Institut für Neuroinformatik Bochum

Internal Report 98–01

## Fusion von Bildanalyseverfahren mittels einer neuronalen Kopplungsstruktur

by

Thomas Schnitger und Uwe Handmann

Ruhr-Universität Bochum Institut für Neuroinformatik 44780 Bochum



IR-INI 98–01 April 1998 ISSN 0943-2752

© 1998 Institut für Neuroinformatik, Ruhr-Universität Bochum, FRG

# Fusion von Bildanalyseverfahren mittels einer neuronalen Kopplungsstruktur

 $\label{eq:constraint} Thomas \ Schnitger^* \ und \ Uwe \ Handmann^\dagger \\ Institut \ für \ Neuroinformatik, \ Ruhr-Universität \ Bochum, \ FRG$ 

## Inhaltsverzeichnis

Abl	kürzungen	iii							
Einleitung									
Die	Die Einzelbildanalyse								
2.1	Die Texturanalyse	2							
	2.1.1 Der mittlere Grauwert	2							
	2.1.2 Die lokale Varianz	3							
	2.1.3 Die Entropie	3							
	2.1.4 Die Kookkurrenz Matrizen	4							
	2.1.5 Die fraktale Geometrie	5							
2.2	Konturbasierte Verfahren	7							
	2.2.1 Das Linienbild	7							
	2.2.2 Die lokale Orientierungskodierung	8							
	2.2.3 Die Schlagschattendetektion	9							
Bild	Bildsequenzanalyse 9								
3.1	Differenzbildanalyse	9							
3.2	Verschiebungsvektorfelder	10							
Multisensor-Analyse 1									
4.1	Die RADAR-Daten	11							
Fus	ion	13							
5.1	Hierarchieebenen der Fusionsprozesse	13							
5.2	Fusion der Merkmale	13							
5.3	Multisensor-Fusion	14							
5.4	Koppelungsnetzwerk	15							
	Able Einil Die 2.1 2.2 Bild 3.1 3.2 Mui 4.1 Fusi 5.1 5.2 5.3 5.4	Abkürzungen         Einleitung         Die Einzelbildanalyse         2.1       Die Texturanalyse         2.1.1       Der mittlere Grauwert         2.1.2       Die lokale Varianz         2.1.3       Die Entropie         2.1.4       Die Kookkurrenz Matrizen         2.1.5       Die fraktale Geometrie         2.1.6       Die kookkurrenz Matrizen         2.1.7       Die Intropie         2.1.8       Linenbild         2.2.1       Das Linienbild         2.2.2       Die lokale Orientierungskodierung         2.2.3       Die Schlagschattendetektion         2.2.3       Die Schlagschattendetektion         3.1       Differenzbildanalyse         3.1       Differenzbildanalyse         3.2       Verschiebungsvektorfelder         4.1       Die RADAR-Daten         5.1       Hierarchieebenen der Fusionsprozesse         5.2       Fusion         5.1       Hierarchieebenen der Fusionsprozesse         5.2       Fusion         5.3       Multisensor-Fusion         5.4       Koppelungsnetzwerk							

\*Schnitger@aol.com

 $^{\dagger} Uwe. Handmann @neuroinformatik.ruhr-uni-bochum.de$ 

6	Künstliche neuronale Netze									
	6.1	Das bi	ologische Vorbild	16						
	6.2	Das Pe	erzeptron	17						
	6.3	Das M	ultilayer-Perzeptron	17						
	6.4	Der Ba	ackpropagation-Lernalgorithmus	19						
7	Praktischer Teil									
	7.1	Ausga	ngspunkt der Arbeit	21						
	7.2	Dateni	material	22						
	7.3	Merkn	alsextraktion	22						
		7.3.1	Erweiterung der Konturmerkmale	22						
		7.3.2	Die lokale Orientierungskodierung als Texturanalyse	24						
		7.3.3	Varianz	24						
		7.3.4	Ortsabhängigkeit der Objektpunkte	25						
		7.3.5	Untersuchung zur Einwirkung einer Gammakorrektur	27						
	7.4	Fusion		27						
		7.4.1	Vorverarbeitung der Merkmale	27						
		7.4.2	Die Rückkopplung des Segmentierungsergebnisses	27						
		7.4.3	Variation der Netz-Topologie	28						
		7.4.4	Aktivierungsfunktion	28						
		7.4.5	Training	29						
		7.4.6	Die Datenbanken	29						
8	Analyse 31									
	8.1	Fehlertoleranz und Robustheit								
	8.2	Sensitivitätsanalyse								
9	Zusammenfassung und Ausblick 34									
0	9.1	.1 Zusammenfassung								
	9.2	Ausblick								

# Abkürzungen

MLP	Multilayer-Perzeptron
RADAR	Radio Detection and Ranging
LIDAR	Light Detection and Ranging
CCD	Charged Couple-Device
FMCW	Frequency modulated continuous Wave
LOC	Lokale Orientierungskodierung (Local Orientation Coding)
$LOC-N_4$	Lokale Orientierungskodierung in einer Vierer-Nachbarschaft
LOC-N <sub>8</sub>	Lokale Orientierungskodierung in einer Achter-Nachbarschaft
$\langle G \rangle$	mittlerer Grauwert
m,n	Koordinaten
P	Anzahl der Bildpunkte
В	Bildbereich
$\langle G \rangle_{m,n}$	mittlerer Grauwert am Punkt $(m, n)$
k, l	Bildpunktindizes
V	Varianz
thr	Schwellwert
$p(q_k)$	relative Häufigkeit des Grauwertes $g_k$
$g_k$	Grauwert am Bildpunkt $k$
$x_i$	zentrierter Bildpunkt eines Bereichs
$G(x_i)$	Entropie
P(i, j)	Kookkurrenz Matrix
d	Abstand
$\alpha$	Winkel
S(i, j)	rotationsinvariante Kookkurrenzmatrix
x, y	Koordinaten
$v_i$	Vektoren
L(s)	Länge eines Poligonenzuges
s	Skaleneinheit
D	fraktale Dimension
R	Referenzpunkt
$\sigma$	Schwellwert
k	Eingangsschicht eines neuronalen Netzes
j	verdeckte Schicht eines neuronalen Netzes
i	Ausgangsschicht eines neuronalen Netzes
a, b	Indizes für Neuronen allgemein
$s_k$	Eingangswerte
$s_j$	Aktivität eines Neurons in der verdeckten Schicht
$s_i$	Aktivität eines Neurons in der Ausgangsschicht
$\sigma(x)$	Aktivierungsfunktion
D	Eingangsdimension
H	Anzahl der Neuronen in der verdeckten Schicht
0	Anzahl der Neuronen in der Ausgabeschicht
x	Eingangsvektor
У	Ausgangssollvektor
ν	Index für Trainingsmuster
p	Anzahl der Trainingsmuster
E	quadratischer Fehler
$w_{ab}$	Gewichte allgemein
$w_{ij}$	Gewichte zwischen der mittleren und der Ausgabeschicht
$w_{jk}$	Gewichte zwischen der Eingabe- und der mittleren Schicht

## 1 Einleitung

Für das sichere Führen von Fahrzeugen im Straßenverkehr ist ein hohes Maß an Informationsverarbeitung notwendig, um aus den zur Verfügung stehenden Informationen, Handlungen für die Fahrzeugsteuerung abzuleiten [26].

Der Mensch löst diese Aufgabe hauptsächlich auf der Basis visueller Informationen. Durch die Arbeitsweise des menschlichen Gehirns motiviert, wird am Institut für Neuroinformatik der Ruhr-Universität Bochum an einer Fahrzeugführung mittels Computer Vision gearbeitet [3, 7].

Fortlaufend oder zumindest in kurzen Abständen müssen hierbei Verkehrsteilnehmer aus den visuellen Informationen extrahiert und danach weiter attributiert werden. Wichtige Eigenschaften sind hierbei: Objektklasse (PKW, LKW, Fußgänger etc.), Abstand, Geschwindigkeit, Bewegungsrichtung und das Gefahrenpotential bezüglich der eigenen Ortsveränderung.

Die Vielzahl der durch die Umwelt aufgestellten Randbedingungen und das aus der Aufgabenstellung implizierte hohe Maß an Sicherheit bedingen ein robustes und flexibles Gesamtsystem. Dieses Gesamtsystem besteht zum einen aus Basis-Algorithmen zur Vorverarbeitung der Eingabedaten und Extraktion von Bildmerkmalen und zum anderen aus darauf aufbauenden Verfahren zur Segmentierung, Klassifizierung und Verfolgung von Fahrzeugen [11].

Der voliegende Artikel ist aus einer Diplomarbeit mit dem Titel Untersuchungen zum Entwurf neuronaler Kopplungsstrukturen zur Fusion von Bildanalyseverfahren entstanden, die am Inititut für Neuroinformatik von Herrn Thomas Schnitger erstellt wurde [24]. Der Schwerpunkt ist die initiale Segmentierung von Grauwertbildsequenzen, d.h. die Zerlegung der einzelnen Bilder in disjunkte Teilbereiche in denen Fahrzeuge abgebildet sind. Hierzu werden zunächst verschiedene Bildverarbeitungsalgorithmen analysiert und ihre Eignung zur Segmentierung von Fahrzeugen vor einem bewegten Hintergrund diskutiert. Durch die Fusion geeigneter Merkmale wird die Segmentierung robust gegenüber Falschinformationen und Rauschen und fehlertolerant gegen den Ausfall von Basismodulen. Redundante Informationen leisten hierbei einen Beitrag, die Sicherheit und Zuverlässigkeit zu erhöhen. Ein Fusions-Konzept zur flexiblen Einkoppelung beliebiger Sensorik wird entwickelt und exemplarisch die Fusion von Radardaten realisiert. Eine abschließende Analyse gibt Aufschluß über den Einfluß einzelner Basismodule.

## 2 Die Einzelbildanalyse

Vor einer Segmentierung müssen aus den Bilddaten die Merkmale extrahiert werden, die eine Relevanz für die gestellte Aufgabe besitzen. Dieses Kapitel gibt eine Übersicht über die Verfahren, die grundsätzlich für eine Segmentierung eines Grauwertbildes geeignet sind. Da im aufgabenspezifischen Fall dieser Arbeit Fahrzeuge vor einem sich bewegenden Hintergrund in relativ kurzer Zeit segmentiert werden sollen, wird die Leistung der einzelnen Algorithmen diesbezüglich diskutiert.

## 2.1 Die Texturanalyse

Der Begriff *Textur* wird in der Literatur verschieden und nur sehr allgemein definiert. In dieser Arbeit soll die Definition von A. KORN [18] Verwendung finden:

Allgemein bezeichnet "Textur" eine Fläche, innerhalb der bestimmte lokale Eigenschaften – die Texturmerkmale – konstant bleiben.

Texturmerkmale beziehen sich, nach dieser Definition, nicht nur auf einen einzelnen Bildpunkt, sondern auch auf dessen Umgebung und haben somit einen integrativen Charakter. Unterscheiden sich die Texturmerkmale oberflächenstrukturierter Objekte von denen anderer Bildbereiche, so wird eine Segmentierung dieser Objekte möglich.

In den folgenden Kapiteln werden verschiedene statistische Maße beschrieben, die zur Texturanalyse in der Literatur vorgeschlagen werden und deren Verwendung zur Lösung aufgabenspezifischer Segmentierungen diskutiert.

#### 2.1.1 Der mittlere Grauwert

Der *mittlere Grauwert* ist ein sehr einfaches, statistisches Mittel, um Aussagen über die lokalen Eigenschaften einer Region zu machen.

Der mittlere Grauwert  $\langle G \rangle$  des Bildpunktes an den Koordinaten (m, n) berechnet sich dabei als arithmetischer Mittelwert der Grauwerte g aller Bildpunkte P des Bildbereiches B:

$$\langle G \rangle_{m,n} = \frac{1}{P} \sum_{k,l \in B} g_{m-k,n-l}.$$
 (1)

Die Koordinaten (k, l) indizieren die Bildpunkte im betrachteten Bildbereich relativ zum Zentralpunkt (m, n).

