Bildverstehen im innerstädtischen Verkehr

U. Franke, D. Gavrila, S. Görzig, F. Lindner, F. Paetzold, C. Wöhler
Daimler-Benz Forschung, T728
{franke, paetzold, goerzig}@dbag.stg.DaimlerBenz.com
{gavrila, lindner, woehler}@dbag.ulm.DaimlerBenz.com

Kurzfassung

Zehn Jahre nach dem Beginn von PROMETHEUS ist es beinahe selbstverständlich geworden, daß mit Kameras ausgestattete Fahrzeuge auf *Autobahnen* ohne Fahrereingriff lenken, präzise den Abstand zum Vordermann halten und ihre Geschwindigkeit an die aktuellen Vorgaben der Verkehrszeichen anpassen. Das automatische Fahren im *städtischen* Umfeld stellt eine neue Herausforderung an bildverstehende System dar. Dieser Beitrag beschreibt den Stand unserer Arbeiten, deren Ziel es ist, die Verkehrsszene soweit zu verstehen, daß unser Versuchsfahrzeug UTA (Urban Traffic Assistant) ohne Verletzung von Verkehrsregeln einem vorausfahrenden Fahrzeug selbständig folgen kann.

1. Einführung

Das Verstehen von Verkehrsszenen aus dem innerstädtischen Bereich stellt signifikant höhere Anforderungen an die Bildverarbeitung als das klar strukturierte Autobahnszenario, bei dem sich die Sehaufgabe auf das Erkennen vorausfahrender Fahrzeuge und einfacher Verkehrszeichen sowie das Vermessen streng reglementierter Spurgeometrien beschränkt.

Um die Aufgabe des automatischen Fahrens in einer solch komplexen Umwelt beherrschbar zu halten, arbeiten wir an einem *intelligenten Stop&Go-System*, das in der Bildsequenz ein vorausfahrendes Fahrzeug detektiert und dem Fahrer die Bereitschaft signalisiert, diesem zu folgen. Nach Übergabe der Kontrolle regelt das System Abstand und laterale Position zum Vordermann selbständig. Kann die Folgefahrt nicht fortgesetzt werden, weil beispielsweise der Vordermann die Spur wechselt, soll nach einer kurzen Phase des Suchens nach einem neuen Partner die Fahraufgabe an den Fahrer zurückgegeben werden.

Ziel ist es, das Fahrzeug nicht blind folgen zu lassen, sondern ihm die Fähigkeit zu vermitteln, sich entsprechend der Verkehrsregeln zu verhalten. Das beinhaltet u.a. die Beachtung von Geboten wie Stoppschilder und rote Ampeln, sowie das Anhalten vor Fußgängerüberwegen, wenn diese von Fußgängern genutzt werden.

Daraus ergeben sich für die Bildverarbeitung folgende Aufgaben, für die Lösungen erarbeitet worden sind:

- Das vorausfahrende Fahrzeug muß detektiert und seine Bewegungsparameter mit einer für eine komfortable Regelung hinreichenden Genauigkeit ermittelt werden.
- Der Spurverlauf muß analysiert werden, auch wenn dieser nicht durch klare Markierungen definiert ist.
- Kleine Verkehrszeichen und Ampeln müssen in einer relativ bunten und stark strukturierten Umgebung detektiert und erkannt werden.
- Sämtliche Verkehrsteilnehmer einschließlich Fußgängern und Fahrradfahrern müssen detektiert und klassifiziert werden.
- Stehende Hindernisse wie beispielsweise parkende Autos, welche die Fahrbahn begrenzen, müssen erkannt werden.

Im folgenden werden die realisierten Bildverarbeitungsmodule vorgestellt und das im Versuchsträger realisierte System beschrieben.

