peripheren Blickfeld beliebige Objekte detektieren und tracken. Die Videokamera hingegen kann einerseits als einziger Sensor die für die Deutung der Gesamtszene wichtige Fahrspurinformation liefern und kann andererseits die vom Lidarsensor detektierten Objekte verifizieren, eine Feinvermessung der Abmessungen durchführen oder durch weitergehende Erkennungsverfahren aufgrund der photometrischen Eigenschaften auch klassifizieren und identifizieren.

Als Basis für die Szeneninterpretation diente der über Jahre in der Praxis bewährte 4D-Ansatz nach Dickmanns, der für die Einbindung des Lidarsensors transparent erweitert wurde. Zur Videoauswertung wurden konturbasierte Meß- und Gruppierungsverfahren vorgestellt. Die an die Vermessung anschließende Assoziation der gemessenen mit den erwarteten Meßwerten bediente sich eines auf der Systematik der Hopfield-Netze basierenden relationalen Zuordnungsverfahrens. Für den Mehrstrahllaser wurde ein einfaches Segmentierungsverfahren beschrieben, welches die Einzelmeßwerte auf benachbarten Strahlen gruppiert. Bewegungs- und Formgrößen wurden mittels rekursiver Schätzverfahren bestimmt.

Das dargestellte Verfahren wurde in ein Programmsystem in C++ auf einem PC umgesetzt und zum Nachweis der Tragfähigkeit an mehreren exemplarischen Szenarien getestet. Die experimentellen Ergebnisse verdeutlichen das Potential des Verfahrens sowie die Vielfalt beherrschbarer Szenen. Sowohl in Autobahnszenarien, in innerstädtischen Kreuzungsszenarien, zur Erkennung von Fahrzeugen, wie auch anderer Objekte liefert die gewählte Sensorkombination tragfähige Resultate.

Der in dieser Arbeit gezeigte Ansatz vermittelt jedoch nur einen ersten Einblick in die Möglichkeiten, die eine enge Kopplung und integrale Interpretation ermöglicht. Durch den Ausbau der Erkennung hin zur Klassifikation und die stärkere Interaktion zwischen der Dateninterpretation der Sensoren (s. o.: Segmentierung) ergeben sich vielfältige Möglichkeiten die Anforderungen zu erfüllen, welche an die Fahrumgebungserfassung hinsichtlich Zuverlässigkeit, Robustheit und Redundanz gestellt werden. Um diese Möglichkeiten voll ausschöpfen zu können, sind jedoch auch Modifikationen vorzunehmen: Zum einen erfordert eine stärkere Interaktion auf Rohdatenbasis eine möglichst zeitsynchrone Erfassung des Videobildes und der Lidarmessung, zum anderen würde eine feinere Auflösung der Laserdaten durch Verwendung eines Laserscanners - wie sie an anderer Stelle schon betrieben wird - eine Möglichkeit bieten, die Erkennungsleistung zu steigern.

Teil IV

Anhang





# Anhang A

## Homogene Koordinaten

Homogene Koordinaten sind aus der Computergraphik bekannt, wo sie seit langem zur effizienten Berechnungen von Koordinatentransformation dienen. Die Hauptvorteile liegen zum einen in der Durchgängigkeit sowohl der rotatorischen wie der translatorischen Transformationen in Form von Matrixmultiplikationen und zum anderen in der gleichfalls als Matrixoperation durchführbaren Projektion.

Bei homogenen Koordinaten werden die Ortsvektoren  $\underline{p} = [x, y, z]^T$  um einen Skalierungsfaktor  $s$  ergänzt zu:

$$\underline{p} = \begin{bmatrix} x \\ y \\ z \\ s \end{bmatrix} \quad (\text{A.1})$$

Um homogene Ortsvektoren vergleichen zu können, müssen diese zunächst auf einen gemeinsamen Skalierungsfaktor normiert werden (meist “1”)<sup>1</sup>.

Rotation  $R_x$  um die x-Achse um den Wankwinkel  $\Phi$ :

$$R_x(\Phi) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos \Phi & -\sin \Phi & 0 \\ 0 & \sin \Phi & \cos \Phi & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (\text{A.2})$$

Rotation  $R_y$  um die y-Achse um den Nickwinkel  $\Theta$ :

$$R_y(\Theta) = \begin{bmatrix} \cos \Theta & 0 & \sin \Theta & 0 \\ 0 & 1 & 0 & 0 \\ -\sin \Theta & 0 & \cos \Theta & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (\text{A.3})$$

---

<sup>1</sup>Durch setzen des Skalierungsfaktors zu “0” lassen sich auf einfache Weise auch Richtungsvektoren transformieren.

Rotation  $R_z$  um die z-Achse um den Gierwinkel  $\Psi$ :

$$R_z(\Psi) = \begin{bmatrix} \cos \Psi & -\sin \Psi & 0 & 0 \\ \sin \Psi & \cos \Psi & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (\text{A.4})$$

Translation  $T$  um den Vektor  $\underline{t} = [t_x, t_y, t_z]^T$ :

$$T(\underline{t}) = \begin{bmatrix} 1 & 0 & 0 & t_x \\ 0 & 1 & 0 & t_y \\ 0 & 0 & 1 & t_z \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (\text{A.5})$$

Skalierung  $S$  mit dem Vektor  $\underline{s} = [s_x, s_y, s_z]^T$ :

$$S(\underline{s}) = \begin{bmatrix} s_x & 0 & 0 & 0 \\ 0 & s_y & 0 & 0 \\ 0 & 0 & s_z & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (\text{A.6})$$

Projektion  $P$  mit der Brennweite  $f$ :

$$P(f) = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & f & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix} \quad (\text{A.7})$$

Seien 2 Koordinatensysteme  $K_1$  und  $K_2$  gegeben. Die Transformationen beschreiben die  $K_2$  aus der Sicht von  $K_1$ . Dann transformiert man einen in  $K_2$  gegebenen Punkt  $P$  durch folgende Konkenation der obigen Transformationsmatrizen in den Punkt  $P'$ , der in  $K_1$  gegeben ist:

$$P' = R_x R_y R_z T P \quad (\text{A.8})$$

# Anhang B

## Rekursive Schätzverfahren

Die Zustandsraumdarstellung ebenso wie die Darstellung des Kalman-Filter-Algorithmus erfolgt in diskreter Zeit, da einerseits die in dieser Arbeit beschriebenen Signale und ihre Verarbeitungsschritte in Digitalrechnern sämtlich diskreter Natur sind. Andererseits lag das Filterverfahren von Kalman ebenfalls zunächst im diskreten Bereich vor, ehe es später in Zusammenarbeit mit Bucy auch auf den kontinuierlichen Bereich ausgedehnt wurde.