Zur Segmentierung ist der mittlere Grauwert nur bei einer konstanten Beleuchtung und dem Vorwissen, wie sich Vorder- und Hintergrund in der Helligkeit unterscheiden, sinnvoll einzusetzen.

Die Abbildung 1 zeigt den mittleren Grauwert einer Straßenverkehrsszene, der jeweils über einen  $(10 \times 10)$  Punkte großen Bildbereich B um ein Zentralpixel (m, n) ermittelt wird.



Abbildung 1: Grauwertbild, mittlere Grauwert über  $(10 \times 10)$  Bildpunkte

An den Bildrändern ist eine Berechnung des mittleren Grauwertes über einen vorgegebenen Bereich B um die Zentralpunkte (m, n) nicht mehr möglich. Diese Bildpunkte werden in Abbildung 1 (rechts) mit dem Wert  $\langle G \rangle = 0$  dargestellt.

Für die Segmentierung von Objekten in einer Straßenverkehrsszene, bei der

• sich die Beleuchtung der Szene in der Bildsequenz verändert (Sonnenstand, Schatten, ...),

- die Objekte (PKW's, LKW's etc.) verschiedene Farben besitzen und deshalb auch mit verschiedenen Grauwerten abgebildet werden,
- der Hintergrund alle Grauwerte beinhaltet

ist der mittlere Grauwert nicht geeignet.

#### 2.1.2 Die lokale Varianz

Ein weiterer einfacher Parameter zur Texturbeschreibung ist, nach dem mittleren Grauwert, die *lokale Varianz* [14] des Grauwertes. Die Berechnung der Grauwertvarianz V ist eine nichtlineare Filteroperation

$$V_{m,n} = \frac{1}{P-1} \sum_{k,l \in B} (g_{m-k,n-l} - \langle G \rangle_{m,n})^2.$$
(2)

Die Summe läuft über alle Bildpunkte P eines Bildbereichs B. Der Ausdruck  $\langle G \rangle_{m,n}$  bezeichnet den Mittelwert des Grauwertes am Punkt (m, n) wie er mit Gleichung 1 in Kapitel 2.1.1 berechnet wird. Die Koordinaten (k, l) indizieren die Bildpunkte im betrachteten Bildbereich B um den Zentralpunkt (m, n).

Das im Rahmen dieser Arbeit erstellte Modul berechnet die Varianz über einen Bereich von  $(7 \times 7)$  Bildpunkten.

Eine Segmentierung auf Basis der lokalen Varianz zeigt Abbildung 2. Durch ein Schwellwertverfahren werden hohe lokale Varianzen dem Vordergund und kleine dem Hintergrund zugeordnet. Mit einer auf das Varianzbild (Abb. 2 mitte) angewendeten Schwelle von thr = 5 ergibt sich, das in Abbildung 2 rechts dargestellte Segmentierungsergebnis.



Abbildung 2: Grauwertbild, lokale Varianz über  $(7 \times 7)$  Bildpunkte, Segmentierung von Bereichen mit hoher lokaler Varianz durch ein Schwellwertverfahren

Die Abbildung 2 motiviert, die lokale Varianz als Eingangsmerkmal für einen Segmentierungsprozeß für Fahrzeuge zu nutzen. Mit dem niedrig angesetzten Schwellwert werden die Fahrzeuge in der Abbildung fast vollständig als Objekte klassifiziert. Die zusätzlich segmentierten Fahrbahnmarkierungen können durch die Wahl eines höheren Schwellwertes unterdrückt werden.

#### 2.1.3 Die Entropie

Basierend auf der Informationstheorie, die C. SHANNON [25] einführte, wird der lokale Informationsgehalt in Bilder geschätzt. Hierbei wird die *Entropie*  $G(x_i)$  als Maß der zu erwartenden Information eines Bildausschnitts *B* herangezogen.

$$G(x_i) = -\sum_k p(g_k) \log p(g_k)$$
(3)

Die Summe läuft über alle Bildpunkte k des Bildausschnitts B.  $p(g_k)$  ist die relative Häufigkeit der Grauwerte  $g_k$ . Die Schätzgüte des Informationsgehaltes hängt von der Größe des Bildbereichs B ab [15].

Bei dem zur Verfügung stehenden Modul, wird dabei über einen Bildausschnitt der Größe  $(20 \times 20)$  Bildpunkte, zentriert im Punkt  $x_i$ , die Entropie nach Gleichung 3 berechnet.

Abbildung 3 zeigt eine Segmentierung auf Basis der Entropie. Im Segmentierungsbild (Abb. 3 rechts) werden alle Bildpunkte ausgeblendet, deren Entropiewert  $G(x_i)$  unterhalb einer vorgegebenen Schwelle (thr = 90) liegt.



Abbildung 3: Grauwertbild, Entropie, Segmentierung von Bereichen mit hoher Entropie durch ein Schwellwertverfahren

Die gesamte Fahrbahn wird durch die Wahl des Schwellwertes dem Hintergrund zugeordnet. Fahrzeuge werden jedoch in einigen Bereichen auch dem Hintergrund zugeordnet.

#### 2.1.4 Die Kookkurrenz Matrizen

Aus den *Kookkurrenz Matrizen* [12] lassen sich unterschiedliche statistische Merkmale berechnen. Diese Merkmale sind zur Beschreibung von Texturen gut geeignet.

Hierbei wird zunächst die Häufigkeit des Auftretens von verschiedenen *Pixelpaaren* in einer Matrix P(i, j) erfaßt. Ein *Pixelpaar* besteht dabei aus zwei Bildpunkten, die sich in einer vorgegebenen relativen Lage zueinander befinden und in der jeder Bildpunkt einen bestimmten Grauwert besitzt. Die Relation dieser Bildpunkte kann dabei durch einen Abstand *d* und einen Winkel  $\alpha$  definiert werden. Definiert sich der Winkel  $\alpha$  wie in Abbildung 4 angegeben, so kann er die Werte  $\alpha \in \{0^{\circ}, 45^{\circ}, 90^{\circ}, 135^{\circ}\}$  annehmen.



Abbildung 4: Richtungsdefinition für die Relation zweier Bildpunkte: Die Zellen 1 und 5 sind in 0° (horizontal), die Zellen 2 und 6 sind in 135°, die Zellen 3 und 7 sind in 90° (vertikal) und die Zellen 4 und 8 sind in 45° nächste Nachbarn zur Zelle X.

Die Abbildung 5 gibt ein Beispiel für die Berechnung einer Kookkurrenzmatrix mit der Relation: Abstand d = 2 und Winkel  $\alpha = 0^{\circ}$  (horizontale Richtung).



Abbildung 5: Bildausschnitt mit 3 verschiedenen Grauwerten und der dazugehörigen Kookkurrenzmatrix für  $\alpha = 0^{\circ}$  und d = 2

Um eine Rotationsinvarianz zu erreichen wird zusätzlich in den meisten technischen Anwendungen die Summe  $S(i, j) = \sum_{\alpha} P(i, j)$  für alle möglichen Winkel  $\alpha \in \{0^{\circ}, 45^{\circ}, 90^{\circ}, 135^{\circ}\}$ mit dem Abstand d = const. berechnet. Die Skalierungsvarianz, die das Berechnungsverfahren impliziert, kann durch die Berechnung mit verschiedenen Distanzen d = 1, 2, ...eingeschränkt werden.

Die Eigenschaften der so gewonnenen Matrix, wie z. B. Energie, Entropie oder größte Kookkurrenz, lassen sich zur Charakterisierung der Textur nutzen ([26], Seite 13). R. M. HARALICK, K. SHANMUGAN U. I. DINSTEIN schlagen in [12] insgesamt 14 solcher Texturmerkmale vor.

Verzichtet man auf die Berechnung der Texturmerkmale und liefert die nichtredundanten Elemente der symmetrischen Matrix direkt an ein neuronales Netz, das eine Segmentierungsaufgabe übernehmen soll, so wird die Dimension des Eingangsvektors sehr groß. Die Ermittlung des Segmentierungsergebnisses würde dadurch ähnlich rechenintensiv werden. (Mit einem solchen Ansatz für eine Klassifikation eines auf 16 Graustufen reduzierten Bildes kamen U. HANDMANN U. T. KALINKE in [10] auf 136 Eingangsneuronen.)

Da die Berechnung der Kookkurrenz Matrizen und die darauf folgende Merkmalsextraktion sehr viel Rechenzeit beansprucht, ist dieses Verfahren trotz seiner Aussagekraft für eine Segmentierung, im aufgabenspezifischen Fall dieser Arbeit, nicht geeignet.

#### 2.1.5 Die fraktale Geometrie

Der Begriff *fraktale Geometrie* [9] wurde um 1975 von Benoit Mandelbrot geprägt. Wesentliches Merkmal der fraktalen Geometrie ist die *Selbstähnlichkeit*. Diese ist wie folgt definiert:

Wird der Maßstab zur Betrachtung eines Fraktals beliebig verkleinert, so findet man immer die gleiche prinzipielle Struktur des Fraktals.

Die Selbstähnlichkeit läßt sich sehr häufig beobachten, z.B. bei Küstenlinien, Pflanzenblättern, Wolken, bestimmten komplexen Zahlen etc. In einem bekannten Beispiel von 1961 nähert L. F. Richardson die Länge einer Küstenlinie durch einen Poligonenzug an. Je kleiner er die zugrundegelegte Skala s machte, desto länger wurde der Polygonenzug L(s). Aus seinen Experimenten leitete er folgenden Zusammenhang ab:

$$L(s) \sim s^{1-D} \tag{4}$$

Wird  $\log(s)$  in einem Koordinatensystem auf der Abszisse und  $\log(L(s))$  auf der Ordinate aufgetragen, so ergibt sich näherungsweise eine Gerade. Mandelbrot erkannte die Steigung dieser Geraden, als eine Näherung für 1 - D und daß D als eine fraktale Dimension betrachtet werden kann.

In der Literatur gibt es verschiedene mathematische Formulierungen, die als fraktale Dimensionen bezeichnet werden. Die wichtigsten fraktalen Dimensionen sind:

- die Hausdorff–Besicovich–Dimension,
- die Kapazitäts–Dimension und
- die Korrelations–Dimension.

In der Diplomarbeit von H. MARBACH [19] wird, ausgehend von einem Algorithmus zur Bestimmung der fraktalen Korrelations–Dimension in Binärbildern<sup>1</sup>, ein auf Grauwertbilder anwendbares Verfahren entwickelt. Dabei wird das Grauwertbild in mehrere Binärbilder aufgespalten, die jeweils verschiedene Grauwertbereiche des Originalbildes repräsentieren. Als Ergebnis ergibt sich ein Dimensionsvektor der fraktalen Geometrie.

Mittels einer euklidischen Abstandsbestimmung verschiedener Dimensionsvektoren ist eine Klassifikation verschiedener Texturen und damit auch eine Segmentierung möglich.

Die in [19] durchgeführten Untersuchungen zur Segmentierung zeigen, daß sich Bilder mit homogenen texturalen Eigenschaften gut analysieren lassen. So konnte in einer natürlichen Szene ein Waldweg mit homogener Textur gut segmentiert werden.



Abbildung 6: Grauwertbild einer natürlichen Szene und Segmentierung eines Waldwegs auf Basis der fraktalen Dimension [19]

Eine Segmentierung von stark strukturierten Fahrzeugen vor einem schwach strukturierten Hintergrund gelang in [26], Seite 14. Der gesamte Fahrbahnbereich konnte in Abbildung 7 ausgeblendet werden. Der LKW im rechten Bildbereich wird aufgrund seiner homogenen Texturbereiche sehr schlecht segmentiert.

<sup>&</sup>lt;sup>1</sup>Binärbilder enthalten nur die Informationen 1 oder 0 (schwarz oder weiß).



Abbildung 7: Grauwertbild mit schwach strukturiertem Hintergrund und Segmentierung auf Basis der fraktalen Dimension [26]

Für stark inhomogen strukturierte Bereiche (wie z.B. PKW's) vor einem ebenso inhomogen strukturierten Hintergrund liefert dieses Verfahren keine verwertbaren Ergebnisse.

