

# Fusion von Basisalgorithmen zur Segmentierung von Straßenverkehrsszenen

Uwe Handmann, Gesa Lorenz und Werner von Seelen

Institut für Neuroinformatik, Ruhr Universität Bochum, 44780 Bochum

## 1 Einleitung

Derzeitige Projekte am Institut für Neuroinformatik in Bochum beschäftigen sich mit der Analyse von Straßenverkehrsszenen mittels Computer Vision [12]. Dies impliziert, wegen der durch die natürliche Umwelt aufgestellten Randbedingungen, hohe Anforderungen an die zu entwickelnden Algorithmen. Im speziellen wird versucht, Verkehrsteilnehmer aus Videobildern zu extrahieren und die so gewonnenen Objekthypothesen weiter zu attributieren (z.B. Objektklasse, Abstand, Geschwindigkeit, Gefahrenpotential hinsichtlich der beabsichtigten Eigentrajektorie etc.), um im Hinblick auf den Einsatz in Assistenzsystemen in Fahrzeugen eine möglichst genaue Beschreibung der Umwelt zu erreichen. Nicht nur die große Vielfalt der unterschiedlichen Umweltszenarien, sondern auch das hohe Maß an Sicherheit, das die gestellte Aufgabe erfordert, bedingen ein breitbandiges und flexibles Gesamtsystem [6]. Ein Lösungsvorschlag wird im folgenden behandelt.

## 2 Fusionsansatz

Im wesentlichen läßt sich das Gesamtsystem in Basisalgorithmen und Algorithmen zur Fusion unterteilen. Als Basis dienen Verfahren, die spezielle Teillösungen bei gegebenen Randbedingungen liefern. Die Ergebnisse der einzelnen Algorithmen sind nicht unabhängig, so daß eine Fusion der Ergebnisse eine Erhöhung der Redundanz zur Folge hat, die ein Gesamtsystem sicher und zuverlässig macht. Die hierbei einzusetzenden Methoden der Fusion sorgen sowohl für ein flexibles Zusammenspiel als auch für eine integrative Ergebnisbildung der Basiskomponenten.

Ein Vorschlag zum Aufbau eines Gesamtsystems unter Verwendung eines Fusionsansatzes ist in Abbildung 1 dargestellt. Hierbei werden auf unterschiedlichen Kopplungsstufen die Ergebnisse der Basisalgorithmen, wie z.B. die Entropie mit den Linien zur Segmentierung oder die Hausdorff-Distanz mit der Kreuzentropie zum Tracking (Objektverfolgung), fusioniert. Im letzten Verarbeitungsschritt wird mit Hilfe symbolischer Informationen das Analyseergebnis bewertet und ein Verhaltensvorschlag erzeugt. Als Kopplungsstruktur werden neuronale Netze eingesetzt, da hier die Kopplung lernbar und flexibel ist [5]. Ein online Training ist ebenso möglich wie eine Rückkopplung über der Zeit.