2. Stereo-Vision

Bild 1: Bildpaar des Stereokamerasystems

Modellbasierte Ansätze für die Detektion von Hindernissen, wie sie auf der Autobahn erfolgreich eingesetzt wurden (z.B. [Ulm94]), sind angesichts der Objektvielfalt im innerstädtischen Bereich wenig erfolgsversprechend. UTA verfügt deshalb über ein Stereokamerasystem mit einer Basisbreite von 30 cm, um Hindernisse zu detektieren und ihre Positionen in der Welt präzise zu bestimmen. Bild 1 zeigt ein typisches Bildpaar dieser Kameras. Die Kameras sind in Epipolargeometrie ausgerichtet und werden vom Rechner über Shutter und Gain in ihrer Helligkeit abgeglichen.

Die Korresondenzanalyse basiert auf einer realzeitfähigen lokalen Merkmalsextraktion. Für jeden Bildpunkt wird geprüft, ob sein Grauwert signifikant größer oder kleiner als der seiner direkten vier Nachbarn ist. Abhängig davon wird er einer von $3^4=81$ Klassen zugewiesen. Punkte gleicher Klassen in beiden Bildern werden unter Berücksichtigung des Epipolarconstraints einander zugeordnet, wobei geringe Disparitäten (=große Entfernungen) priorisiert werden [Fra96]. Dieses einfache Vorgehen verhindert nahe Phantomobjekte. Es zeigt sich, daß sich nicht vermeidbare Fehlzuordnungen in der Praxis nicht negativ auswirken. Zusätzlich reduziert die Berücksichtigung des „stereo ordering constraints“ störende Fehlkorrespondenzen.

Bild 2: Entfernungsbild. Die Helligkeit der Merkmalspunkte ist proportional zum Abstand

Das Verfahren liefert ein relativ dichtes Tiefenbild, wie in Bild 2 gezeigt. Eliminiert man alle Merkmale, die in der Straßenebene liegen, erhält man nach Projektion der verbleibenden

Merkmale die in Bild 3 gezeigte Tiefenkarte (Histogramm), in der potentielle erhabene Hindernisse einfach detektiert werden können. Die dazu notwendigen Parameter Kamerahöhe und Nickwinkel lassen sich on-line aus dem Tiefenbild schätzen [Fra97].

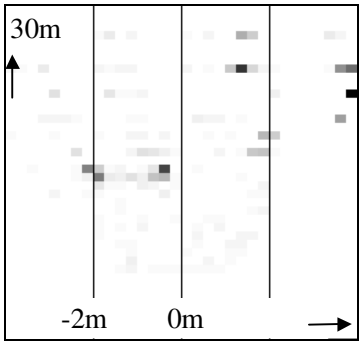


Bild 3: Tiefenkarte, die Maxima sind durch das Auto und die Bäume hervorgerufen

Nach ihrer Detektion können Hindernisse einfach durch die Sequenz verfolgt und ihrer Abmaße sowie ihre Position und Bewegungsparameter mittels Kalman-Filtern geschätzt werden. Unter normalen Sichtbedingungen werden Fahrzeuge ab einer Entfernung von 30 Metern, Fußgänger ab 20-25 Metern sicher in einem Bild der Auflösung 384x256 Bildpunkten detektiert.

Das realisierte Verfahren ist in der Lage, mehrere Objekte in Realzeit zu verfolgen. Gemessener Abstand sowie geschätzte Relativgeschwindigkeit und Relativbeschleunigung sind Eingangsgrößen für die Längs- und Querregelung.

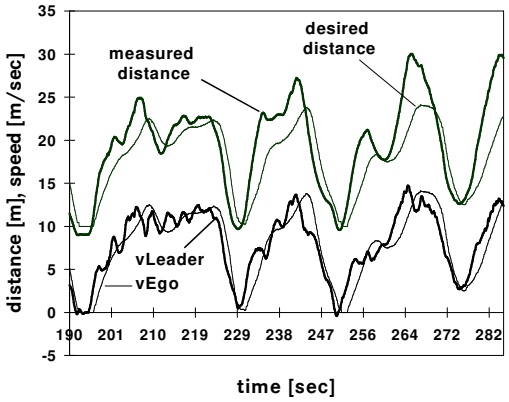


Bild 4: Folgefahrt durch die Esslinger Innenstadt.