Die Skalierungs- und Rotationsinvarianz der fraktalen Dimension motiviert, sie zur Segmentierung zu nutzen. Da sich die zur Verfügung stehenden Verfahren aber iterativ der fraktalen Dimension annähern, benötigen sie sehr viel Rechenzeit. Die Segmentierung liefert zudem keine guten Ergebnisse. Der Entwicklungsstand dieses Modules ist noch nicht so weit fortgeschritten um zur Unterstüzung der gestellten Segmentierungsaufgabe beizutragen.

## 2.2 Konturbasierte Verfahren

### 2.2.1 Das Linienbild

An Kanten von künstlichen Objekten treten typischerweise Linien auf. Eine Anhäufung von Linienelementen liefert einen Hinweis auf ein Objekt.

Die am Institut für Neuroinformatik verwendeten Linienbilder werden mit Hilfe einer *Countour Matching Hardware* der Firma *ELTEC* aus Grauwertbildern erzeugt. Sie erzeugt aus den Grauwertbildern zunächst mit Hilfe eines Sobelfilteroperators ein Konturbild. Daraus entsteht durch Ausdünnung und Konkatination das Linienbild.



Abbildung 8: Grauwertbild und dazugehöriges Linienbild

An PKW's, LKW's etc. treten Linien am Dach, an den Seiten, im Bereich der Stoßstange und im Bereich des Schlagschattens auf (siehe Abb. 8). Gerade die horizontalen Linien sind sehr markante Merkmale dieser Objekte.

Da für dieses Verfahren bereits eine Hardware besteht, die das Linienbild sehr effizient berechnet, wird es als ein Modul für die Segmentierung in dieser Arbeit genutzt.

#### 2.2.2 Die lokale Orientierungskodierung

Der von C. GOERICK [8] vorgeschlagene Algorithmus zur *lokalen Orientierungskodierung* ( $\text{LOC}^2$ ) kodiert Grauwertdifferenzen in einer Vierer- ( $N_4$ ) oder Achter-Nachbarschaft ( $N_8$ ). Den Bildpunkten in der Nachbarschaft des Referenzbildpunktes R werden dabei folgende Koeffizienten zugeordnet:

$$\begin{bmatrix} 0 & 1 & 0 \\ 2 & R & 4 \\ 0 & 8 & 0 \end{bmatrix} \begin{bmatrix} 1 & 2 & 4 \\ 8 & R & 16 \\ 32 & 64 & 128 \end{bmatrix} \stackrel{n}{\frown}_{\to m}$$
(5)  
LOC-N<sub>4</sub> LOC-N<sub>8</sub>

Die Grauwerte der Bildpunkte aus der Nachbarschaft werden mit dem Grauwert des Referenzpunktes R verglichen und ihre Koeffizienten (siehe Abb. 5) aufsummiert, wenn ihre Differenz kleiner als ein vorgegebener Schwellwert ist. Um die Operation sensitiv für das Auffinden von Kanten und Ecken zu machen, können mit Hilfe einer implementierten Schwellwertbestimmung nur Grauwertsprünge genügender Höhe betrachtet werden. Das Beispiel in Abbildung 9 soll die Kodierung mit der LOC-N<sub>4</sub> Operation verdeutlichen.



Abbildung 9: Kodierung des LOC-N<sub>4</sub>

Alle Konturen im Bild können mit Hilfe dieses einfachen Operators markiert werden. Aus der Verteilung der Merkmale kann auf mögliche Objekte geschlossen werden. So treten an Fahrzeugen, aufgrund ihrer Geometrie, horizontale Konturen überproportional häufig auf (siehe Abb. 10).

<sup>&</sup>lt;sup>2</sup>Akronym: Local Orientation Coding



Abbildung 10: Grauwertbild, LOC-N<sub>4</sub>: alle Merkmale markiert, LOC-N<sub>4</sub>: nur horizontale Merkmale markiert.

### 2.2.3 Die Schlagschattendetektion

In natürlichen Straßenverkehrsszenen kommt das Licht immer von oben. Dadurch ensteht unterhalb der Fahrzeuge ein *Schlagschatten*. Dieses sehr markante Merkmal von Fahrzeugen versucht ein am Institut für Neuroinformatik entwickeltes Modul zur Schlagschattendetektion zu finden [11].

Es durchsucht die Bilder von unten nach oben nach horizontalen Konturen. Konturen dessen positiver Grauwertgradient nach unten zeigt, werden als untere Kante eines Schlagschattens interpretiert. Über den so gefundenen Schlagschatten wird aufgrund der Geometrie von Fahrzeugen ein Quadrat gesetzt. Im Ergebnisbild werden dadurch die Bereiche, in denen wahrscheinlich ein Fahrzeug abgebildet ist, markiert.



Abbildung 11: Grauwertbild, Schlagschattendetektion, Segmentierungsergebnis auf Basis der Schlagschattendetektion

Das zur Verfügung stehende Modul findet PKW's, LKW's und Transporter mit einer hohen Treffsicherheit. Der Hintergrund beinhaltet aber auch viele horizontale Konturen, die als Schlagschatten bewertet werden. Da es, im Gegensatz zu den anderen bildbasierten Verfahren, Vorwissen über die Art der vorkommenden Objekte benutzt, ist es gut geeignet, zu einer robusteren Segmentierung von Fahrzeugen beizutragen.

## 3 Bildsequenzanalyse

## 3.1 Differenzbildanalyse

Viele Segmentierungsalgorithmen verwenden die Differenzen der Grauwerte zweier aufeinanderfolgender Bilder zur Detektion von Bewegung. In den Bereichen, in denen sich die Grauwerte ändern, befindet sich ein Objekt. Objekte die sich nicht bewegen, können nicht segmentiert werden. Ein hierauf aufbauender Algorithmus, der von R. MECH U. M. WOLLHORN in [20] vorgestellt wurde, eignet sich zur Datenkompression von Videobildern und wurde auch für den aufkommenden Standard ISO/MPEG-4 diskutiert.

Ein Bild mit einem einmal abgespeicherten Hintergrund zu vergleichen, ist eine etwas andere Art dieser Technik. Diese kann zur Verkehrsüberwachung mit einer stationären Kamera genutzt werden, wie es in [6] oder [17] gezeigt wird.

Zur Segmentierung von Objekten vor einem dynamischen Bildhintergrund, der auch von einer bewegten Kamera verursacht sein kann, ist dieses Verfahren nicht geeignet.



Abbildung 12: Beispiel für ein Differenzbild

#### 3.2 Verschiebungsvektorfelder

Ein Verschiebungsvektorfeld gibt an, wie in einem Bild einer Sequenz Bildbereiche verschoben werden müssen, damit das darauf folgende Bild entsteht [9]. Da sich Verschiebungsvektoren nur an markanten Punkten berechnen lassen, müssen diese vorher durch einen Deskriptor<sup>3</sup> ermittelt werden. Diese Punkte, die starke Grauwertänderungen markieren, können mit unterschiedlichen Verfahren ermittelt werden.

Ein Verfahren, das als Deskriptor die lokale Orientierungskodierung (siehe Kap. 2.2.2) verwendet, wird in [26], Seite 7, als  $Flu\beta$ feldanalyse vorgestellt. Dabei wird zunächst die



Abbildung 13: Flußfeld [26]

 $LOC-N_4$  zweier aufeinanderfolgender Bilder berechnet und danach die markanten Punkte des ersten Bildes in einer Achter-Nachbarschaft der zweiten Bildes gesucht. Zur Erfassung größerer Verschiebungen wird hierbei mit drei verschiedenen Auflösungen gearbeitet.

 $<sup>^{3}\</sup>mathrm{Deskriptor}$  bezeichnet allgemein einen Algorithmus zur Punktfindung

Dieses Verfahren eignet sich, überholende oder einscherende Fahrzeuge auf Autobahnen oder Landstraßen zu detektieren. Eine sichere Segmentierung besonders dieser Fahrzeuge ist wichtig, da sie ein erhöhtes Gefahrenpotential für die beteiligten Verkehrsteilnehmer darstellen. Die Kamerageometrie, der Fahrbahnverlauf und die Eigenbewegung müssen hierfür aber bekannt oder abschätzbar sein.

## 4 Multisensor-Analyse

Sensordaten liefern zusätzliche Informationen, die in den Kamerabildern nicht enthalten sind. Sie können verwendet werden, um eine robustere Segmentierung zu erhalten. Als Sensoren eigenen sich beispielsweise RADAR<sup>4</sup>-, LIDAR<sup>5</sup>- oder Infrarotsensoren.

Im folgenden wird das Radarverfahren näher erläutert, da es exemplarisch für externe Sensoreinkoppelungen in dieser Arbeit verwendet wird.

#### 4.1 Die RADAR-Daten

Unter dem Begriff *RADAR* werden die Methoden zur Entdeckung von Objekten und zur Bestimmung ihrer räumlichen Lage sowie ihres Bewegungszustandes mit Hilfe elektromagnetischer Wellen zusammengefaßt. Radarverfahren sind in der Lage, unabhängig von den Witterungsbedingungen, Entfernungsangaben über sehr große Reichweiten zu liefern [23]. Beim klassischen Pulsradar wird die Entfernung eines Objektes aus der Impulslaufzeit zum Radarziel und zurück ermittelt. Die Richtung des Radarziels wird aus der Winkelstellung einer gut bündelnden Antenne ermittelt.

Ein weiteres sehr häufig eingesetztes Verfahren ist das FMCW<sup>6</sup>-Radar. Bei diesem Verfahren wird ein in der Frequenz zwischen zwei Grenzen variierendes Mikrowellensignal verwendet. Aus der gemessenen Übertragungsfunktion der Meßstrecke wird durch eine Fouriertransformation die Impulsantwort errechnet. Die Impulsantwort kann genauso ausgewertet werden wie beim Pulsradar.



Abbildung 14: Prinzip einer phasengesteuerten Antenne

<sup>&</sup>lt;sup>4</sup>Akronym: RAdio Detection And Ranging

<sup>&</sup>lt;sup>5</sup>Akronym: LIght Detection And Ranging

<sup>&</sup>lt;sup>6</sup>Akronym: Frequency Modulated Continuous Wave

Als Antennen werden häufig phasengesteuerte Antennen verwendet. Sie bestehen aus vielen Strahlerelementen, deren Phasenlage einzeln eingestellt wird (siehe Abb. 14). Dadurch ist es möglich, die Hauptstrahlungsrichtung zu ändern.

Zur Erzeugung der Radardaten, die in dieser Arbeit verwendet werden, wurde ein Radarsystem benutzt, das nach einer Raumabtastung drei Objekte erfaßt und mit einer schwenkbaren Antennenkeule verfolgt. Die vorliegenden Daten sind bereits so aufbereitet, daß sie direkt die Position des Fußpunktes und die Breite des Objektes, im dazugehörigen CCD<sup>7</sup>-Kamerabild, beschreiben. Die Synchronisation zwischen den einzelnen Bildern und den Radardaten erfolgt dabei über einen Time-Code<sup>8</sup>. Zur Auswertung wird angenommen, daß sich die erfaßten Objekte oberhalb des Fußpunktes befinden. Aufgrund der Geometrie der Fahrzeuge werden sie mit der ermittelten Breite quadratisch abgebildet.

Die Qualität der Daten ist relativ schlecht. Häufig erfassen alle drei Antennenkeulen das gleiche Objekt, wie in Abbildung 15 und nicht immer finden sie die Mitte der Objekte. Zudem wurde der Fahrbahnverlauf bei der Datenaufbereitung nur unzureichend berücksichtigt.



Abbildung 15: Grauwertbild, Radarbild, Segmentierung auf Basis des Radarbildes. Alle Antennenkeulen erfassen das gleiche Objekt

Desweiteren werden in komplexeren Szenen nicht alle im Bild vorkommenden Objekte erfaßt, da systembedingt nur maximal drei Objekte verfolgt wurden (siehe Abb. 16).