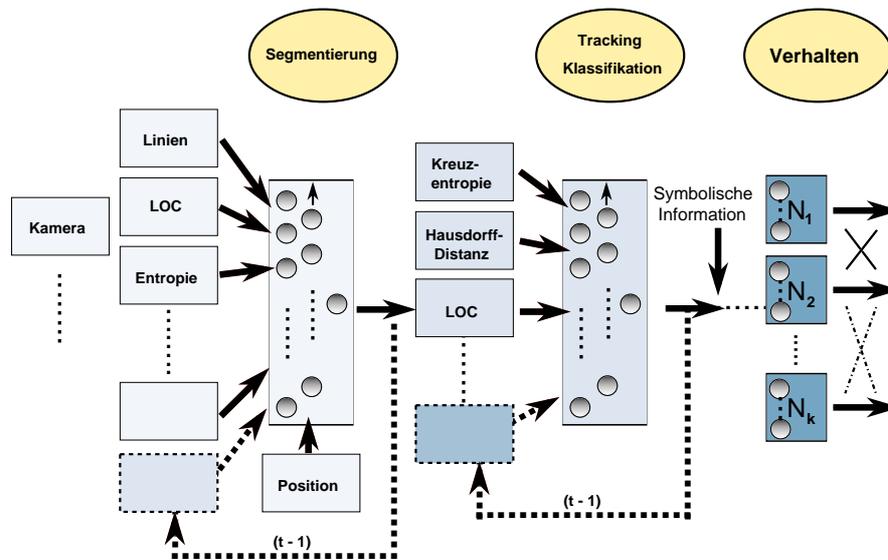


Abbildung1. Kopplungsmodell

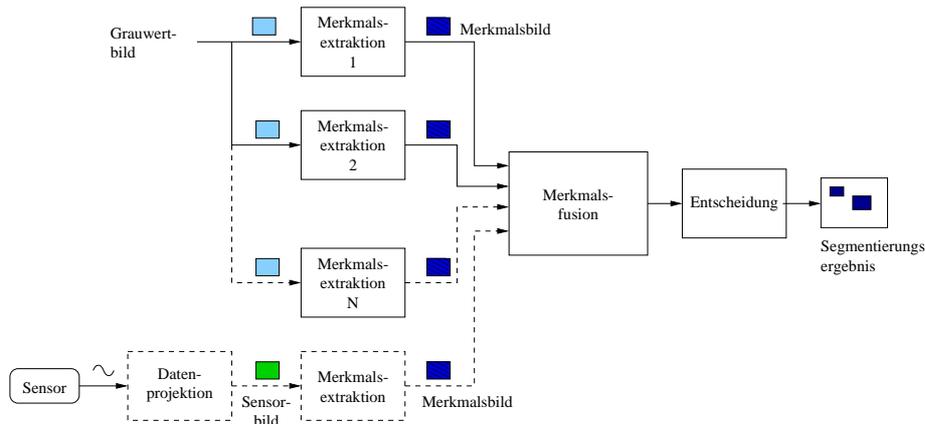
Einen wesentlichen Teilaspekt der automatischen Szenenanalyse stellt die Bildsegmentierung dar. Dabei wird unter Segmentierung eine Aufteilung des Bildes in Objekthypothesen und Hintergrund verstanden. Die generierten Objekthypothesen werden innerhalb des Gesamtsystems in weiteren Verarbeitungsschritten klassifiziert (verifiziert oder verworfen) und gegebenenfalls über der Zeit verfolgt. Die Fusion von Basisalgorithmen wird im folgenden exemplarisch am Beispiel der Bildsegmentierungsaufgabe durchgeführt.

Ein Fusionsprozeß kann auf verschiedenen hierarchischen Ebenen erfolgen ([1], Seite 6ff). Allgemein wird zwischen einer Fusion auf Daten-, Merkmals- oder Entscheidungsebene unterschieden.

Im vorliegenden Fall wird ein Fusionsprozeß auf der Merkmalsebene gewählt (Abbildung 2), um den Vorteil der Datenreduktion im Vergleich zu einer Fusion auf der Datenebene zu nutzen. Dazu werden relevante Ergebnismerkmale der Basisalgorithmen zu einem Merkmalsvektor zusammengesetzt und mit Hilfe eines neuronalen Netzes fusioniert. Eine Schwellwertbildung am Ausgang (Entscheidung) ordnet die einzelnen Pixel des Bildmaterials dem Hintergrund bzw. einem relevanten Segment zu.

### 3 Segmentierung

Am Beispiel der Segmentierung werden im folgenden Abschnitt zunächst die verwendete Kopplungsstruktur sowie die Basisalgorithmen vorgestellt und motivierende Segmentierungsergebnisse präsentiert. In einem weiteren Abschnitt



**Abbildung 2.** Prinzip einer Multisensor-Fusion auf der Merkmalsebene

wird am Beispiel der Einkopplung von Farbinformation gezeigt, daß die Kopplungsstruktur flexibel erweiterbar ist und die integrative Ergebnisbildung die Qualität der Segmentierung erhöht.

### 3.1 Segmentierung basierend auf Grauwertbildsequenzen

Im diesem Ansatz koppeln differenzierende und integrierende Basisalgorithmen (Abbildung 3) in ein neuronales Netz ein, um die Segmentierungsaufgabe zu lösen. Das Linienbild (Polygonzugapproximation der Kontur) und die Lokale Orientierungskodierung (LOC [3] [4]) finden als differenzierende Merkmale Verwendung. Die Entropie [8] wird als integrierendes Merkmal eingesetzt. Für jedes Pixel wird ein zwölfdimensionaler Eingangsvektor