Bild 4 zeigt die Abstands- und Geschwindigkeitsverläufe einer geregelten Fahrt durch Esslingen. Bei diesem Versuch setzte sich der Sollabstand aus einem Sicherheitsabstand von 10 Metern und einem Zeitabstand von 1 Sekunde zusammen. Die Grafik offenbart noch vorhandene Schwächen beim Beschleunigen, die auf die unterlagerte Regelung des Fahrzeugs zurückzuführen sind. Gleichzeitig zeigt sich, daß der Sicherheitsabstand problemlos reduziert werden kann.

3. Erkennung der Fahrbahnmarkierungen

Die Spurerkennung verwendet das Bild einer der beiden Stereokameras. Sie basiert auf einer schnellen Kantenextraktion und unterscheidet zwischen Markierungen und Fahrbahnbegrenzungen in Form von Bordsteinkanten etc. Dafür werden die extrahierten Konturen polygonal approximiert und eine Datenbank aufgebaut, die schnelle Zugriffe auf die gefundenen Kanten über die Schlüssel Länge, Orientierung und Ort erlaubt. Bei der Suche nach Kanten mit speziellen Eigenschaften können zusätzlich Relationen wie Parallelität und Kollinearität berücksichtigt werden (vgl. [Enk95]). Durch entsprechend kombinierte Suchanfragen werden potentielle Fahrbahnbegrenzungen detektiert sowie Stopplinien und Fußgängerüberwege gefunden.

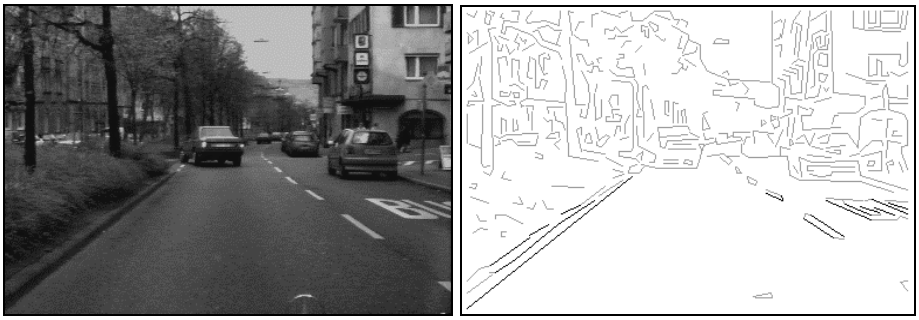


Bild 5: links: Innenstadtszene in Stuttgart, rechts: polygonal approximiertes Kantenbild. Die extrahierten Spurbegrenzungen sind dunkel hervorgehoben

Das Bild 5 zeigt eine Straßenszene in Stuttgart und die gefundenen Kandidaten für die Spurbegrenzung. In Bild 6 ist ein erkannter Zebrastreifen gekennzeichnet.

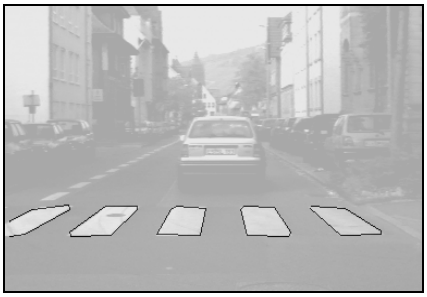


Bild 6: Erkannter Zebrastreifen

Auf einem 200 MHz PowerPC erfordern Kantenextraktion und Verdünnung eine Rechenzeit von ca. 100msec und sind damit (noch) nicht in Videoechtzeit zu leisten. Aus diesem Grund steht ein Konturverfolgungsalgorithmus zur Verfügung, der auf den bekannten modellgetriebenen Trackingprinzipien aufbaut und in wenigen Millisekunden die für die Regelung notwendigen Zustände „laterale Position“ und „Gierwinkelfehler“ liefert [Pae98].