Abbildung 16: Grauwertbild, Radarbild, Segmentierung auf Basis des Radarbildes. Es werden maximal drei Objekte erfaßt.

Dennoch liefen die Radardaten Zusatzinformationen, die zu einer robusteren Segmentierung beitragen. Zukünftige, zuverlässigere Radarsensoren werden dies begünstigen.

 $^{8}\mathrm{Der}$  Time-Code des Bildes ist im Grauwertbild in Abb. 16 oben links als störende Linie sichtbar.

<sup>&</sup>lt;sup>7</sup>Akronym: Charged Couple-Device

## 5 Fusion

Durch die Fusion geeigneter Bildmerkmale wird das Segmentierungsergebnis gegenüber den Einzelalgorithmen verbessert [10]. Dem Entscheidungsprozeß liegen mehr unabhängige Informationen vor und seine Entscheidung wird dadurch qualifizierter.

Da allen Bildanalyseverfahren die gleiche Datenbasis zugrunde liegt, sind die resultierenden Bildmerkmale untereinander korreliert. Dies entspricht einer Redundanz. Sie wird verwendet um das Segmentierungsergebnis *robuster*, d.h. unanfälliger gegen falsche Informationen zu machen. So können beispielsweise konturbasierte Verfahren, die typischerweise differenzierende Eigenschaften besitzen und damit anfällig gegen Bildfehler und Rauschen sind, sich gegenseitig unterstützen.

Zusätzliche Sensorik ergänzt die Information aus den Bildanalyseverfahren und verbessert so den Entscheidungsprozeß.

## 5.1 Hierarchieebenen der Fusionsprozesse

Die Fusion von Eingabevariablen kann auf verschiedenen *hierarchischen Ebenen* erfolgen [5]. Da allgemein zwischen der Daten-, der Merkmals- und der Entscheidungsebene unterschieden wird, heißen die Fusionsprozesse entsprechend:

- Daten-Fusion (data-fusion),
- Merkmals-Fusion (feature-fusion),
- Entscheidungs-Fusion (decision-fusion).

Um die Begriffe zu verdeutlichen, ist es sinnvoll, sich am Beispiel eines Grauwertbildes dem Ablauf einer Segmentierung auf der Basis eines einzelnen Bildmerkmals zuzuwenden. Die CCD<sup>9</sup>-Kamera liefert die Rohdaten, die als Bildpunkte auf einer Koordinatenebene abgebildet werden. Aus dem entstehenden Grauwertbild wird ein Merkmalsbild extrahiert und zur Segmentierung einem Klassifikator übergeben. Der Klassifikator weist jedem Bildpunkt die Klasse "Objekt" oder "Hintergrund" zu und besteht in diesem Fall aus einem einfachen Schwellwertverfahren (siehe Abb. 17).



Abbildung 17: Prinzip einer Segmentierung auf Basis eines Bildmerkmals

## 5.2 Fusion der Merkmale

Die Fusion mehrerer Merkmale ist durch ein Koppelungsnetzwerk in der Merkmalsebene und in der Entscheidungsebene möglich. Bei einer Entscheidungs-Fusion, wie sie Abbildung 18 zeigt, muß das Koppelungsnetzwerk aufgrund der Zuverlässigkeit der einzelnen

<sup>&</sup>lt;sup>9</sup>Akronym: Charged Couple-Device

Klassifikatorentscheidungen sein Ergebnis bestimmen. Das Koppelungsnetzwerk kann im einfachsten Fall aus logischen Elementen bestehen, die beispielsweise eine Mehrheitsentscheidung herbeiführen.



Abbildung 18: Prinzip einer Entscheidungs-Fusion

Die Merkmals-Fusion (siehe Abb. 19) maximiert den Beitrag der unterschiedlichen Informationen, die in das Koppelungsnetzwerk einfließen. Sie ist daher der Entscheidungs-Fusion vorzuziehen [16].



Abbildung 19: Prinzip einer Merkmals-Fusion

#### 5.3 Multisensor-Fusion

Die Fusion der Bildmerkmale mit weiteren Sensordaten verbessert den Entscheidungsprozeß durch zusätzliche Informationen, die unabhängig von den Bilddaten sind. Als Sensoren eigenen sich beispielsweise RADAR-, LIDAR- oder Infrarotsensoren. Wird, wie in dieser Arbeit, eine bildpunktbezogene Fusion angestrebt, so müssen die Rohdaten dieser Sensoren in die gleiche bildpunktorientierte Basisrepräsentation projiziert werden, die für das Grauwertbild und den daraus extrahierten Merkmalen gewählt wurde. So läßt sich die Fusion der Bildmerkmale auf eine Multisensor-Fusion verallgemeinern (siehe Abb. 20). Je nach Sensortyp ist eventuell eine Merkmalsextraktion erforderlich (in Abb. 20 gestrichelt dargestellt).



Abbildung 20: Prinzip einer Multisensor-Fusion auf der Merkmalsebene

### 5.4 Koppelungsnetzwerk

Das *Koppelungsnetzwerk* fusioniert auf der Merkmalsebene die eingehenden Informationen. Zur Lösung der gestellten Aufgabe soll dabei das Netzwerk folgende Anforderungen erfüllen:

- Da die Entwicklung neuer Bildverarbeitungsverfahren zur Merkmalsextraktion ständig fortschreitet, soll eine *zukünftige Einkoppelung* neuer Verfahren einfach realisierbar sein.
- Stehen zusätzliche Informationen aus externen Sensoren zur Verfügung, so sollen sie *flexibel* eingekoppelt werden können.
- Das Koppelungsnetzwerk soll seine Entscheidung in möglichst kurzer Zeit treffen.
- Das Koppelungsnetzwerk soll *robust* sein gegen Störungen. Rauschen, Übertragungsfehler oder der Ausfall einzelner Module darf nicht zum Ausfall des gesamten Systems führen.
- Gibt ein Modul fehlerbehaftete oder ungenaue Ergebnisse, so soll das Gesamtergebnis weiterhin richtig sein. In dieser Hinsicht soll es *fehlertolerant* sein.

Aufgrund dieser Forderungen wird als Koppelungsnetzwerk ein Multilayer-Perzeptron mit einer verdeckten Schicht verwendet, wie es im Kapitel 6.3 erklärt wird. Vorwärtsgerichtete Netze mit einer verdeckten Schicht können selbst komplizierte Funktionen gut approximieren [2]. Die Synapsengewichte können flexibel gelernt werden. Durch einfachen Austausch aller Gewichte kann das Netz an geänderte Bedingungen angepaßt werden. Durch die Möglichkeit einer hochgradig parallelen Verarbeitung, kann eine hardwaremäßig parallele Implementation die Ergebnisberechnung sehr schnell durchführen.

## 6 Künstliche neuronale Netze

Bei neuronalen Netzen handelt es sich im allgemeinen um Strukturen, die durch einen Lernvorgang zur Lösung einer bestimmen Aufgabe befähigt werden. Sie setzten sich meist

aus Grundelementen zusammen, die geeignet untereinander verbunden sind und nach einem Selbstorganisationsvorgang – dem Lernvorgang – als Ganzes die gestellte Aufgabe bewältigen. Mit künstlichen neuronalen Netzen werden die biologischen neuronalen Netze als informationsverarbeitende Systeme nachgeahmt [22]. Ein wesentlicher Unterschied zu herkömmlichen Verfahren der Informationsverarbeitung, ist die hochgradig parallele Verarbeitung. Es gibt verschiedene Arten künstlicher neuronaler Netze, wie beispielsweise das Multilayer-Perzeptron, das radiale Basisfunktions-Netz, das Hopfield-Netz, Kohonenkarten, etc. Im folgenden wird nach einer kurzen biologischen Einführung, daß in dieser Arbeit verwendete Multilayer-Perzeptron vorgestellt.

#### 6.1 Das biologische Vorbild

Das biologische Vorbild Gehirn besteht aus vielen vernetzten Nervenzellen, den Neuronen. Beispielsweise besteht das menschliche Gehirn aus  $10^{12}$  Neuronen [13].

An einem typischen Neuron lassen sich drei Hauptstrukturen unterscheiden, der *Dendritenbaum*, der *Zellkörper* und das *Axon*. Ihnen können die Aufgaben Eingabe, Verarbeitung und Ausgabe zugeordnet werden [21]. Die Abbildung 21 zeigt den schematischen Aufbau eines im menschlichen Gehirn am häufigsten vorkommenden Neurons, der Pyramidenzelle.



Abbildung 21: Schematischer Aufbau einer Pyramidenzelle [21].

Der Dendritenbaum bildet den Haupteingabepfad dieser Zelle. Er summiert die Ausgangssignale der umgebenden Neuronen in Form eines elektrischen Potentials, das er dem Zellkörper zuführt. Überschreitet dieses Potential einen gewissen Schwellenwert, so erzeugt der Zellkörper einen elektrischen Nadelimpuls, der von einer Nervenfaser, dem Axon, weitergeleitet wird. Das Axon, das Bruchteile von Millimetern bis zu mehreren Metern lang sein kann, führt den Impuls an bis zu mehrere tausend Zielneuronen. Die Kontaktstellen des Axons befinden sich meist auf dem Dendritenbaum oder dem Zellkörper der Zielneuronen und heißen Synapsen. Die meisten Synapsen sind chemische Kontakte. Der elektrische Impuls des Axons bewirkt an der Synapse die Ausschüttung eines Überträgerstoffes (Neurotransmitter), der dann wiederum zu einer Potentialänderung am Dendritenbaum bzw. am Zellkörper des Zielneurons führt. Je nach Art und Zustand der Synapse bewirkt ein eintreffender Impuls eine mehr oder weniger starke Potentialerhöhung (erregende Synapse) oder Potentialerniedrigung (hemmende Synapse).

#### 6.2 Das Perzeptron

Das Perzeptron wurde 1957 von F. ROSENBLATT entwickelt und ist das einfachste Netz zur Klassifikation von Eingabemustern [13]. Es besteht grundsätzlich aus einem einzigen Neuron mit veränderbaren Synapsengewichten und einem Schwellwert (siehe Abb. 22). Es



Abbildung 22: Perzeptron mit einer Schicht (Single-Layer-Perzeptron)

ist in der Lage Musterklassen zu trennen, die *linear separierbar* sind, also Musterklassen die durch eine Hyperebene im p-dimensionalen Raum getrennt werden können. Mit nur einem Neuron ist es ausschließlich in der Lage, zwischen zwei Musterklassen zu unterscheiden. Mit mehreren Neuronen in der Ausgangsschicht, die mit den gleichen Eingängen verbunden sind, kann zwischen mehreren Klassen unterschieden werden. Für jedes Perzeptron müssen die unterschiedlichen Klassen aber weiterhin linear separierbar bleiben.

#### 6.3 Das Multilayer-Perzeptron

Das Multilayer-Perzeptron (MLP) [13] ist eine Verallgemeinerung des Single-Layer-Perzeptrons aus Kapitel 6.2. Es besteht aus Eingangszellen<sup>10</sup>  $s_k$ , einer oder mehreren verdeckten Neuronenschichten mit den Neuronenaktivitäten  $s_j$  und einer Neuronenschicht zur Ausgabe mit den Aktivitäten  $s_i$ . Die Eingangsmuster werden bei diesem Netztyp Schicht für Schicht durch das Netz propagiert. Die Abbildung 23 zeigt solch ein MLP mit einer verdeckten Schicht.

Jedes Neuron kann über ein Synapsengewicht w durch den Bias  $s_0$  voraktiviert werden. Dies ist äquivalent zum Schwellwert  $\sigma$  des Perzeptrons von Rosenblatt (vergl. Abb. 22). Ein entscheidender Unterschied zum Perzeptron von Rosenblatt ist die nichtlineare Aktivierungsfunktion  $\sigma(x)$  am Ausgang der Neuronen (siehe Abb. 24).

Die Aktivität eines Neurons  $s_j$  in der verdeckten Schicht j ist gegeben durch:

$$s_j = \sigma\left(\sum_k w_{jk} s_k\right). \tag{6}$$

<sup>&</sup>lt;sup>10</sup>Die Eingangszellen sind die sensorischen Einheiten des MLP und nicht zu verwechseln mit den Neuronen.