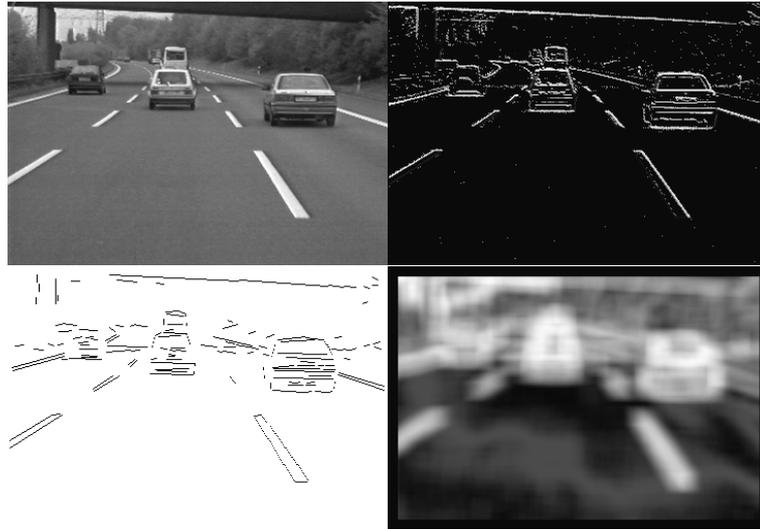
$$\mathbf{u}(x, y) = (\mathbf{u}_1(x, y)^T, x, y)^T$$

für das Kopplungsnetz erzeugt. Der 10-dimensionale Vektor  $\mathbf{u}_1(x, y)$  beinhaltet die Kontur- und Texturinformationen, die Variablen  $x$  und  $y$  repräsentieren die Pixelkoordinaten. Der Vektor  $\mathbf{u}_1(x, y)$  ist definiert durch

$$\mathbf{u}_1(x, y) = \sum_{(i,j) \in R} \mathbf{v}(i, j),$$

wobei  $R$  die lokale Nachbarschaft ( $9 \times 9$ ) des Bildpunktes  $(x, y)$  und  $\mathbf{v}(i, j)$  einen binären Vektor beschreibt. Die Vektorelemente  $v_1(x, y), \dots, v_4(x, y)$  kodieren reduzierte LOC-Merkmale,  $v_5(x, y), \dots, v_9(x, y)$  kodieren die Entropiewerte und  $v_{10}(x, y)$  wird gesetzt, wenn der Bildpunkt  $(x, y)$  Teil eines Linielements ist.

Als Kopplungsstruktur wird ein Multilayer Perzeptron (MLP, [7], Seite 138ff) eingesetzt. Es findet eine 12-5-1-Struktur der Neuronen Verwendung. Die zwölfdimensionalen Merkmalsvektoren werden über zwölf Eingangsneuronen und fünf



**Abbildung 3.** Grauwert, LOC, Linien, Entropie

versteckte Neuronen zum Ausgangsneuron vorwärts durch das Netz propagiert. Die Aktivität des Ausgangsneurons beschreibt die Zugehörigkeit der Pixel zu den initialen Segmenten. Durch eine Schwellwertbildung wird die Entscheidung getroffen.

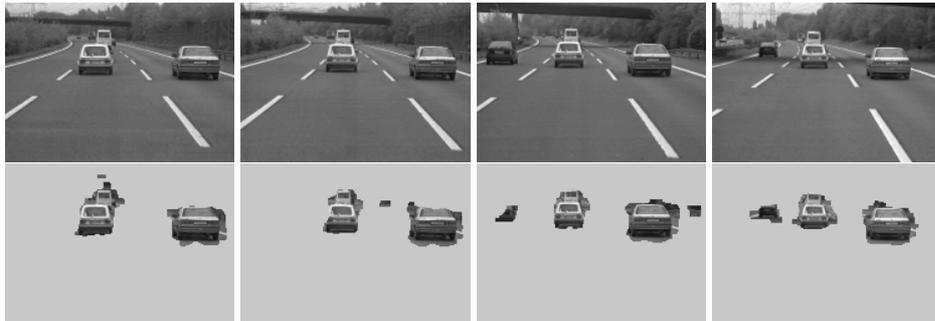
Das Training des Kopplungsnetzes findet überwacht mit Hilfe des Back Propagation Algorithmus statt ([7], Seite 142ff). Mit von Hand segmentierten Daten wird die Fehlerfunktion zwischen Ausgangsaktivität und dem Sollzustand am Ausgang ermittelt und rückwärts durch das Netz propagiert.