### B.1 Zustandsraumdarstellung

In der Zustandsraumdarstellung wird das zeitliche Verhalten eines dynamischen Systems durch einen Satz von Differentialgleichungen für den Systemzustand  $\underline{x}$  beschrieben. Diese lassen sich - mit der Vereinbarung  $t_k = k$  - in der Form

$$\underline{x}_{k+1} = f_k(\underline{x}_k, \underline{u}_k) + \underline{w}_k \quad (\text{B.1})$$

mit dem Zustandsvektor  $\underline{x}$ , dem auf das System wirkenden Steuervektor  $\underline{u}$  und dem (unbekannten) Systemrauschen  $\underline{w}$  darstellen.

In dieser Arbeit werden nur lineare, zeitvariante oder zeitinvariante Systeme der Form

$$\underline{x}_{k+1} = \underline{\Phi}_k \underline{x}_k + \underline{\Gamma}_k \underline{u}_k + \underline{w}_k \quad (\text{B.2})$$

mit der Systemmatrix  $\underline{\Phi}$  und der Steuermatrix  $\underline{\Gamma}$  betrachtet.

Die nichtlineare Meßgleichung für ein diskretes System lautet

$$\underline{y}_k = g_k(\underline{x}_k) + \underline{v}_k \quad (\text{B.3})$$

mit dem Meßvektor  $\underline{y}$  und dem (unbekannten) Meßrauschen  $\underline{v}$ . Im Falle einer nichtlinearen Abbildung  $g()$ , wie sie beispielsweise bei der nichtlinearen perspektivischen Projektion von 3D-Merkmalen auf die 2D-Bildebene entstehen, erfolgt eine Linearisierung um die Nominaltrajektorie mit dem Ergebnis

$$\Delta \underline{y}_k = \underline{C}_k \Delta \underline{x}_k + \underline{v}_k \quad (\text{B.4})$$

mit

$$\underline{C}_k = \left. \frac{\partial}{\partial \underline{x}} g(\underline{x}) \right|_{\underline{x}=\underline{x}_k} \quad (\text{B.5})$$

als der Jacobischen Meßmatrix.

## B.2 Kalman-Filter

Die in den Gl. B.1/B.2 und B.3/B.4 berücksichtigten Rauschterme  $\underline{w}$  und  $\underline{v}$  machen es unmöglich, den “wahren” Zustand eines dynamischen Systems zu bestimmen. Hinter den Rauschtermen verbergen sich nicht allein Rauschprozesse im engeren Sinn, sondern auch Abweichungen zwischen dem realen System und dem gewählten Systemmodell.

Grundvoraussetzung für die Rekonstruktion des Systemzustands aus den Messungen ist die Beobachtbarkeit (nach [Kalman 60]) des Systems. Aufgrund der auftretenden Störungen läßt sich der Systemzustand nicht eindeutig bestimmen; hierzu dienen vielmehr Schätzverfahren, welche auf der Basis von Modell- und Optimalitätsannahmen einen besten Schätzwert für den Systemzustand  $\underline{x}$  liefern. Den Grundstein dieser Verfahren legte C. F. Gauß im Jahr 1801 mit seiner Methode der kleinsten Fehlerquadrate im Rahmen der Bestimmung von Planetenbahnen. Das Kalman-Filter kann als eine Verallgemeinerung der Gauß’schen Ausgleichsrechnung auf lineare bzw. um einen Arbeitspunkt linearisierte dynamische Systeme angesehen werden. Auf der Basis der Modellannahmen läßt sich das Filter als (rekursive) Folge eines Prädiktions- und eines Innovationsschrittes darstellen.

### Voraussetzungen

Die Rauschterme  $\underline{w}$  und  $\underline{v}$  in den Gl. B.1/B.2 und B.3/B.4 werden als weiß, gaußverteilt, mittelwertfrei sowie als zeitlich untereinander und mit dem Anfangszustand unkorreliert angenommen.

$$E\{\underline{w}_k\} = \underline{0} \quad (\text{B.6})$$

$$E\{\underline{v}_k\} = \underline{0} \quad (\text{B.7})$$

$$E\{\underline{w}_k \underline{w}_j\} = \underline{0} \quad (\text{B.8})$$

$$E\{\underline{v}_k \underline{v}_j\} = \underline{0} \quad (\text{B.9})$$

$$E\{\underline{w}_k \underline{x}_0\} = \underline{0} \quad \forall k \neq j \quad (\text{B.10})$$

$$E\{\underline{v}_k \underline{x}_0\} = \underline{0} \quad \forall k \neq j \quad (\text{B.11})$$

$$E\{\underline{w}_k \underline{v}_j\} = \underline{0} \quad (\text{B.12})$$

$$E\{\underline{w}_k \underline{w}_k^T\} = \underline{Q}_k \quad (\text{B.13})$$

$$E\{\underline{v}_k \underline{v}_k^T\} = \underline{R}_k \quad (\text{B.14})$$

$$(\text{B.15})$$

## Prädiktion

Ausgehend von einem vorhandenen besten Schätzwert  $\hat{x}$  zur Zeit  $k$  erfolgt eine Vorhersage des Zustandsvektor  $\underline{x}^*$  für die Zeit  $k + 1$  über

$$\underline{x}_{k+1}^* = \Phi_k \hat{x}_k \quad (\text{B.16})$$

und der Schätzfehlerkovarianz  $\underline{P}^*$  über

$$\underline{P}_{k+1}^* = \Phi_k \hat{P}_k \Phi_k^T + \underline{Q}_k \quad (\text{B.17})$$

Damit sind der a-priori Schätzwert für den Zustand  $\underline{x}^*$  und die zugehörige Schätzfehlerkovarianz  $\underline{P}^*$  vor dem Eintreffen der Messung zum Zeitpunkt  $k + 1$  gegeben. Durch das Systemrauschen  $\underline{Q}$  wächst die Schätzunsicherheit an. Über die Meßgleichung (s. Gl. B.3 und Gl. B.4) erhält man ohne Berücksichtigung des unbekannten Meßrauschens  $v$  die prädierten Referenzwerte  $\underline{y}^*$  für die Messung

$$\Delta \underline{y}_{k+1}^* = \underline{C}_{k+1} \Delta \underline{x}_{k+1}^* \quad (\text{B.18})$$

mit  $\underline{C}_{k+1}$  als der um die prädierte Nominaltrajektorie linearisierten Jacobi'schen Meßmatrix

$$\underline{C}_{k+1} = \left. \frac{\partial}{\partial \underline{x}} g(\underline{x}) \right|_{\underline{x}=\underline{x}_{k+1}^*}. \quad (\text{B.19})$$

Die Abbildung der prädierten Schätzfehlerkovarianz in den Meßbereich, dem sich additiv das Meßrauschen überlagert, ergibt die sog. Innovationskovarianz  $\underline{V}$

$$\underline{V}_{k+1} = \underline{C}_{k+1} \underline{P}_{k+1}^* \underline{C}_{k+1}^T + \underline{R}_{k+1} \quad (\text{B.20})$$

Die Innovationskovarianz gibt die erwartete Streuung der Meßwerte  $\underline{y}$  um den prädierten Referenzwert  $\underline{y}^*$  an. Aufgrund ihrer zentralen Rolle bei der Meßwertauswahl und -zuordnung insbesondere in der Umfelderkennung wird später noch auf sie eingegangen.