Abbildung 23: Zweischichtiges Multilayer-Perzeptron



Abbildung 24: Nichtlineares Modell eines Neurons

Die Aktivität eines Neurons  $s_i$  in der Ausgabeschicht i ist entsprechend gegeben durch:

$$s_i = \sigma\left(\sum_j w_{ij} s_j\right). \tag{7}$$

Sie versetzt das MLP in die Lage, nach einem Trainingsprozeß, verschiedenste komplizierte Probleme zu lösen. Selbst nicht linear separierbare Musterklassen können erfolgreich unterschieden werden. Der verbreitetste Trainingsalgorithmus für ein MLP ist der *Backpropagation Algorithmus*. Dieser wird im folgenden Kapitel näher erläutert. Er propagiert ein Fehlersignal rückwärts durch das Netz und verändert dabei die Synapsengewichte w, um den Fehler zu minimieren. Voraussetzung dafür ist eine stetige und überall differenzierbare Aktivierungsfunktion der Neuronen. Im allgemeinen wird die sigmoide nichtlineare Fermifunktion  $\sigma(x)$  benutzt:

$$\sigma(x) = \frac{1}{1 + e^{-x}}.\tag{8}$$

Der Graph der Fermifunktion ist zur Veranschaulichung in Abbildung 25 dargestellt. Sie bewegt sich im Bereich  $0 < \sigma(x) < 1 \forall x \in \mathbb{R}$  und kann zur Modellierung von Sättigungseffekten eingesetzt werden.



Abbildung 25: Fermifunktion

#### 6.4 Der Backpropagation-Lernalgorithmus

Der *Backpropagation-Lernalgorithmus* ist das Standardlernverfahren für vorwärtsgekoppelte Netze. Er wird hier am Beispiel eines zweischichtigen Netzes erklärt. Der Algorithmus läßt sich auf beliebigschichtige Netze verallgemeinern.

Im folgenden soll k als Index für die Netzeingänge benutzt werden, j für die verdeckte Schicht und i für die Ausgabeschicht,  $(i, j, k \in \mathbb{N})$  siehe Abb. 23). Der Eingang hat die Dimension D  $(D \in \mathbb{N})$ . H und O beschreiben die Zahl der Neuronen in der verdeckten Schicht und der Ausgabeschicht  $(H, O \in \mathbb{N})$ . Die Eingangsvariablen und der Ausgangssollvektor werden mit **x** bzw. **y** bezeichnet und s beschreibt die Aktivität eines Neurons ( $\mathbf{x} \in \mathbb{R}^D$ ,  $\mathbf{y} \in \mathbb{R}^O$ ,  $s \in \mathbb{R}$ ). Neuronen allgemein werden mit a, b indiziert,  $(a, b \in \mathbb{N})$ . Den Bezug zu einzelnen Trainingsmustern beschreibt  $\nu$  ( $\nu \in \{1, 2, ..., p\}$ ). Die Anzahl der Trainingsmuster entspricht p.

Werden die Eingabemuster  $\mathbf{x}^{\nu}$  während des Trainingsprozesses in Vorwärtsrichtung vom Netz verarbeitet, so ergeben sich Differenzen zwischen den berechneten und vorgegebenen Ausgangsmustern  $\mathbf{y}^{\nu}$ . Der über alle Mustepaare  $(\mathbf{x}^{\nu}, \mathbf{y}^{\nu})$  erstreckte quadratische Fehler E wird als Gütekriterium des Netzes benutzt.

$$E = \frac{1}{2} \sum_{\nu=1}^{p} \sum_{i} \left( y_{i}^{\nu} - s_{i}(\mathbf{x}^{\nu}) \right)^{2}.$$
 (9)

Der Proportionalitätsfaktor 1/2 in Gleichung 9 ist zur Vereinfachung der folgenden Rechnung eingeführt. Das Ziel ist E zu minimieren, bzw. die Gewichte  $w_{ij}$  und  $w_{jk}$  so anzupassen, daß E ein Stück in Richtung des negativen Gradienten  $\partial E/\partial w_{ab}$  verschoben wird. Dazu ändert man alle Gewichte  $w_{ab}$  wiederholt um:

$$\Delta w_{ab} = -\alpha \cdot \frac{\partial E}{\partial w_{ab}}.$$
(10)

Für ein hinreichend kleines  $\alpha > 0$  bewegt man sich in Richtung des steilsten Abfalls von *E*. Für die Fehleränderung  $\Delta E$  gilt bei einem solchen Schritt näherungsweise [21]:

$$\Delta E \approx \sum_{ab} \frac{\partial E}{\partial \Delta w_{ab}} \Delta w_{ab} = -\alpha \sum_{ab} \left(\frac{\partial E}{\partial w_{ab}}\right)^2 \le 0.$$
(11)

Falls die Aktivierungsfunktion ableitbar ist, ergibt sich der Gradient  $\partial E / \partial w_{ab}$  durch die Anwendung der Kettenregel. Für die Gewichte  $w_{ij}$  zwischen der mittleren und der Ausgabeschicht ergibt sich:

$$\frac{\partial E}{\partial w_{ij}} = -\sum_{\nu} \left( y_i^{\nu} - s_i(\mathbf{x}^{\nu}) \right) \cdot \sigma' \left( \sum_{j'} w_{ij'} s_{j'} \right) \cdot s_j \tag{12}$$

Für die Gewichte  $w_{jk}$  zwischen der Eingabe- und der mittleren Schicht wird über alle Ausgänge i summiert. Es ergibt sich:

$$\frac{\partial E}{\partial w_{jk}} = -\sum_{\nu} \sum_{i} \left( y_{i}^{\nu} - s_{i}(\mathbf{x}^{\nu}) \right) \cdot \sigma' \left( \sum_{j'} w_{ij'} s_{j'} \right) \cdot w_{ij} \cdot \frac{\partial s_{j}}{\partial w_{jk}}$$

$$= -\sum_{\nu} \sum_{i} \left( y_{i}^{\nu} - s_{i}(\mathbf{x}^{\nu}) \right) \cdot \sigma' \left( \sum_{j'} w_{ij'} s_{j'} \right) \cdot w_{ij}$$

$$\cdot \sigma' \left( \sum_{k'} w_{jk'} s_{k'} \right) \cdot s_{k}$$
(13)

Beide Gleichungen bilden die Summe aus den Beiträgen aller Musterpaare  $(\mathbf{x}^{\nu}, \mathbf{y}^{\nu})$ . Daraus folgt nur ein kleiner Unterschied (für hinreichend kleine  $\alpha > 0$ ), wenn für jede Wiederholung der Gewichtsänderungen (Gl. 10) nur ein Musterpaar  $(\mathbf{x}^{\nu}, \mathbf{y}^{\nu})$  berücksichtigt wird und alle Musterpaare im Mittel gleich oft trainiert werden.

Mit der Substitution  $\epsilon_i^{\nu} = y_i^{\nu} - s_i(\mathbf{x}^{\nu})$  für den Ausgabefehler beim  $\nu$ -ten Musterpaar  $(\mathbf{x}^{\nu}, \mathbf{y}^{\nu})$  und der für die Fermifunktion gültigen Substitution  $\sigma'(x) = \sigma(x)(1 - \sigma(x))$ , lassen sich die Änderungen der Gewichte  $w_{ij}$  und  $w_{jk}$  wie folgt beschreiben:

$$\Delta w_{ij} = \alpha \cdot \epsilon_i^{\nu} \cdot s_j s_i (1 - s_i) \tag{15}$$

$$\Delta w_{jk} = \alpha \cdot \sum_{i} \epsilon_{i}^{\nu} \cdot s_{k} s_{i} (1 - s_{i}) \cdot w_{ij} \cdot s_{j} (1 - s_{j})$$

$$\tag{16}$$

Die Fehlerfunktion E kann eine sehr komplexe Funktion darstellen, die von allen Gewichten  $w_{ab}$  abhängt und mehrere lokale Minima besitzen kann. Deshalb ist das, mittels diesem Verfahren gefundene Minimum des Fehlers E, nicht unbedingt auch das globale Minimum. Ob ein gutes Minimum gefunden wird, hängt stark von den Initialwerten der Gewichte ab. Dadurch ist es hin und wieder notwendig, ein Netz, daß den Fehler E schlecht minimiert, mehrfach neu zu initialisieren und zu trainieren.

## 7 Praktischer Teil

#### 7.1 Ausgangspunkt der Arbeit

Ausgangspunkt dieser Arbeit ist die Fusion von integrativen und differentiellen Bildmerkmalen auf der Merkmalsebene (siehe Abb. 26). Integratives Merkmal ist hierbei die Entropie. Die differentiellen Merkmale sind der LOC-N<sub>4</sub> und das Linienbild. Aus der Entro-



Abbildung 26: Ausgangspunkt der Arbeit

pie werden durch Schwellwertoperationen fünf Merkmale abgeleitet. Die 16 verschiedenen Konturcodes des LOC-N<sub>4</sub> werden auf vier Richtungscodes reduziert. Danach wird über einen  $(9 \times 9)$  Punkte großen Bildbereich *B* die Häufigkeit jedes Merkmals ermittelt. Der so ermittelte Merkmalsvektor wird zusätzlich um die Zentrumskoordinaten (m, n) des Bildbereichs *B* erweitert, den er beschreibt. Da die Fahrzeuge aufgrund des Fahrbahnverlaufs und der Kamerageometrie nicht in allen Bildbereichen abgebildet werden können, leistet diese Information einen zusätzlichen Beitrag zur Unterscheidung von Objekt und Hintergrund. Das Koppelungsnetzwerk ist ein Multilayer-Perzeptron mit 12 Eingangszellen, 5 Neuronen in einer verdeckten Schicht und einem Ausgangsneuron ([26], Seite 20).

Das System liefert motivierende Ergebnisse bei einfachen Szenen (siehe Abb. 27). Es zeigt aber deutliche Schwächen bei der Segmentierung von Fahrzeugen in kontrastarmen Bildregionen. Bei komplizierteren Szenen werden oft Fahrbahnbegrenzungen segmentiert. LKW's werden nur teilweise erkannt.



Abbildung 27: Segmentierungsbeispiel

Die motivierenden Ergebnisse sollen im folgendem hinsichtlich der Separierbarkeit von Objekten und Hintergrund verbessert werden. Das System soll zusätzlich robuster gegen

Störungen und fehlertolerant gegen den Ausfall eines Moduls werden. Eine möglichst flexible Einkoppelung externer Sensorik soll hierzu einen Beitrag liefern.

### 7.2 Datenmaterial

Das vorliegende Bildmaterial wurde ausschließlich mit einer CCD-Kamera aufgenommen, die in der Position des Rückspiegels in einem Fahrzeug angebracht worden ist. Die Daten wurden als Halbframes auf einer Bildplatte analog gespeichert. Zum Auslesen der Daten werden diese durch den Digitalisierungsvorgang in 256 Graustufen quantisiert. Die daraus resultierenden Gauwertbilder besitzen eine Auflösung von (736 × 256) Bildpunkten. Zur Lösung der Segmentierungsaufgabe werden sie auf eine Größe von (184 × 128) Bildpunkten verkleinert.

In dieser Arbeit wurden drei verschiedene Bildsequenzen betrachtet:

Golf-Szene:	Einfache Autobahnszene mit einem Golf als vorausfahrendem Fahrzeug.
A42-Szene:	Autobahnszene mit Schallschutzmauer und LKW's. Die Fahrzeuge befinden sich häufig nur am Horizont.
Paris-Szene:	Szene einer dreispurigen Autobahn mit LKW's. Die Kamera ist weit nach unten geneigt. Dies soll eine Erkennung des Fahrbahnverlaufs ermöglichen. (Die Szene ist dunkel.)

Für die Paris-Szene stehen Radardaten zur Verfügung. Sie werden über einen Time-Code mit den Bildern synchronisiert. Die Radardaten werden ausführlich in Kapitel 4.1 vorgestellt.