Diese Kopplungsstruktur ist flexibel erweiterbar und auf neue Probleme leicht adaptierbar. Abbildung 4 verdeutlicht das Segmentierungsergebnis von Fahrzeugen an einer Sequenz von 200 Bildern einer Straßenverkehrsszene, wobei jedes fünfzigste Bild dargestellt ist. Es werden alle relevanten Objekte stabil segmentiert. Partielle Fehlsegmentierungen bei Objekten mit geringem Kontrast oder fehlender Struktur sind jedoch möglich (Abbildung 4, Spalte 3).

Sollen komplexere Szenarien analysiert werden, so reicht es deshalb nicht aus, eine Analyse nur mit Standardkameras wie den herkömmlichen CCD<sup>1</sup>-Kameras durchzuführen. Einerseits ist der Dynamikbereich dieser Kameras gering und andererseits wird nur die Lichtintensität der Szenarien ausgewertet. Abhilfe kann der Einsatz höherwertiger Kameras (z.B. HDRC<sup>2</sup>, Farbkamera) und die Hinzunahme weiterer Sensoren schaffen.

<sup>1</sup> Charged Couple-Device

<sup>2</sup> High Dynamic Range Camera



**Abbildung 4.** Segmentierung mit Hilfe eines MLP-Kopplungsnetzes

### 3.2 Integration von Farbsättigung zur Segmentierung

Ein wichtiges Merkmal, welches der Mensch zur Szenenanalyse und Steuerung seiner Aufmerksamkeit verwendet, ist die Farbinformation. Sowohl in der Natur als auch in der vom Menschen geschaffenen Umwelt wird die Signalwirkung der Farbe gezielt zur Erregung von Aufmerksamkeit eingesetzt. So werden wichtige Objekte im Straßenverkehr, wie Verkehrsschilder, Bremslichter, Rettungsfahrzeuge, etc. mit Signalfarben versehen. Zur Detektion dieser für die Verkehrsteilnehmer wichtigen Objekte wird der Farbwert erfolgreich zur Segmentierung eingesetzt [10].

Bei allgemeineren Segmentierungsaufgaben, hier der Segmentierung von Fahrzeugen, wurde Farbe als Merkmal bisher selten genutzt, da Farbe im Unterschied zur Form i.a. nicht objektspezifisch ist. Fahrzeuge zeichnen sich vor allem durch eine besondere Geometrie, wie der Anordnung von horizontalen und vertikalen Linien aus, die in Linien und LOC-Merkmalen kodiert sind. Zusätzlich besitzen sie eine im Vergleich zur umgebenden Straße große Strukturiertheit, die durch die lokale Bildentropie ausgedrückt wird. Ein Teil der Fahrzeuge zeichnet sich zudem durch einen markanten Farbwert aus. Wird die Farbinformation in Verbindung mit den anderen Merkmalen verwendet, so stellt sie ein wichtiges ergänzendes Merkmal dar. Sie ist in der Lage große Objekte wie LKWs, die im allgemeinen eine geringe Struktur aufweisen, zu segmentieren.

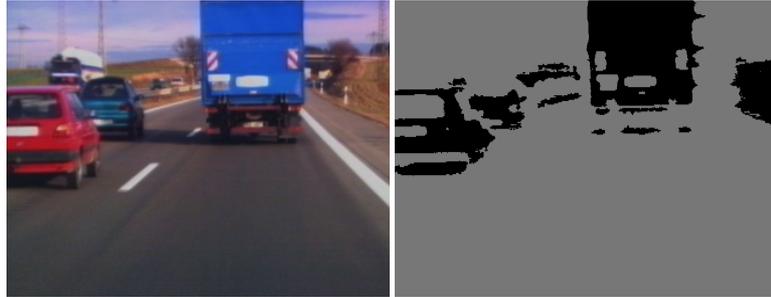
Das hier vorgeschlagene Fusionskonzept bietet die Möglichkeit, die Farbinformation neben den bereits vorhandenen differentiellen und integrativen Merkmalen in das Kopplungsnetz zu integrieren.

Farbe ist eine dreidimensionale Größe. Zu ihrer eindeutigen Beschreibung muß daher ein 3-D Koordinatensystem definiert werden. Häufig handelt es sich dabei aus Rechenzeitgründen um das hardwareorientierte RGB Koordinatensystem. Ein an der menschlichen Farbvorstellung orientierter Farbraum ist der HSV-Raum, der eine Farbe durch die Attribute Farbton, Sättigung und Helligkeit beschreibt ([2], Seite 590ff). Für die Extraktion farbiger Objekte, kann die drei-

dimensionale Farbbeschreibung auf die Verwendung der Sättigungscoordinate

$$S = \frac{\max(R, G, B) - \min(R, G, B)}{\max(R, G, B)}$$

reduziert werden.