## Innovation

Nach dem Eintreffen der Messungen  $\underline{y}$  zum Zeitpunkt  $k + 1$  erfolgt die Innovation, die eine Verbesserung der Schätzung in Abhängigkeit des Informationsgehalts der Messungen bedeutet. Der innovierte Zustand  $\hat{x}$  ergibt sich zu

$$\hat{x}_{k+1} = \underline{x}_{k+1}^* + \underline{K}_{k+1} \overbrace{(\underline{y}_{k+1} - \underline{y}_{k+1}^*)}^{\gamma_{k+1}} \quad (\text{B.21})$$

sowie die Schätzfehlerkovarianz  $\hat{P}$  zu

$$\hat{P}_{k+1} = \underline{P}_{k+1}^* - \underline{K}_{k+1} \underline{C}_{k+1} \underline{P}_{k+1}^* = (\underline{I} - \underline{K}_{k+1} \underline{C}_{k+1}) \underline{P}_{k+1}^* \quad (\text{B.22})$$

Darin ist  $\underline{K}$  die Kalman-Filter Verstärkungsmatrix, die sich in der Form

$$\underline{K}_{k+1} = \underline{P}_{k+1}^* \underline{C}_{k+1}^T \underline{V}_{k+1}^{-1} \quad (\text{B.23})$$

schreiben läßt.

Wie man erkennt, spiegelt die Filter-Matrix  $\underline{K}$  im wesentlichen das Verhältnis von Schätzunsicherheit zu Meßunsicherheit wider. Über die Meßmatrix  $\underline{C}$  geht die Beobachtbarkeit des Systems und damit der Informationsgehalt der Messungen ein. Aus dem Verhältnis folgt die Gewichtung von Messung zu prädictiertem Modell. Je geringer die Meßunsicherheit, je besser die Beobachtbarkeit und je höher die Systemunsicherheit, desto größer wird der Einfluß der Messung sein und vice versa.

# Literaturverzeichnis

- [Ballard, Brown 82] D. H. Ballard, C. M. Brown, *Computer Vision*, Prentice-Hall, Englewood Cliffs, N. J. , 1982
- [Bar Shalom, Fortman 88] Y. Bar Shalom, T. E. Fortman, *Tracking and Data Association*, Mathematics in Science and Engineering, Academic Press, 1988
- [Bardeen, Brattain, Shockley 47] J. Bardeen, W. Brattain, W.B. Shockley, *U.S. Patente 2,502,488; 2,524,035*
- [Batavia et. al. 97] P.H. Batavia, D. Pommerleau, C. Thorpe, *Overtaking Vehicle Detection Using Implicit Optical Flow*, Proc. IEEE Conference on Intelligent Transportation Systems ITSC, Boston, 1997
- [Batavia et. al. 98] P.H. Batavia, D. Pommerleau, C. Thorpe, *Predicting Lane Position for Roadway Departure Prevention*, Proc. IEEE International Conference on Intelligent Vehicles, pp. 5-10, Stuttgart, 1998
- [Becker 01] J. Becker, *Fusion der Daten der objekterkennenden Sensoren eines autonomen Straßenfahrzeugs*, Dissertation, Technische Universität Braunschweig, Fakultät für Maschinenbau und Elektrotechnik, 2001
- [Behringer et. al. 92] R. Behringer, V. v. Holt, D. Dickmanns, *Road and relative ego-state recognition*, Proc. International Symposium on Intelligent Vehicles, Detroit, 1992
- [Behringer 96] R. Behringer, *Visuelle Erkennung und Interpretation des Fahrbahnverlaufs durch Rechnersehen für ein autonomes Straßenfahrzeug*, Dissertation, Universität der Bundeswehr München, Fakultät für Luft- und Raumfahrttechnik, 1996
- [Bertozzi et. al. 97] M. Bertozzi, A. Broggi, G. Conte, A. Fascioli, *Obstacle and Lane Detection on ARGO*, Proc. IEEE Conference on Intelligent Transportation Systems ITSC, Boston, 1997
- [Bertozzi et. al. 98a] M. Bertozzi, A. Broggi, A. Fascioli, *An extension to the Inverse Perspective Mapping to handle non-flat roads*, Proc. IEEE International Conference on Intelligent Vehicles, pp. 305-310, Stuttgart, 1998
- [Bertozzi et. al. 98b] M. Bertozzi, A. Broggi, A. Fascioli, *Experiments using MMX-based processors for real-time image processing on the ARGO vehicle*, Proc. IEEE International Conference on Intelligent Vehicles, pp. 505-510, Stuttgart, 1998
- [Bertsekas et. al. 93] D. P. Bertsekas, D. A. Castanon, H. Tsaknakis, *Reverse Auction and the Solution of Asymmetric Assignment Problems*, *SIAM J. on Optimization*, Vol. 3, 1993, pp. 268-299