### 7.3 Merkmalsextraktion

Die Bildanalyseverfahren stellen die Basisinfomationen bereit, die für eine Segmentierung benötigt werden. Die *Merkmalsextraktion* extrahiert aus diesen Basisinformationen die Merkmale, die eine Relevanz für den Entscheidungsprozeß besitzen und übergibt sie an das Koppelungsnetzwerk.

#### 7.3.1 Erweiterung der Konturmerkmale

Als Konturmerkmale wurden bisher das Linienbild und vier Richtungsmerkmale des LOC- $N_4$  zu Fusion verwendet (siehe Kap. 7.1). Eine Erweiterung dieses Merkmale soll eine Verbesserung des Ergebnisses bringen.

**Hypothese 1** Je mehr unabhängige Informationen ein Entscheidungproze $\beta$  erhält, desto qualifizierter ist sein Urteil.

Auf Grund dieser Hypothese, wurde die lokale Orientierungskodierung in einer Vierer-Nachbarschaft (LOC-N<sub>4</sub>), gegen eine lokale Orientierungskodierung in einer Achter-Nachbarschaft (LOC-N<sub>8</sub>) ausgetauscht. Aus den nun möglichen 256 verschiedenen Konturcodes werden die betrachtet, die von ihren Eigenschaften Teil einer geraden Konturlinie sein können.

Zum LOC-N<sub>4</sub> gibt es äquivalente Eigenschaften, wie horizontale, vertikale, im  $45^{\circ}$  Winkel steigende und fallende Konturen. Darüber hinaus werden Konturen mit differierenden Steigungen kodiert. Zusätzlich steigt die Auflösung bei einer Achter-Nachbarschaft. Schon

die ersten Versuche mit einfachen Merkmalseinteilungen erzielten eine Verbesserung des Ergebnisses.

Im Laufe der Arbeit wurden die Merkmalsunterteilungen erweitert, variiert und vervollständigt. Mit 14 LOC-N<sub>8</sub> Merkmalen wurde eine Häufigkeitsanalyse durchgeführt. Eine Zusammenfassung nur selten auftretender Merkmale wurde möglich.

Die Häufigkeitsanalyse ergab starke Unterschiede zwischen Merkmalen mit positiven und negativen Steigungen, der sonst nur an der vertikalen Achse gespiegelten Konturcodes. Dies wurde auf eine einseitige Beleuchtung der Trainings-Szenen (das Sonnelicht kommt dort stets von links) und schlecht markierte Objekte in den Datenbanken zurückgeführt.

Um eine Generalisierung auch für Szenen mit einer anderen Beleuchtungsrichtung zu erreichen, wurde zwischen den Merkmalen der Kontur mit positiven und negativen Steigungen nicht mehr unterschieden. Sieben Merkmale konnten so extrahiert werden. Abbildung 28 zeigt die Zusammenfassung der Konturcodes zu diesen Merkmalen. Mit den so gewählten Merkmalen wurden die besten Ergebnisse erzielt.



Abbildung 28: Reduktion der Merkmale einer achter Nachbarschaft

#### 7.3.2 Die lokale Orientierungskodierung als Texturanalyse

Das Verhalten des LOC ist stark von dem Schwellwert abhängig. Der Schwellwert gibt den kleinsten Grauwertsprung an, der noch detektiert wird. Durch eine sehr keine Schwelle werden auch kleine Grauwertsprünge erfaßt, die häufig in Texturen vorkommen.

Ein LOC-N<sub>4</sub> soll auf diese Weise, mit einer fest vorgegebenen Schwelle thr = 4, die texturorientierten Verfahren unterstützen. Eine kleinere Schwelle ist nicht sinnvoll, da sonst hauptsächlich Rauschen analysiert wird. Aus den 16 möglichen Konturcodes wurden vier Richtungsmerkmale extrahiert, die Konturen von 0° (horizontal), 45°, 90° (vertikal) und 135° beschreiben (siehe Abb. 29).



Abbildung 29: Reduktion der Merkmale einer vierer Nachbarschaft

Das Segmentierungsergebnis verbesserte sich hierdurch weiter.

#### 7.3.3 Varianz

Ein weiteres texturorientiertes Verfahren, das den Merkmalsraum erweitern soll, ist die Varianz. Bildbereiche mit sehr schwachen Grauwertschwankungen besitzen eine kleine Varianz, wie beispielsweise der Fahrbahnbereich. In Bildbereichen mit stärkeren Grauwertschwankungen befinden sich die Objekte (größere Varianz). Auf Grund dieser Beobachtung wurde untersucht, ob eine Segmentierung auf Basis der Varianz sinnvoll ist.

Die maximale Varianz ergibt sich nach Gleichung 2, wenn die Hälfte aller betrachteten Bildpunkte den maximal und die andere Hälfte den minimal möglichen Grauwert annimmt, zu:

$$V_{max} = \frac{1}{P-1} \left[ \frac{P}{2} \left( \max(g) - \langle G \rangle \right)^2 + \frac{P}{2} \left( \min(g) - \langle G \rangle \right)^2 \right]$$
(17)

Der mittlere Grauwert ergibt sich dabei nach Gleichung 1 zu:

$$\langle G \rangle = \frac{1}{P} \left[ \frac{P}{2} \max(g) + \frac{P}{2} \min(g) \right] = \frac{\max(g) + \min(g)}{2}$$
(18)

Mit einem minimal möglichen Grauwert  $\min(g) = 0$  ergibt sich für Gleichung 17:

$$V_{max} = \frac{1}{P-1} \left[ \frac{P}{2} \left( \max(g) - \frac{\max(g)}{2} \right)^2 + \frac{P}{2} \left( -\frac{\max(g)}{2} \right)^2 \right]$$
(19)

$$= \frac{1}{P-1} \left[ \frac{P}{2} \left( \frac{\max(g)}{2} \right)^2 + \frac{P}{2} \left( -\frac{\max(g)}{2} \right)^2 \right]$$
(20)

$$= \frac{1}{P-1} \left[ P\left(\frac{\max(g)}{2}\right)^2 \right]$$
(21)

$$= \frac{P}{P-1} \left(\frac{\max(g)}{2}\right)^2 \tag{22}$$

Wird der Bildbereich B sehr groß, so daß die Punktanzahl  $P \to \infty$  geht, bedeutet dies:

$$\lim_{P \to \infty} \{V_{max}\} = \lim_{P \to \infty} \left\{ \frac{P}{P-1} \left(\frac{\max(g)}{2}\right)^2 \right\}$$
(23)

$$= \left(\frac{\max(g)}{2}\right)^2 \tag{24}$$

Ein Varianz-Modul, daß als Teil dieser Arbeit implementiert wurde, arbeitet auf einem  $(7 \times 7)$  großen Bildausschnitt. P = 49 Bildpunkte werden berücksichtigt. Hierfür ergibt sich aus Gleichung 22 für ein Grauwertbild mit einem Grauwertbereich von  $g \in \{0, 1 \cdots 255\}$  eine theoretische maximale Varianz  $V_{max,th}$  von:

$$V_{max,th} = \frac{49}{49 - 1} \left(\frac{255}{2}\right)^2 = 16594,9 \tag{25}$$

Die höchste in den Bildern tatsächlich vorkommende Varianz ist deutlich niedriger und liegt bei

$$V_{max,pr} \approx 6242. \tag{26}$$

Eine Normierung mit der höchsten theoretischen Varianz  $V_{max,th}$  nach Gleichung 23 ist deshalb nicht sinnvoll. Zuviele Informationen gingen durch eine Quantisierung auf 256 Werte verloren. Daher wurde die Varianz so skaliert, daß sie einen Wertebereich von etwa  $g \in \{0, 1 \cdots 240\}$  umfaßt. Der Bereich wurde so gewählt, daß für Bilder, die eine etwas höhere Varianz aufweisen, noch eine Reserve vorhanden ist.

Eine Segmentierung auf Basis der Varianz ergibt nur sinnvolle Ergebnisse, wenn ein sehr kleiner Segmentierungsschwellwert vorgegeben wird. In Abbildung 2 (S. 3) wurde ein Schwellwert von 5 verwendet. (Zum Vergleich: Die Segmentierung auf Basis der Entropie in Abbildung 3 (S. 4) wurde mit einem Schwellwert von 90 erstellt.)

#### 7.3.4 Ortsabhängigkeit der Objektpunkte

Auf Grund der Kamerageometrie, -neigung und -position im Fahrzeug kann es Bildbereiche geben, in denen kein Fahrzeug abgebildet sein kann. Abbildung 30 zeigt eine Bild einer solchen Szene.

Ein neurorales Netz kann in einer solchen Szene lernen, in welchen Bildbereichen die Abbildung von Fahrzeugen wahrscheinlich ist, wenn zusätzlich zu den Bildmerkmalen am



Abbildung 30: Beispiel für eine Szene mit einer Ortsabhängigkeit der Objektpunkte. Im oberen Bildbereich können keine Fahrzeuge abgebildet sein.

Punkt (x, y) die Koordinaten (x, y) dem Netz übergeben werden (vgl. Kap. 7.1). Es lernt aber auch die Bildbereiche zu bevorzugen, in denen sich während des Trainings die meisten Bildpunkte befinden. Die Interpretation der Einflußnahme anderer Bildmerkmale wird dadurch erschwert. Wird beispielsweise ein Fahrzeug am linken Bildrand nicht segmentiert, so kann es dafür folgende Ursachen geben:

- Das neuronale Netz benachteiligt den linken Bildbereich, da zu viele rechte Objektbildpunkte oder linke Hintergundbildpunkte gelernt wurden.
- Das Fahrzeug wird nicht segmentiert, weil es als schwarzes Fahrzeug gegenüber einem dunklen Hintergrund keine großen Konturen besitzt.

Soll die Wahrscheinlichkeit des Vorhandenseins von Objekten in den verschiedenen Bildbereichen gelernt werden, müssen die Objektpunkte in den Trainingsdaten, möglichst ausgewogen, alle möglichen Positionen repräsentieren. Dies ist nur mit sehr großen Trainingsdaten zu erreichen und führt damit zu langen Trainingszeiten.

In anderen Bildsequenzen gibt es keine Ortsabhängigkeit der Objektpunkte. Die Kamera ist dabei so stark nach unten gerichtet, daß theoretisch an jeder Bildkoordinate Fahrzeuge abgebildet sein können, wie in Abbildung 31 dargestellt. Für eine solche Szene sind die



Abbildung 31: Beispiel für eine Szene ohne Ortsabhängigkeit der Objektpunkte. Im gesammten Bildbereich können Fahrzeuge abgebildet sein.

Koordinaten irrelevant. Werden sie dennoch dem neuronalen Netz übergeben, wird sich bei einer unausgewogenen Statistik der gelernten Objekt- und Hintergrundpositionen das Segmentierungsergebnis verschlechtern. Der während des Trainings gefundene Zusammenhang zwischen den Koordinaten und den Objekten ist nicht sinnvoll.

Um die Adaptabilität eines ausgelernten, neuronalen Netzes zu beurteilen, kann ein Test mit einer unbekannten Szene durchgeführt werden. Da alle drei verwendeten Szenen mit

unterschiedlichen Kameraneigungen aufgenommen wurden, ist eine gute Adaption an diese Szenen nur möglich, wenn keine Ortsabhängigkeiten gelernt wurden. Aus diesen Gründen wurden die Koordinaten aus dem Merkmalsvektor entfernt.

### 7.3.5 Untersuchung zur Einwirkung einer Gammakorrektur

Eine Gammakorrektur, die eine nichtlineare Transformation der Grauwerte in einem Grauwertbild durchführt, wurde implementiert und der Einfluß auf das Gesamtsystem untersucht. Wenn eine Szene ohne Gammakorrektur gelernt wurde, waren die Ergebnisse ohne Gammakorrektur die besten. Mit Hilfe der Gammakorrektur konnte zwar die Größe der segmentierten Bereiche variiert werden, aber eine deutliche Verbesserung wurde nicht erzielt. Wurde ein neuronales Netz mit den gleichen Synapsengewichten auf eine andere Szene angewendet, so wurden die besten Ergebnisse erzielt, wenn der Gammawert an die gelernte Szene angepaßt wurde.