**Abbildung 5.** Links: Originalbild, Rechts: Binarisiertes Sättigungsbild

Das Sättigungsbild  $\mathbf{S}$  wird durch eine Schwellwertoperation (Schwelle  $s_{min}$ ) binarisiert. Das binarisierte Bild  $\mathbf{W}$  wird durch

$$w(x, y) = \begin{cases} 1 & \text{falls } s(x, y) > s_{min} \\ 0 & \text{sonst} \end{cases},$$

beschrieben und als zusätzliches Merkmal

$$u_2(x, y) = \sum_{(i, j) \in R} w(i, j),$$

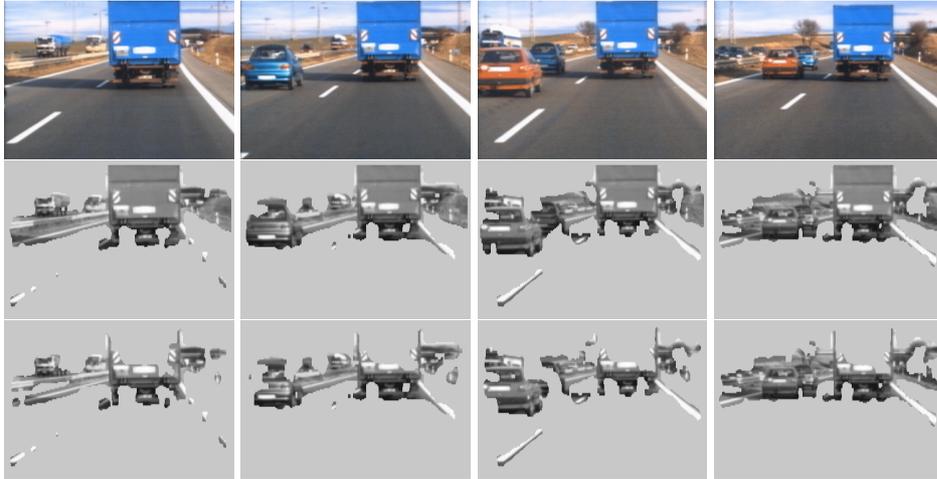
in das Kopplungsnetzwerk eingespeist, wobei die lokale Nachbarschaft  $R$  eines Bildpunktes ebenfalls Berücksichtigung findet. Es ergibt sich für jedes Pixel ein 13-dimensionaler Eingangsvektor

$$\mathbf{u}(x, y) = (\mathbf{u}_1(x, y)^T, u_2(x, y), x, y)^T$$

für das Kopplungsnetz.

Die Wahl des Schwellwertes erweist sich als unkritisch für die anschließende Weiterverarbeitung und kann auch in unterschiedlichen Bildsequenzen konstant gehalten werden. Abbildung 5 zeigt das Segmentierungsergebnis auf der Basis der Farbsättigung.

Es werden vor allem diejenigen Objektbereiche detektiert, welche durch textur- und kantenbasierte Verfahren aufgrund mangelnder Struktur häufig dem Hintergrund zugeordnet werden, wie z.B. die LKW-Rückfront. Es werden allerdings



**Abbildung 6.** Segmentierung mit Hilfe eines MLP-Kopplungsnetzes mit/ohne Bewertung der Farbsättigung

auch Hintergrundbereiche segmentiert, wie die Vegetation am Randstreifen. Dieses Problem kann jedoch bei bekanntem Fahrbahnverlauf, geeigneten zusätzlichen Basisalgorithmen oder durch Einkopplung weiterer Sensoren (z.B. RADAR [11], LIDAR- oder Infrarotsensoren [9]) gelöst werden.

Abbildung 6 zeigt eine Gegenüberstellung der Ergebnisse am Beispiel der Segmentierung von PKWs und LKWs einer Sequenz von 200 Bildern einer Straßenverkehrsszene, wobei jedes fünfzigste Bild dargestellt ist. Während bei beiden Ansätzen die Segmentierung der PKWs bei guten Kontrastverhältnissen gelingt, zeigt die Integration von Farbsättigung zur Segmentierung eine deutliche Steigerung der Ergebnisse bei schlecht ausgeleuchteten Bereichen (Abbildung 6, Spalte 3), sowie der Segmentierung von LKWs mit stark homogenen Flächen (Abbildung 6, Zeile 3).