- [Bertsekas 01] D. P. Bertsekas, *Auction Algorithms*, Encyclopedia of Optimization, Kluwer, 2001
- [Biermann 77] G. J. Biermann, *Factorization Methods for Discrete Sequential Estimation*, Academic Press, 1977
- [Brammer, Siffling 75] K. Brammer, G. Siffling, *Kalman-Bucy-Filter - Deterministische Beobachtung und stochastische Filterung*, Oldenbourg, Wien, 1975
- [Brockhaus] F. A. Brockhaus, *Lexikon in 20 Bänden*
- [Broggi et. al. 99] A. Broggi, M. Bertozzi, A. Fascioli, *The 2000 km Test of the ARGO Vision-Based Autonomous Vehicle*, IEEE Intelligent Systems, pp. 55-64, 1999
- [Brooks 81] R. A. Brooks, *Symbolic Reasoning among 3-D Models and 2-D Images*, Artificial Intelligence, 17, 1981
- [Brüdigam 94] C. Brüdigam, *Intelligente Fahrmanöver sehender autonomer Fahrzeuge in autobahnähnlicher Umgebung*, Dissertation, Universität der Bundeswehr München, Fakultät für Luft- und Raumfahrttechnik, 1994
- [Burkhard 00] H.-D. Burkhard, *Software-Agenten*, in Handbuch der KI, Oldenbourg, München, Wien, 2000
- [Canny 86] J. Canny, *A Computational Approach to Edge Detection*, PAMI 8, pp. 679-698, 1986
- [Caveney et.al. 02] D. Caveney, D. Dickmanns, J.K. Hedrick, *Vor dem Regler: Ein Verfahren zur Mehrfach-Objektverfolgung für Fahrerassistenzsysteme (Before the Controller: A Multiple Target Tracking Routine for Driver Assistance Systems)*, at Automatisierungstechnik, Volume 51, 5/2003, Oldenbourg, 2003
- [Choi et. al. 95] D. Choi, S.Y. Oh, K.I. Kim, *Fitness-Based Modular Vision Steering Architecture for Super Cruise Control of Automobiles*, IEEE Intelligent Vehicles, pp. 170-175, Detroit, 1995
- [Coué et. al. 02] C. Coué, T. Fraichard, P. Bessière, E. Mazer, *Multi-Sensor Data Fusion Using Bayesian Programming: an Automotive Application*, Proc. IEEE International Conference on Intelligent Vehicles, Detroit, 2002
- [Cugiani, Giubbolini 98] C. Cugiani, L. Giubbolini, *Millimetre-Wave Imaging Radar in the Near Field for Anti-Collision, A Multi-DSP-Architecture for the Real-Time 3-D Image Reconstruction*, 31st International Symposium on Automotive Technology and Automation, Düsseldorf, 1998
- [Darms, Winner 04] M. Darms, H. Winner, *Eine Sensorfusionsarchitektur zur Fusion von Umfelddaten*, Tagung "Aktive Sicherheit durch Fahrerassistenz", TÜV Süddeutschland, München 11.-12.3.2004

- [Dickmanns 80] E.D. Dickmanns, *Untersuchungen und Arbeitsschritte zum Thema künstliche Intelligenz: Rechnersehen und -steuerung dynamischer Systeme*, Interner Bericht, Institut für Systemdynamik und Flugmechanik, Universität der Bundeswehr München, 1980
- [Dickmanns, Zapp 86] E.D. Dickmanns, A. Zapp, *A curvature-based scheme for improving road vehicle guidance by computer vision*, Mobile Robots, Vol. 727, SPIEE, pp. 161-168
- [Dickmanns 87] E.D. Dickmanns, *4D-Szenenanalyse mit integralen raum-/zeitlichen Modellen*, DAGM Symposium Mustererkennung, Braunschweig, 1987
- [Dickmanns, Graefe 88a] E.D. Dickmanns, V. Graefe, *Applications of dynamic monocular machine vision*, Machine Vision Applications 1, pp. 241-261
- [Dickmanns, Graefe 88b] E.D. Dickmanns, V. Graefe, *Dynamic monocular machine vision*, Machine Vision Applications 1, pp. 223-241
- [Dickmanns, Mysliwetz 92] E.D. Dickmanns, B. Mysliwetz, *Recursive 3-D state and relative ego-state recognition*, Pattern Analysis and Machine Intelligence 14(2), pp. 199-213
- [Dickmanns et. al. 93] E.D. Dickmanns, R. Behringer, C. Brüdigam, D. Dickmanns, F. Thomanek, V. v. Holt, *An all-transputer visual autobahn-autopilot/copilot*, Proc. ICCV, Berlin, 1993
- [Dickmanns et. al. 94] E.D. Dickmanns, R. Behringer, D. Dickmanns, T. Hildebrandt, M. Maurer, J. Schiehlen, F. Thomanek, *The seeing passenger car VaMoRs-P*, International Symposium on Intelligent Vehicles, Paris, 1994
- [Dickmanns 99] E.D. Dickmanns, *An Expectation-based, Multi-focal, Saccadic (EMS) Vision System for Vehicle Guidance*, 9th International Symposium on Robotics and Research, 1999
- [D. Dickmanns 94] D. Dickmanns, *KRONOS Benutzerhandbuch*, Universität der Bundeswehr München, 1994
- [D. Dickmanns 97] D. Dickmanns, *Rahmensystem für visuelle Wahrnehmung veränderlicher Szenen durch Computer*, Dissertation, Universität der Bundeswehr München, 1997
- [Doi et. al. 97] Doi, Ayumu, T. Yoshioka, T. Butsuen, *Introduction of the Mazda Advanced Safety Vehicle*, 5th ITS World Congress, Berlin, 1997
- [Duda, Hart 72] R. O. Duda, P. E. Hart, *Use of the Hough Transformation to Detect Lines and Curves in Pictures*, ACM 15, pp. 1-15, 1972
- [Eberl 87] G. Eberl, *Automatischer Landeanflug durch Rechnersehen*, Dissertation, Universität der Bundeswehr München, Fakultät für Luft- und Raumfahrttechnik, 1987
- [Eibert et.al. 98] M. Eibert, H. Hopfmüller, C. Schaefer, K.R. Schulz, *Laserkamera zur 3D-Vermessung*, Dornier GmbH, 1998