### 7.4 Fusion

#### 7.4.1 Vorverarbeitung der Merkmale

Zur Fusion können drei Arten von Merkmalen unterschieden werden:

- Differentielle Merkmale, die Bildpunkte markieren an denen bestimmte Grauwertsprünge vorkommen (Linien, LOC).
- Integrative Merkmale, deren Werte über einen Bildbereich ermittelt wurden (Entropie, Varianz).
- Direkte Merkmale, die Bildpunkte markieren welche mit hoher Wahrscheinlichkeit zu einer Fahrzeugabbildung gehören (Radar, Schlagschatten, Segmentierungsergebnis des vorherigen Bildes).

Die markierten Bildpunkte der differentiellen Merkmale werden jeweils über einen  $(9 \times 9)$  großen Bildbereich gezählt. Das Ergebnis wird dem Fusionsnetzwerk übergeben. Alle anderen Merkmale werden direkt übergeben um die Informationen für den Entscheidungsprozeß zu maximieren.

Die Entropiewerte wurden, zu Beginn dieser Arbeit, in fünf Wertebereiche eingeteilt. Die Histogrammwerte über einen  $(9 \times 9)$  Punkte großen Bildbereich wurden darauf dem neuronalen Netz übergeben. Dies ermöglichte eine einheitliche Behandlung aller Bildmerkmale. Durch die grobe Unterteilung in die fünf Wertebereiche gingen aber Informationen verloren. Eine zusätzliche Unschärfe wurde durch die integrativ wirkende Histogrammbildung erzeugt. Die Entropie, als integratives Merkmal, wird bereits über einen Bildbereich von  $(20 \times 20)$  Bildpunkten errechnet.

Eine direkte Übergabe der Entropiewerte verkleinerte den Merkmalsvektor. Die Trainingszeiten des neuronalen Netzes wurden deutlich kleiner. Dies deutet auf eine bessere Repräsentation der Bildmerkmale im Merkmalsvektor hin.

#### 7.4.2 Die Rückkopplung des Segmentierungsergebnisses

Da die natürliche Umwelt stetig ist, wird sich ein Objekt nach einer kurzen Zeitspanne dt an einer nur leicht veränderten Position befinden. Um diese Eigenschaft zu nutzen, wird

das Segmentierungsergebnis eines jeden Bildes zum Zeitpunkt (t-1) dem neuronalen Netz als zusätzliches Eingangsmerkmal zum Zeitpunkt t zugeführt.

Beim Training dieses Merkmals werden, statt einer vorherigen Segmentierung, die Objektund Hintergrunddaten den Datenbanken entnommen. Um eine Überbewertung zu vermeiden, werden die Daten mit Pixelrauschen überlagert. Dabei wird einem Drittel aller Bildpunkte eine falsche Klasse (Objekt, Hintergrund) zugeordnet. Die Wahrscheinlichkeit, daß ein Bildpunkt die richtige Klasse angibt, wird damit auf maximal 66% reduziert. Höhere Werte scheinen nicht sinnvoll, weil dadurch die initiale Segmentierung neu im Bild vorkommender Objekte behindert wird.

Eine zukünftige Koppelung mit einem robusten Klassifikatormodul kann die Bereiche, die als "kein Objekt" klassifiziert wurden, vor der Rückkopplung unterdrücken.

#### 7.4.3 Variation der Netz-Topologie

Die Netztopologie des Multilayer-Perzeptrons wurde im Laufe der Arbeit mehrfach variiert. Es wurde versucht mit zwei Ausgangsneuronen das Segmentierungsergebnis zu verbessern. Jedes Ausgangsneuron repräsentierte dabei eine der Klassen "Objekt" bzw. "Hintergrund". Der höchste Wert entschied über die Klassenzuordnung des betrachteten Bildpunktes. Eine Verbesserung konnte nicht erzielt werden.

Die Anzahl der Neuronen in der verdeckten Schicht wurde variiert. Es konnte gezeigt werden, daß auch mit weniger Neuronen die Aufgabe gelöst wurde. Da die Funktion, die durch das Netz gelernt werden soll aber nicht bekannt ist, sollte die Anzahl dieser Neuronen nicht zu klein gewählt werden.

Da in [4, 2] gezeigt wird, daß ein zweischichtiges Netzwerk ausreicht um beliebige Funktionen zu approximieren, wurden keine Untersuchungen an mehrschichtigen Netzen vorgenommen.

#### 7.4.4 Aktivierungsfunktion

Innerhalb dieser Arbeit wurde als Aktivierungsfunktion aller Neuronen die Fermifunktion gewählt (Gl. 8). Sie stellt eine *Quetschfunktion* dar und eignet sich gut zur Modellierung von Sättigungseffekten [1].

Alternativ kann auch eine tanh Funktion verwendet werden:

$$\sigma(x) = \tanh(x) \tag{27}$$

Sie unterscheidet sich von der Fermifunktion nur durch eine Linear-Transformation. Der Eingang x der Fermifunktion  $\sigma(x) = \frac{1}{1+e^{-x}}$  wird hierbei durch  $\tilde{x} = x/2$  und der Ausgang durch  $\tilde{\sigma} = 2 \sigma - 1$  transformiert. Hieraus ergibt sich  $\tilde{\sigma}(\tilde{x})$  zu:

$$\tilde{\sigma}(\tilde{x}) = \frac{2}{1+e^{-2\tilde{x}}} - 1$$

$$= \frac{2}{1+e^{-2\tilde{x}}} - \frac{1+e^{-2\tilde{x}}}{1+e^{-2\tilde{x}}}$$

$$= \frac{1-e^{-2\tilde{x}}}{1+e^{-2\tilde{x}}}$$

$$= \frac{e^{\tilde{x}} - e^{-\tilde{x}}}{e^{\tilde{x}} + e^{-\tilde{x}}} \equiv \tanh(\tilde{x})$$

Ein Netz mit gleicher Topologie aber einer tanh-Aktivierungsfunktion für die Neuronen in der verdeckten Schicht ist folglich äquivalent. Es hat aber andere Werte für die Synapsengewichte und Biases [2].

Als Aktivierungsfunktion am Ausgangsneuron könnte auch eine lineare Funktion eingesetzt werden. Eine sigmoide Funktion am Ausgang, hat aber den Vorteil, daß auf Grund des Ausgangswertes direkt Aussagen darüber gemacht werden können, wie sicher ein Merkmalsvektor einer Ausgangsklasse<sup>11</sup> zugeordnet wird. Ausgangswerte im Sättigungsbereich der Funktion gelten als sicher erkannt; Werte im linearen Bereich sind mit einer Unsicherheit behaftet.

Da die Fermifunktion, zum Ausgangspunkt dieser Arbeit, auch als Aktivierung der Eingangsknoten verwendet wurde, bildete sie den möglichen Wertebereich  $G \in \{0, 1, \dots 255\}$ der Eingangsmerkmale auf einen Bereich  $G' \in \{0.5 \dots 1\}$  ab. Werte über  $G \ge 4$  unterscheiden sich nach dieser Transformation nur noch unwesentlich, da sie sich im Sättigungsbereich der Funktion befinden. Die Transformation der Eingangsmerkmale wirkt sich nachteilig auf das Segmentierungsergebnis aus und wurde deshalb deaktiviert.

#### 7.4.5 Training

Vor dem Training des Multilayer-Perzeptrons müssen Trainingsdaten erstellt werden. Sie bestehen aus den Eingangsvektoren und den dazugehörigen Ausgangssollvektoren der Trainingsmenge (vgl. Kap. 6.4). Die Ausgangsvektoren sind hier nur eindimensional und besitzen die Werte  $\mathbf{y}^{\text{Obj}} = [1]$  für Objektbildpunkte und  $\mathbf{y}^{\text{Hg}} = [0]$  für Hintergrundbildpunkte. Das Verhältnis zwischen Objekt- und Hintergrundbildpunkten in den Trainingsdaten hat einen Einfluß auf das Segmentierungsergebnis. Allgemein sollte ein Verhältnis angestrebt werden, wie es auch in der später zu lösenden Aufgabe vorkommt. Da die Extraktion von Fahrzeugen aus einer Straßenverkehrsszene möglichst sicher erfolgen muß, werden mehr Objektbildpunkte gelernt. Mit einem Verhältnis von 1 : 1, bei etwa 10000 Trainingsdatensätzen, wurden gute Ergebnisse erzielt.

Desweiteren konnte im Laufe der Arbeit die Anzahl der Lernzyklen von ca. 1600 auf 180 reduziert werden. Dies deutet auf eine bessere Repräsentation der Bildmerkmale im Merkmalsvektor hin.

#### 7.4.6 Die Datenbanken

Für die Erstellung von Trainingsdatensätzen für neuronale Netze, die zur Segmentierung oder zur Klassifizierung genutzt werden sollen, wird auf Datenbanken zugegriffen. Sie beinhalten Informationen über

- die Anzahl der Objekte im Bild,
- die Koordinaten der objektumschließenden Rechtecke,
- die Klasse der Objekte (PKW, LKW, Transporter, Fahrrad, Fußgänger oder Hintergrund),
- die Helligkeit des Bildes (sehr dunkel, dunkel, normal, hell, sehr hell),

<sup>&</sup>lt;sup>11</sup>Für die Segmentierungaufgabe gibt es nur die Klassen: Objekt und Hintergrund.

- den Fahrbahnverlauf (stark links, links, geradeaus, rechts, stark rechts),
- und die Koordinaten zweier Geraden, die die Fahrbahnbegrenzung markieren.

Da bei der Erstellung der Datenbanken nur rechteckige Bereiche ausgewählt werden können, enthalten die Objektrechtecke hin und wieder Bildpunkte, die idealerweise dem Hintergrund zugeordnet werden müßten. Bei PKW's die am Bildrand abgebildet werden, kommen diese Punkte beispielsweise über der Motorhaube vor (siehe Abb. 32). Die Bereiche unterhalb der Fahrzeuge werden mit als Objektbildpunkte markiert, weil der Schlagschatten ein wichtiges Erkennungsmerkmal für ein Fahrzeug darstellt (siehe Kap. 2.2.3).



Abbildung 32: Beispiel: Objekt und Hintergrundbereiche in einer Datenbank

Da desweiteren die Erstellung einer solchen Datenbank manuell erfolgt und sehr viel Zeit in Anspruch nimmt, liegen Objektinformationen nur für jedes 50. und Hintergrundinformationen nur für jedes 250. Bild vor. Die, im Rahmen dieser Arbeit, erstellte Objektdatenbank für die Paris-Sequenz enthält abweichend davon jeweils drei aufeinander folgende Bilder. Mit dieser Bilderfolge ist es möglich, die Rückkopplung des Segmentierungsergebnisses in der Trainingsphase des neuronalen Netzes zu simulieren (vergl. Kap. 7.4.2).

## 8 Analyse

## 8.1 Fehlertoleranz und Robustheit

Ein Ziel dieser Arbeit ist es, die Segmentierung fehlertolerant gegen den Ausfall einen Moduls und robust gegen Störungen zu machen. Eine Segmentierungssequenz von 2750 Bilder ist in Abbildung 33 verdeutlicht.



Abbildung 33: Segmentierungen im Abstand von 250 Bildern.

Duch gezieltes Abschalten der einzelnen Basismodule wird die Fehlertoleranz und Robustheit untersucht. Die Abbildung 34 zeigt die Segmentierungsergebnisse. Bis auf Bild 34h zeigen sich keine wesentlichen Verschlechterungen der Segmentierung. Allein die Abschaltung der Konturmerkmale des LOC-N<sub>8</sub> bewirkt, daß zwei PKW's nur noch teilweise extrahiert werden. Bei dieser Abschaltung werden gleich 7 von 20 Merkmalen im Merkmalsvektor zu Null gesetzt. Die rechteckige Form des mittleren, gut segmentierten Fahrzeuges zeigt den Einfluß, den jetzt die Radardaten nehmen. Vergleiche hierzu das Ergebnis des Radarmoduls in Abbildung 15.