## 4 Diskussion

Es wird ein System zur Fusion von Basisalgorithmen zur Segmentierung vorgestellt und die Funktionsfähigkeit anhand zweier Beispiele demonstriert. Durch Nutzung eines breiten Merkmalspektrums (Kanten, Textur, Farbe) wird eine verbesserte Segmentierung in natürlichen Umwelten erreicht. Die Wahl eines neuronalen Netzes erlaubt eine Adaption auf weiterführende Aufgabenstellungen (z.B. Innenstadtverkehr). Seine Flexibilität erlaubt die einfache Integrationsmöglichkeit von Farbinformation, sowie anderer Verfahren, um eine weitere Verbesserung der Segmentierung zu erreichen. Der Austausch der Netzgewichte, sowie die Möglichkeit des online Lernens ist möglich. Weitergehend ist eine

Fusion verschiedener Sensordaten denkbar (z.B. RADAR-, LIDAR- oder Infrarotsensoren), um auftretende Fehlsegmentierungen im Randbereich eliminieren zu können.

Das vorgestellte Fusionsnetz ist Teil eines Gesamtsystems. Um eine zuverlässige Analyse von Straßenverkehrsszenen zu erreichen, müssen segmentierte Bereiche klassifiziert (Hintergrund, Hindernis, Fahrzeug) und über der Zeit verfolgt werden. Diese Analyseergebnisse können einfach in das Fusionsnetz integriert werden und so das Segmentierungsergebnis weiter stabilisieren.

## Literatur

1. B. V. Dasarathy. *Decision Fusion*. IEEE Computer Society Press, Los Alamitos, 1994.
2. J.D. Foley, A. van Dam, S.K. Feiner und J.F. Hughes. *Computer Graphics: Principles and Practice*. Addison-Wesley Publishing Company, Inc., 1990.
3. C. 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.
4. C. Goerick, D. Noll und M. Werner. Artificial Neural Networks in Real Time Car Detection and Tracking Applications. *Pattern Recognition Letters*, 1995.
5. U. Handmann und T. Kalinke. Fusion of texture and contour based methods for object recognition. In *ITSC'97, IEEE Conference on Intelligent Transportation Systems 1997*, Boston, 1997. IEEE. Session 35: Intelligent Vehicles: Vision(3).
6. U. Handmann, T. Kalinke, C. Tzomakas, M. Werner und W. von Seelen. Computer Vision for Driver Assistance Systems. In *Proceedings of SPIE Vol. 3364*, Seite 136 – 147, Orlando, 1998. SPIE. Session Enhanced and Synthetic Vision 1998.
7. S. Haykin. *Neural Networks*. Macmillan College Publishing Company, New York, 1994.
8. T. Kalinke und W. von Seelen. Entropie als Maß des lokalen Informationsgehalts in Bildern zur Realisierung einer Aufmerksamkeitssteuerung. In *Mustererkennung 1996*, Seite 627–634, 1996.
9. L. C. Ramac, M. K. Uner, P. K. Varshney, M. Alford und D. Ferris. Morphological filters and wavelet based image fusion for concealed weapons detection. In *Proceedings of SPIE Vol. 3376*, Orlando, USA, 1998. SPIE. Session Sensor Fusion 1998.
10. W. Ritter, F. Stein und R. Janssen. Traffic Sign Recognition Using Colour Information. *Mathl. Comput. Modelling*, 22(4–7):149–161, 1995.
11. T. Schnitger und U. Handmann. Fusion von Bildanalyseverfahren mittels einer neuronalen Kopplungsstruktur. Internal Report IRINI 98-01, Institut für Neuroinformatik, Ruhr-Universität Bochum, D-44780 Bochum, Germany, April 1998.
12. W. von Seelen, W. Gillner, C. Goerick, U. Handmann, T. Kalinke, D. Kastrup, I. Leefken, G. Lorenz, D. Noll, C. Tzomakas, B. Völpel, M. Werner und C. Winkel. Image Processing of Dynamic Scenes. Internal Report IRINI 97-14, Institut für Neuroinformatik, Ruhr-Universität Bochum, Germany, D-44780 Bochum, Germany, Juli 1997.