- [Enkelmann 97] W. Enkelmann, *Robust Obstacle Detection and Tracking by Motion Analysis*, IEEE Conference on Intelligent Transportation Systems ITSC, Boston, 1997
- [Fagerer 96] C. Fagerer, *Automatische Teleoperation eines Tracking- und Greifvorgangs im Weltraum basierend auf Bilddatenauswertung*, Dissertation, Universität der Bundeswehr München, Fakultät für Luft- und Raumfahrttechnik, Fortschrittsberichte VDI Reihe 8, Nr. 541, 1996
- [Föllinger] O. Föllinger, *Regelungstechnik*, 6. Auflage, Hüthig, Heidelberg, 1990
- [Forgeber 98] E. Forgeber, *High Resolution 3D-Camera*, Daimler-Benz Aerospace, Bremen, 1998
- [Franke et. al. 96] U. Franke, I. Kutzbach, *Fast Stereo based Object Detection for Stop&Go Traffic*, IEEE Conference on Intelligent Transportation Systems, pp. 339-344, Tokyo, 1996
- [Franke et. al. 97] U. Franke, S. Görzig, F. Lindner, D. Mehren, F. Paetzold, *Steps Towards an Intelligent Vision System for Driver Assistance in Urban Traffic*, IEEE Conference on Intelligent Transportation Systems ITSC, pp. 601-605, Boston, 1997
- [Franke et. al. 98a] U. Franke, D. Gavrilu, S. Görzig, F. Lindner, F. Paetzold, C. Wöhler, *Autonomous Driving Goes Downtown*, IEEE Intelligent Systems & their applications, 13(6), pp. 40-48, 1998
- [Franke et. al. 98b] U. Franke, D. Gavrilu, S. Görzig, F. Lindner, F. Paetzold, C. Wöhler, *Bildverstehen im innerstädtischen Verkehr*, Autonome Mobile Systeme, pp. 37-44, Springer, Berlin, 1998
- [Fürst 97] S. Fürst, *Cronos Benutzerhandbuch*, Universität der Bundeswehr München, 1997
- [Fuerstenberg et. al. 02] K.C. Fuerstenberg, K.C.J. Dietmayer, V. Willhoeft, *Pedestrian Recognition in Urban Traffic using a vehicle based Multilayer Laserscanner*, Proc. IEEE International Conference on Intelligent Vehicles, Paris, 2002
- [Gehrig, Franke 03] S.K. Gehrig, U. Franke, *Intersection Assistance Fusing Image, Map, and GPS information*, Workshop Fahrerassistenzsysteme, Leinsweiler, 2003
- [Gelb 84] A. Gelb, *Applied optimal estimation*, The M.I.T. Press, Cambridge, 1984
- [Giubbolini 97] L. Giubbolini, *3-D Full Range Millimeter Imaging Technique based on stationary holographic Radar and Pulse Compression*, 6th European Congress Lightweight and Small Cars, pp. 881-890, 1997
- [Görzig 03] S. Görzig, *Eine generische Software-Architektur für Multi-Agentensysteme und ihr Einsatz am Beispiel von Fahrerassistenzsystemen*, Dissertation, Universität Stuttgart, 2003

- [Goldbeck et. al. 98] J. Goldbeck, G. Dräger, B. Hürtgen, S. Ernst, F. Willms, *Lane Following Combining Vision and DGPS*, Proc. IEEE International Conference on Intelligent Vehicles, pp. 451-456, Stuttgart, 1998
- [Graefe 84] V. Graefe, *Two Multi-Processor Systems for Low-Level Real-Time Vision*, in: J.M. Brady, L.A. Gerhardt, H.F. Davidson: Robotics and Artificial Intelligence, Springer Verlag, 1984, pp. 301-308
- [Gregor et. al. 97] R. Gregor, S. Baten, V. v. Holt, M. Lützel, M. Maurer, A. Rieder, E. D. Dickmanns, *A low-cost vision system for automotive applications*, Proc. 30th ISATA, Florence, Italy, June, 16-19, 1997, pp. 195-201
- [Gregor 02] R. Gregor, *Fähigkeiten zur Missionsdurchführung und Landmarkennavigation*, Dissertation, Universität der Bundeswehr München, Fakultät für Luft- und Raumfahrttechnik, 2002
- [Handmann et. al. 98a] U. Handmann, T. Kalinke, C. Tzomakas, M. Werner, W. v. Seelen, *An Image Processing System for Driver Assistance*, Proc. IEEE International Conference on Intelligent Vehicles, pp. 481-486, Stuttgart, 1998
- [Handmann et. al. 98b] U. Handmann, G. Lorenz, T. Schnitger, W. v. Seelen, *Fusion of different Sensors and Algorithms for Segmentation*, Proc. IEEE International Conference on Intelligent Vehicles, pp. 499-504, Stuttgart, 1998
- [Heimes et. al. 98] F. Heimes, H.-H. Nagel, T. Frank, *Model-Based Tracking of Complex Innercity Road Intersections*, Mathematical and Computer Modelling, 27 (9-11): 189-203, 1998
- [Hock 94] C. Hock, *Wissenbasierte Fahrzeugführung mit Landmarken für autonome Roboter*, Dissertation, Universität der Bundeswehr München, Fakultät für Luft- und Raumfahrttechnik, 1994
- [Hoffman 98] D.D. Hoffman, *Visual Intelligence (How We Create What We See)*, Norton, 1998
- [Hofmann et. al. 01] U. Hofmann, A. Rieder, E.D. Dickmanns, *Radar and Vision Data Fusion for Hybrid Adaptive Cruise Control on Highways*, Proc. of Computer Vision Systems, Vancouver, Lecture Notes in Computer Science 2095, Springer, 2001
- [von Holt 91] V. v. Holt, *System zur autonomen Fahrzeugführung auf Transputerbasis*, in Parallele Datenverarbeitung mit dem Transputer, 3. Transputer-Anwender-Treffen Aachen, Springer, Berlin, 1991
- [von Holt 94] V. v. Holt, *Tracking And Classification Of Overtaking Vehicles On Autobahnen*, in Proc. Int. Symp. on Intelligent Vehicles, IEEE, Paris, 1994
- [von Holt, Baten 98] V. v. Holt, S. Baten, *Perceptual Architecture for a Vision System of Autonomous Vehicles*, Proc. IEEE International Conference on Intelligent Vehicles, Stuttgart, 1998