Abbildung 34: Segmentierungsergebnisse nach dem Abschalten von Basismodulen:

- a) Grauwertbild
- c) ohne Entropie
- e) ohne Schagschattendetektion
- g) ohne Radardaten
- i) ohne Segmentierungsergebnis des vorherigen Bildes
- b) Segmentierung mit allen Modulend) ohne Varianz
- f) ohne Linienbild
- $h) \quad {\rm ohne\ Konturmerkmale\ des\ LOC-N_8} \\ j) \quad {\rm ohne\ Texturmerkmale\ des\ LOC-N_4}$

## 8.2 Sensitivitätsanalyse

In diesem Kapitel soll der Einfluß der Bildmerkmale auf die Segmentierung untersucht werden. Dazu wird eine *Sensitivitäts-Analyse* durchgeführt. *Sensitivität* wird in [1] wie folgt definiert:

Sensitivität ist eine Maßzahl, die das Ausmaß der Veränderung am Ausgang eines Output-Knotens in Abhängigkeit von der Veränderung eines spezifischen Input-Knotens zu einem bestimmten Pattern<sup>12</sup> kennzeichnet.

Um ein Maß zu erhalten, daß nicht an einen bestimmten Eingangsvektor gebunden ist, sondern eine allgemeinere Aussage zuläßt, wird das arithmetische Mittel der Sensitivitätszahlen über eine Stichprobe von 50 Eingangsvektoren ermittelt. Abbildung 35 zeigt die so ermittelten Werte in einem Balkendiagramm.



Abbildung 35: Sensitivitätsanalyse

Ein Vergleich der Merkmale untereinander ist nur zwischen gleichartigen Merkmalen möglich (vgl. Kap. 7.4.1).

Von den direkten Merkmalen hat die Radarinformationden größten, die Schlagschattendetektion einen mittleren und das Segmentierungsergebnis des vorherigen Bildes den kleinsten Einfluß auf die Segmentierung. Ein größerer Einfluß der Segmentierung des vorherigen Bildes wurde erfolgreich durch die Überlagerung mit Pixelrauschen vermieden (vgl. Kap. 7.4.2). Von den Integrativen Merkmalen beeinflußt die Varianz das Ergebnis am stärksten.

Bei den differentiellen Merkmalen fällt auf, daß der Ausgang auf Änderungen häufig in Objekten vorkommenden Merkmale (wie die horizontalen Richtungscodes des LOC), weniger sensibel reagiert als auf seltener vorkommende Merkmale.

<sup>&</sup>lt;sup>12</sup>In [1] bezeichnet "Pattern" den Eingangsvektor.

## 9 Zusammenfassung und Ausblick

#### 9.1 Zusammenfassung

In der vorliegenden Arbeit wurden zunächst verschiedene Bildanalysealgorithmen vorgestellt und ihre Eignung zur Segmentierung von Fahrzeugen vor einem bewegten Hintergrund analysiert und beurteilt. Positiv bewertete Verfahren wurden verwendet um den Merkmalsvektor problemangepaßt zu erweitern. Redundante Informationen leisten hierbei einen Beitrag zu einer robusten und fehlertoleranten Segmentierung. Die Struktur und die Wahl des Koppelungsnetzwerkes zur Fusion der Eingangsmerkmale wurde diskutiert und ein Konzept zur flexiblen Einkoppelung beliebiger Sensorik entwickelt. Am Beispiel von Radardaten wurde eine Sensor-Fusion auf der Merkmalsebene realisiert.

Es konnte gezeigt werden, daß die Separierbarkeit zwischen Objekten und Hintergrund verbessert und das Segmentierungsergebnis robuster gegen Störungen und fehlertoleranter gegen den Ausfall einzelner Basismodule wurde.

### 9.2 Ausblick

Die reale Umwelt stellt bedeutend mehr Randbedingungen an die Segmentierung, als die Auswahl der zur Verfügung stehenden Szenen. Daher ist eine Erweiterung der Eingangsmerkmale, hinsichtlich neuer unbekannter Szenen notwendig. Das erstellte Segmentierungsmodul bietet hierfür eine gute Basis.

Als zusätzliches Merkmal könnte beispielsweise, mit Hilfe des Fahrbahnverlaufs, eine Ortsabhängigkeit für Fahrzeugabbildungen integriert werden. Dies ist nur bei Aufnahmen mit einer nicht zu stark nach unten geneigten Kamera sinnvoll (vgl. Kap. 7.3.4). Der Fahrbahnverlauf und seine unmittelbare Umgebung könnten dem Koppelungsnetzwerk, wie in Abbildung 36 gezeigt, übergeben werden.



Abbildung 36: Der Fahrbahnverlauf als zusätzliches Merkmal zur Segmentierung: Die grauen Bereiche werden mit g = 1 kodiert, die weißen mit g = 0.

Wenn zusätzlich zum Fahrbahnverlauf auch die Eigenbewegung bekannt ist, kann das Flußfeld zusätzliche Merkmale liefern (vgl. Kap. 3.2).

Weitere Verbesserungen kann der Einsatz von Kameras mit einem hohen Dynamikbereich bringen. Konturen, die mit herkömmlichen CCD-Kameras in dunklen Bildbereichen durch Rauschen überlagert werden und in hellen Bereichen durch die Sättigung verloren gehen, können hiermit erfaßt werden [27].

Zukünftige Frame-Grabber<sup>13</sup> beinhalten Standardverfahren der Bildverarbeitung, wie z.B. Varianzanalyse oder Kantendetektoren, die die Berechnung beschleunigen.

<sup>&</sup>lt;sup>13</sup>Frame Grabber akquirieren Bilder aus einem Videosignal.

Da sich die Leistungsfähigkeit aller Teilmodule des Gesamtsystems noch ständig weiterentwickelt, ist es vorstellbar, daß ein modellbasierter Klassifikator sehr sicher die Bereiche, die kein Objekt enthalten, erkennt. Damit könnten diese Bereiche aus dem rückgekoppelten Segmentierungsergebnis entfernt werden (vgl. Kap. 7.4.2).

Fehlersignale von Sensoren könnten in Zukunft automatisch einen Austausch der Synapsengewichte bewirken, so daß das neuronale Netz nicht durch redundante Informationen den Sensor ersetzt muß, sondern wieder optimal an die gegebenen Informationsquellen angepaßt ist. Auch ein Online-Lernen ist denkbar.

## Literatur

- [1] SENN Software Entwicklungsumgebung für Neuronale Netze. Benutzerhandbuch, Version 2.1, Siemens Nixdorf.
- [2] Christopher M. Bishop. Neural Networks for Pattern Recognition. Oxford University Press Inc., New York, 1995.
- [3] M.E. Brauckmann, C. Goerick, J. Groß und T. Zielke. Towards all around automatic visual obstacle sensing for cars. In *Proceedings of the Intelligent Vehicles '94* Symposium, Paris, France, Seite 79–84, 1994.
- [4] Rüdiger Brause. Neuronale Netze. Teubner Verlag, Stuttgart, 2 Ausgabe, 1995.
- [5] Belur V. Dasarathy. Decision Fusion. IEEE Computer Society Press, Los Alamitos, 1994.
- [6] Brian Gloyer, Hamid K. Aghajan, Kai-Yeung Siu und Thomas Kailath. Video-based freeway monitoring system using recusive vehicle tracking. In IS&T/SPIE Symposium on Electronic Imaging: Science & Technologie – Image and Video Processing, 1995.
- [7] C. Goerick, D. Noll und M. Werner. Artificial Neural Networks in Real Time Car Detection and Tracking Applications. *Pattern Recognition Letters*, 1996.
- [8] Christian Goerick. Local orientation coding and adaptive thresholding for real time early vision. Internal Report IRINI 94-05, Institut für Neuroinformatik, Ruhr-Universität Bochum, D-44780 Bochum, Germany, Juni 1994.
- [9] Peter Haberäcker. Praxis der digitalen Bildverarbeitung und Mustererkennung. Carl Hanser Verlag, München, Wien, 1995.
- [10] Uwe Handmann und Thomas Kalinke. Fusion of texture and contour based methods for object recognition. In *ITSC'97*, Boston, 1997. IEEE. Session 35: Intelligent Vehicles: Vision(3).
- [11] Uwe Handmann, Thomas Kalinke, Christos Tzomakas, Martin Werner und Werner von Seelen. Computer vision for driver assistences systems. In *Proceedings of SPIE*, Orlando, 1998. SPIE.
- [12] Robert M. Haralick, K Shanmugan und Its'hak Dinstein. Textural Features for Image Classification. IEEE Transactions on Systems, Man and Cybernetics, 1973.

- [13] Simon Haykin. Neural Networks. Macmillan College Publishing Company, New York, 1994.
- [14] Bernd Jähne. Digitale Bildverarbeitung. Springer Verlag, Berlin, Heidelberg, New York, 3 Ausgabe, 1993.
- [15] Thomas Kalinke. Entropie als Maß des lokalen Informationsgehalts in Bildern zur Realisierung einer Aufmerksamkeitssteuerung. Internal Report IRINI 96-07, Institut für Neuroinformatik, Ruhr-Universität Bochum, D-44780 Bochum, Germany, 1996.
- [16] David H. Kil und Frances B. Shin. Pattern recognition and prediction with applications to signal characterization: signal projection, feature analysis, decision architecture. American Institut of Physics, Woodbury, New York, 1996.
- [17] Dieter Koller, Joseph Weber und Jitendra Malik. Robust multiple car tracking with occlusion reasoning. In *Third European Conference on Computer Vision, Stockholm, Sweden*, Seite 189 – 196, Berlin, Heidelberg, New York, 1994. Springer Verlag.
- [18] Axel Korn. Bildverarbeitung durch das visuelle System. Springer Verlag, Berlin, Heidelberg, New York, 1982.
- [19] Hubertus Marbach. Objekterkennung mittels Methoden der fraktalen Geometrie und nichtlinearer Dynamik. Diplomarbeit, Institut f
  ür Neuroinformatik, Ruhr-Universit
  ät Bochum, D-44780 Bochum, 1997.
- [20] R. Mech und M. Wollhorn. A noise robust method for segmentation of moving objects in video sequences. In *ICASSP'97 Proceedings*, Seite 2657 – 2660, Munich, Germany, 1997.
- [21] Helge Ritter, Thomas Martinetz und Klaus Schulten. Neuronale Netze: Eine Einführung in die Neuroinformatik selbstorganisierter Netzwerke. Addison-Wesley, Bonn, München, Reading, Massachusetts (u.a.), 2 Ausgabe, 1991.
- [22] Raúl Rojas. Theorie der neuronalen Netze. Springer Verlag, Berlin, Heidelberg, New York, 1993.
- [23] Burkhard Schiek. Einführung in die Radartechnik. Skript zur Vorlesung an der Ruhr-Universität Bochum, 1996.
- [24] Thomas Schnitger. Untersuchungen zum Entwurf neuronaler Kopplungsstrukturen zur Fusion von Bildanalyseverfahren. Diplomarbeit, Institut für Neuroinformatik, Ruhr-Universität Bochum, D-44780 Bochum, 1998.
- [25] Claude E. Shannon. A mathematical theory of communication. Bell Systems Technical Journal, 27:379–423,623–656, 1948.
- [26] Werner von Seelen, Walter Gillner, Christian Goerick, Uwe Handmann, Thomas Kalinke, David Kastrup, Iris Leefken, Gesa Lorenz, Detlev Noll, Christos Tzomakas, Bernd Völpel, Martin Werner und Carsten Winkel. Technische Bildverarbeitung dynamischer Szenen. Internal Report IRINI 97-13, Institut für Neuroinformatik, Ruhr-Universität Bochum, D-44780 Bochum, Germany, Juli 1997.
- [27] Keiichi Yamada, Tomoaki Nakano und Shin Yamamoto. Effectiveness of video camera dynamic range expansion for lane mark detection. In *ITSC*'97, Boston, 1997. IEEE.