- [von Holt 04a] V. v. Holt, *Funktion und Sensorfunktion im Wechselspiel*, TÜV-Konferenz "Sensorik in Sicherheits- und Fahrerassistenzsystemen", TÜV-Rheinland, Köln, 10.-11.2.2004
- [von Holt 04b] V. v. Holt, *Vernetzte Sensorsysteme zur Interpretation komplexer Fahrszenen*, Kooperationsforum "Fahrerassistenzsysteme - erkennen, vernetzen, unterstützen", BAIKA, Regensburg, 25.05.2004
- [von Holt, Maurer 04] V. v. Holt, M. Maurer, *Aktive Sicherheitssysteme mit maschineller Wahrnehmung - Anforderungen, Potentiale und Einführungshemmnisse*, Tagung "Aktive Sicherheit durch Fahrerassistenz", TÜV Süddeutschland, München 11.-12.3.2004
- [Hopfield 82] J. Hopfield, *Neural Networks And Physical Systems With Emergent Collective Computational Abilities*, Proc. of the National Academy of Sciences, Vol. 79, pp. 2554-2558
- [Ising 25] E. Ising, *Beitrag zur Theorie des Ferromagnetismus*, Zeitschrift für Physik, Vol. 31, Nr. 253
- [Jochem et. al. 95] T. Jochem, D. Pommerleau, B. Kumar, J. Armstrong, *PANS: A Portable Navigation Platform*, Proc. IEEE Symposium on Intelligent Vehicles, Detroit, 1995
- [Kaempchen et. al. 04] N. Kaempchen, K. Fuerstenberg, K. Dietmayer, *Ein Sensorfusions-system für automotive Sicherheits- und Komfortfunktionen*, Tagung "Aktive Sicherheit durch Fahrerassistenz", TÜV Süddeutschland, München 11.-12.3.2004
- [Kalman 60] R.E. Kalman, *A New Approach to Linear Filtering and Prediction Problems*, Transactions of the ASME, series D, Journal of Basic Engineering, 1960
- [Kamiya et. al. 97] Kamiya, Hiroyuki, Y. Fujita, T. Tsuruga, Y. Nakamura, S. Matsuda, K. Enomoto, *Intelligent Technologies of Honda ASV*, 5th ITS World Congress, Berlin, 1997
- [Kant] I. Kant, *Werke in zehn Bänden*, Wissenschaftliche Buchgesellschaft, Darmstadt, 1983
- [Kastrinaki et. al. 03] V. Kastrinaki, M. Zervakis, K. Kalaitzakis, *A survey of video processing techniques for traffic applications*, in Image and Vision Computing 21:359-381, Elsevier, 2003
- [Kim et. al. 95] K. Kim, S.Y. Oh, S.W. Kim, H. Jeong, J. Han, C.N. Lee, B.S. Kim, C.S. Kim, *An Autonomous Land Vehicle PRV II: Progresses and Performance Enhancement*, IEEE Conference on Intelligent Vehicles, pp. 264-269, Detroit, 1995
- [Kinzel 94] W. Kinzel, *Präattentive and attentive Bildverarbeitungsschritte zur visuellen Erkennung von Fußgängern*, Dissertation, Universität der Bundeswehr München, Fakultät für Luft- und Raumfahrttechnik, 1994

- [Kirchner 00] A. Kirchner, *Sensordatenverarbeitung eines Laserscanners für autonome Fahrfunktionen von Kraftfahrzeugen*, Dissertation, Universität der Bundeswehr Hamburg, Fortschrittberichte VDI Reihe 12, Nr. 438, 2000
- [Koller 92] D. Koller, *Detektion, Verfolgung und Klassifikation bewegter Objekte in monokularen Bildfolgen am Beispiel von Straßenverkehrsszenen*, Dissertation, Universität Karlsruhe, infix, Sankt Augustin, 1992
- [Kopischke 00] S. Kopischke, *Entwicklung einer Notbremsfunktion mit Rapid Prototyping Methoden*, Dissertation, Volkswagen AG, TU Braunschweig, 2000
- [Kosecka et. al. 97] J. Kosecka, R. Blasi, C.J. Taylor, J. Malik, *Vision-Based Lateral Control of Vehicles*, IEEE Conference on Intelligent Transportation Systems ITSC, Boston, 1997
- [Kosecka et. al. 98] J. Kosecka, R. Blasi, C.J. Taylor, J. Malik, *A Comparative Study of Vision-Based Lateral Control Strategies for Autonomous Highway Driving*, IEEE International Conference on Robotics and Automation, pp. 1903-1908, Leuven, 1998
- [Kosko 92] B. Kosko, *Neural Networks And Fuzzy Systems*, Prentice-Hall, Englewood Cliffs, N. J. , 1992
- [Krüger et. al. 95] W. Krüger, W. Enkelmann, S. Rössle, *Real-Time Estimation and Tracking of Optical Flow Vectors for Obstacle Detection*, IEEE Symposium on Intelligent Vehicles, pp. 304-309, Detroit, 1995
- [Kruse et al. 91] R. Kruse, E. Schwecke, J. Heinsohn, *Uncertainty And Vagueness in Knowledge-Based Systems: Numerical Methods*, Springer, Berlin, 1991
- [Kruse et. al. 94] R. Kruse, D. Nauck, F. Klawonn, *Neuronale Netze und Fuzzy-Systeme*, Vieweg, Braunschweig / Wiesbaden, 1994
- [Kuhnert 88] K. Kuhnert, *Zur Echtzeit-Bildfolgenanalyse mit Vorwissen*, Dissertation, Universität der Bundeswehr München, Fakultät für Luft- und Raumfahrttechnik, 1988
- [Lee et. al. 95] D.M. Lee, D.H. Kim, B.S. Kim, S.H. Moon, M.H. Han, *Development of Autonomous Test Vehicles Using Image Processing*, IEEE Symposium on Intelligent Vehicles, pp. 409-414, Tokyo, 1996
- [Loffeld 90] O. Loffeld, *Estimationstheorie*, Oldenbourg, 1990
- [Lorei et. al. 99] M. Lorei, C. Stiller, *Visual Sensing in Electronic Truck Coupling*, Proc. IEEE Conference on Image Processing ICIP, Kobe, 1999
- [Lowe 85] D.G. Lowe, *Perceptual Organization and Visual Recognition*, The Kluwer Int. Series in Engineering and Computer Science, Kluwer Academic Publishers, 1985
- [Lützelner 02] M. Lützelner, *Fahrbahnerkennung zum Manövrieren auf Wegenetzen mit aktivem Sehen*, Dissertation, Universität der Bundeswehr München, Fakultät für Luft- und Raumfahrttechnik, Fortschrittsberichte VDI Reihe 12, Nr. 493, 2002

- [Maurer 00] M. Maurer, *Flexible Automatisierung von Straßenfahrzeugen mit Rechnern*, Dissertation, Universität der Bundeswehr München, Fakultät für Luft- und Raumfahrttechnik, Fortschrittsberichte VDI Reihe 12, Nr. 443, 2000
- [Maybeck 79] P.S. Maybeck, *Stochastic Models, Estimation and Control*, Band 1-3, Academic Press, 1979
- [Mende 99] R. Mende, *Radarsysteme zur automatischen Abstandsregelung in Automobilen*, Dissertation, Technische Universität Braunschweig, 1999
- [Meinecke 01] M.M. Meinecke, *Zum optimalen Sendesignalentwurf für Automobilradare*, Dissertation, Technische Universität Hamburg-Harburg, 2001
- [Meissner 82] H.G. Meissner, *Steuerung dynamischer Systeme aufgrund bildhafter Information*, Dissertation, Universität der Bundeswehr München, Fakultät für Luft- und Raumfahrttechnik, 1982
- [Mimuro et. al. 97] Mimuro, Tetshuhi, Y. Miichi, T. Maemura, K. Hayafune, *Mitsubishi's Advanced Safety Vehicle (ASV)*, 5th ITS World Congress, Berlin, 1997
- [Moravec 83] H.P. Moravec, *The Stanford Cart and the CMU Rover*, in Proc. of the IEEE 71:872-884, 1983
- [Müller 96] N. Müller, *Autonomes Manövrieren und Navigieren mit einem sehenden Straßenfahrzeug*, Dissertation, Universität der Bundeswehr München, Fakultät für Luft- und Raumfahrttechnik, 1996
- [Mysliwetz 90] B.D. Mysliwetz, *Parallelrechner-basierte Bildfolgen-Interpretation zur autonomen Fahrzeugführung*, Dissertation, Universität der Bundeswehr München, Fakultät für Luft- und Raumfahrttechnik, 1990
- [Naab 04] K. Naab, *Sensorik- und Signalverarbeitungsarchitekturen für Fahrerassistenz und Aktive Sicherheit*, Tagung "Aktive Sicherheit durch Fahrerassistenz", TÜV Süddeutschland, München 11.-12.3.2004
- [Nagel et. al. 98a] H.-H. Nagel, A. Gehrke, *Bildbereichsbasierte Verfolgung von Straßenfahrzeugen durch adaptive Schätzung und Segmentierung von Optischen-Fluß-Feldern*, 20. DAGM Symposium Mustererkennung, pp. 314-321, Stuttgart, 1998
- [Nagel et. al. 98b] H.-H. Nagel, F. Heimes, S. Noltemeier, M. Haag, *Quantitative Comparison between Trajectory Estimates Obtained from a Binocular Camera Setup within a Moving Road Vehicle and from the Outside by a Stationary Monocular Camera*, IEEE International Conference on Intelligent Vehicles, pp. 423-429, Stuttgart, 1998
- [Nilsson 84] N. Nilsson, *Shakey the robot*, A.I. Center Technical Note 323, SRI International, 1984
- [Obara et. al. 97] Obara, Hideo, M. Kaneda, J. Koreishi, T. Shirato, *Nissan's Advanced Safety Vehicle*, 5th ITS World Congress, Berlin, 1997

- [Özgüner et. al. 97] U. Özgüner, B. Baertlein, C. Cavello, D. Farkas, C. Hatipoglu, S. Lytle, J. Marin, F. Paynter, K. Redmill, S.J. Schneider, E. Walton, J. Young, *The OSU Demo '97*, IEEE Conference on Intelligent Transportation Systems ITSC, Boston, 1997
- [Pellkofer 03] M. Pellkofer, *Verhaltensentscheidung für autonome Fahrzeuge mit Blickrichtungssteuerung*, Dissertation, Universität der Bundeswehr München, Fakultät für Luft- und Raumfahrttechnik, 2003
- [Pommerleau 95] D. Pommerleau, *RALPH: Rapidly Adapting Lateral Position Handler*, in Proc. IEEE Symposium on Intelligent Vehicles, Detroit, 1995
- [Pommerleau 97] D. Pommerleau, *Visibility Estimation from a Moving Vehicle Using the RALPH Vision System*, Proc. IEEE Conference on Intelligent Transportation Systems ITSC, Boston, 1997
- [Pontois et. al. 97] J. Pontois, P. Deloof, A.M. Jolly-Desodt, D. Jolly, *Collision Avoidance System based on a multisensor set-up*, 5th ITS World Congress, Berlin, 1997
- [Pontois et. al. 98] J. Pontois, A.M. Jolly-Desodt, D. Jolly, *Perception of the road environment based on a multisensor set-up*, Proc. IEEE International Conference on Intelligent Vehicles, pp. 493-498, Stuttgart, 1998
- [Pschyrembel 82] W. Pschyrembel, *Klinisches Wörterbuch*, Walter de Gruyter, Berlin, 1982
- [RAL 84] Forschungsgesellschaft für Strassen- und Verkehrswesen, Arbeitsgruppe Strassenentwurf, *Richtlinie für die Anlage von Straßen (RAS-L-1)*, 1984
- [Rieder 00] A. Rieder, *Fahrzeuge sehen*, Dissertation, Universität der Bundeswehr München, Fakultät für Luft- und Raumfahrttechnik, 2000
- [Risack et. al. 98] R. Risack, P. Klausmann, W. Krüger, W. Enkelmann, *Robust lane recognition embedded in a real-time driver assistance system*, Proc. IEEE International Conference on Intelligent Vehicles, pp. 35-40, Stuttgart, 1998
- [Roberts 63] L.G. Roberts, *Machine Perception of Three-dimensional Solids*, in J.P. Toppel et.al.(Hrsg.), *Optical and Electro-Optical Information Processing*, S. 159-197, MIT Press, Cambridge, MA, 1963
- [Schell 92] F.R. Schell, *Bordautonomer automatischer Landeanflug aufgrund bildhafter und inertialer Meßdatenauswertung*, Dissertation, Universität der Bundeswehr München, Fakultät für Luft- und Raumfahrttechnik, 1992
- [Schick 92] J. Schick, *Gleichzeitige Erkennung von Form und Bewegung durch Rechnersehen*, Dissertation, Universität der Bundeswehr München, Fakultät für Luft- und Raumfahrttechnik, 1992
- [Schiehlen 95] J. Schiehlen, *Kameraplattformen für aktiv sehende Fahrzeuge*, Dissertation, Universität der Bundeswehr München, Fakultät für Luft- und Raumfahrttechnik, 1995
- [Schischkoff 82] G. Schischkoff, *Philosophisches Wörterbuch*, Kröner, Stuttgart, 1982



- [Schmid 94] M. Schmid, *3D-Erkennung von Fahrzeugen in Echtzeit aus monokularen Bildfolgen*, Dissertation, Universität der Bundeswehr München, Fakultät für Luft- und Raumfahrttechnik, 1994
- [Schneider et. al. 98] S.J. Schneider, U. Özgüner, *A Framework for Data Validation and Fusion, Fault Detection and Isolation for Intelligent Vehicle Systems*, Proc. IEEE International Conference on Intelligent Vehicles, pp. 528-533, Stuttgart, 1998
- [Schulze 97] M. Schulze, *CHAUFFEUR - The European Way towards an Automated Highway System*, Proc. of the 4th World Congress on Intelligent Transportation Systems, 1997
- [Shirai 87] Y. Shirai, *Three Dimensional Computer Vision*, Series Symbolic Computation, Springer, Berlin, 1987
- [Siedersberger 03] K.H. Siedersberger, *Komponenten zur automatischen Fahrzeugführung in (semi-)autonomen Fahrzeugen*, Dissertation, Universität der Bundeswehr München, Fakultät für Luft- und Raumfahrttechnik, 2003
- [Stiller et. al. 97] C. Stiller, W. Pöchmüller, B. Hürtgen, *Stereo Vision on Driver Assistance Systems*, IEEE Conference on Intelligent Transportation Systems ITSC, Boston, 1997
- [Stiller et. al. 98] J. Hipp, C. Rössig, A. Ewald, *Multisensor Obstacle Detection and Tracking*, Proc. IEEE International Conference on Intelligent Vehicles, pp. 451-456, Stuttgart, 1998
- [Stüker 04] D. Stüker, *Heterogene Sensordatenfusion zur robusten Objektverfolgung im automobilen Straßenverkehr*, Dissertation, Universität Oldenburg, 2004
- [Suganthan et. al. 93] P.N. Suganthan, E.K. Teoh, D.P. Mital, *Programming Hopfield Network for Object Recognition*, Proc. International Conference on Systems, Man and Cybernetics, Paris, 1993
- [Suganthan et. al. 94] P.N. Suganthan, E.A. Teoh, D.P. Mital, *On Attributed Relational Graph Matching Using Hopfield Network*, Proc. 11th European Conference on Artificial Intelligence, pp. 236-240, Wiley, 1994
- [Tan et. al. 94a] T.N. Tan, G.D. Sullivan, K.D. Baker, *Fast Vehicle Localisation and Recognition Without Line Extraction and Matching*, Proc. 5th British Machine Vision Conference, York, BMVA Press, 1994
- [Tan et. al. 94b] T.N. Tan, G.D. Sullivan, K.D. Baker, *Vehicle Localisation and Discrimination in Outdoor Traffic Scenes*, Proc. 2nd Asian Conference on Computer Vision, Singapore, Springer, 1995
- [Thomanek 96] F. Thomanek, *Visuelle Erkennung und Zustandsschätzung von mehreren Straßenfahrzeugen zur autonomen Fahrzeugführung*, Dissertation, Universität der Bundeswehr München, Fakultät für Luft- und Raumfahrttechnik, Fortschrittsberichte VDI Reihe 12, Nr. 272, 1996

- [Thomanek, von Holt 93] F. Thomanek, V. v. Holt, *An all-transputer visual Autobahn-autopilot*, IEEE Int. Conf. on Computer Vision, Berlin, 1993
- [Thorpe et. al. 97] C. Thorpe, T. Jochem, D. Pommerleau, *The 1997 Automated Highway Free Agent Demonstration*, IEEE Conference on Intelligent Transportation Systems ITSC, Boston, 1997
- [Tsinas 96] L. Tsinas, *Zur Auswertung von Farbinformationen beim maschinellen Erkennen von Verkehrssituationen in Echtzeit*, Dissertation, Universität der Bundeswehr München, Fakultät für Luft- und Raumfahrttechnik, 1996
- [Tsugawa 93] S. Tsugawa, *Vision-Based Vehicles in Japan: The Machine Vision Systems and Driving Control Systems*, IEEE International Symposium on Industrial Electronics, pp. 278-285, 1993
- [Tsugawa et. al. 93] S. Tsugawa, K. Tomita, A. Hosaka, M. Taniguchi, *Steering Control Algorithms for Intelligent Vehicles with Machine Vision*, IFAC 12th Triennial World Congress, pp. 117-122, Sydney, 1993
- [Ulmer 94a] B. Ulmer, *Autonomous Automated Driving in Real Traffic*, Proc. of the first world congress on applications of transport telematics and intelligent vehicle-highway systems, 1994
- [Ulmer 94b] B. Ulmer, *VITA II - Active Collision Avoidance in Real Traffic*, Symposium on Intelligent Vehicles, pp. 1-6, 1994
- [Unbehauen] H. Unbehauen, *Regelungstechnik I-III*, Vieweg, Braunschweig / Wiesbaden
- [Weisser et. al. 98] H. Weisser, P.J. Schulenberg, R. Bergholz, U. Lages, *Autonomous Driving on Vehicle Test Tracks: Overview, Motivation and Concept*, Proc. IEEE International Conference on Intelligent Vehicles, pp. 439-443, Stuttgart, 1998
- [Werner 97] S. Werner, *Machinelle Wahrnehmung für den bordautonomen automatischen Hubschrauberflug*, Dissertation, Universität der Bundeswehr München, Fakultät für Luft- und Raumfahrttechnik, 1997
- [Wertheimer 23] M. Wertheimer, *Untersuchungen zur Lehre von der Gestalt II*, Psychologische Forschung 23(4), 1923
- [Widmann et. al. 98] G.R. Widmann, W.A. Bauson, S.W. Alland, *Development of Collision Avoidance Systems at Delphi Automotive Systems*, Proc. IEEE International Conference on Intelligent Vehicles, pp. 353-358, Stuttgart, 1998
- [Willersinn et. al. 97] D. Willersinn, W. Enkelmann, *Robust Obstacle Detection and Tracking by Motion Analysis*, IEEE Conference on Intelligent Transportation Systems ITSC, Boston, 1997
- [Williamson et. al. 98] T. Williamson, C. Thorpe, *Detection of Small Obstacles at Long Range Using Multibaseline Stere*, Proc. IEEE International Conference on Intelligent Vehicles, pp. 311-316, Stuttgart, 1998

- [Wisselmann et. al. 04] D. Wisselmann, K. Gresser, H. Spannheimer, K. Bengler, A. Huesmann, *Connected Drive - ein methodischer Ansatz für die Entwicklung zukünftiger Assistenzsysteme*, Tagung "Aktive Sicherheit durch Fahrerassistenz", TÜV Süddeutschland, München 11.-12.3.2004
- [Wünsche 87] H.J. Wünsche, *Erfassung und Steuerung von Bewegungen durch Rechnersehen*, Dissertation, Universität der Bundeswehr München, Fakultät für Luft- und Raumfahrttechnik, 1987
- [Yamamoto et. al. 99] M. Yamamoto, Y. Kagawa, A. Okuno, *Robust Control for Automated Lane Keeping Against Lateral Disturbances*, IEEE Conference on Intelligent Transportation Systems ITSC, pp. 240-245, Tokyo, 1999
- [Zapp 88] A. Zapp, *Automatische Straßenfahrzeugführung durch Rechnersehen*, Dissertation, Universität der Bundeswehr München, Fakultät für Luft- und Raumfahrttechnik, 1988
- [Zimbardo 83] P.G. Zimbardo, *Psychologie*, Springer, Berlin, 1983
- [Zhang et. al. 93] S. Zhang, G.D. Sullivan, K.D. Baker, *The Automatic Construction of a View-Independent Relational Model for 3-D Object Recognition*, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 15, No. 6, 1993