Für die Analyse der Richtungspfeile wurde ein flächenbasiertes Verfahren entwickelt. Kandidaten werden durch eine schnelle histogrammbasierte Segmentierung mit nachgeschalteter Formanalyse der resultierenden Regionen detektiert. Für die eigentliche Erkennung wurde ein Radial-Basisfunction (RBF)-Klassifikator anhand einer Vielzahl von Beispielen trainiert. Abbildung 7 gibt ein Beispiel. Es zeigt das Originalbild, das Resultat der Segmentierung und das Ergebnis der Klassifikation. Die zwei Regionen, in die der Pfeil durch die relativ einfache Segmentierung zerfallen ist, wurden von der Formanalyse vor der Klassifikation zu einem Objektkandidaten fusioniert.

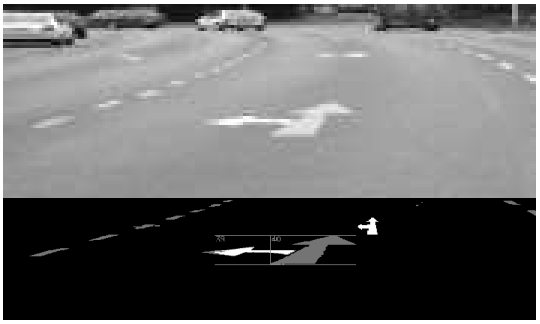


Bild 7: Oben: Straßenszene mit Bodenpfeil. Unten: Resultat der Segmentierung und Klassifizierungsergebnis.

Unsere Experimente zeigen, daß nahezu alle Pfeile detektiert und mit einer Rate besser 95% korrekt klassifiziert werden. Probleme treten lediglich bei starken Reflexionen infolge nasser Fahrbahn auf, bei denen der menschliche Fahrer ebenfalls Schwierigkeiten hat.

4. Verkehrszeichenerkennung

Bild 8: Erkannte Verkehrszeichen bei großem Blickwinkel und hoher Dynamik

Für die Erkennung der Verkehrszeichen wurde ein auf der Autobahn erprobtes farbbasiertes Modul an die Belange der Innenstadt angepaßt [Jan93]. Es besteht aus:

1. einer pixelweisen Farbklassifikation, die „interessante" Bildpunkte herausfiltert, gefolgt von
2. einer Formanalyse, die Kandidaten für Verkehrszeichen anhand von Form- und Lagerelationen liefert, und
3. einem nachgeschalteten RBF-Klassifikator, der die Kandidaten einem Verkehrszeichen zuweist bzw. sie als Fehldetektionen verwirft.

Für die Stadt mußten vor allem die Palette der bekannten Zeichen erweitert werden und die Farbklassifikation an die Bedingungen der Innenstadt angepaßt werden. Darüber hinaus ist eine Adaption an den deutlich größeren Blickwinkel vorgenommen worden. Bild 8 zeigt erkannte Verkehrszeichen im Bild einer Farbkamera.

Parallel dazu entsteht ein grauwertbasierter Ansatz, bei dem Objekte wie z.B. Verkehrszeichen anhand ihrer Form detektiert werden [Gav98]. Wie die Spurerkennung verwendet dieser Ansatz eine schnelle Gradientenbildung als Ausgangspunkt der Verarbeitung. Die im Bild zu findenden Formen werden durch binäre Muster repräsentiert, die mit dem distanztransformierten Gradientenbild korreliert werden (Chamfer-Matching). In Bild 9 ist das Kantenbild der hier betrachteten Vorlage, das dazugehörige Distanzbild und die detektierten Verkehrszeichen gezeigt. Durch die Bildung einer Musterhierarchie und einer Grob-zu-Fein-Suche wird eine erhebliche Effizienzsteigerung gegenüber Methoden erreicht, die die Mustern jeweils getrennt an allen Bildpunkten korrelieren.

Derzeit liegt die Detektionsrate bei 90% bei 4-6% „false positives". Mehr als 95% dieser zuviel detektierten Verkehrszeichenkandidaten werden von dem nachgeschalteten RBF-Klassifikator zurückgewiesen.

Bild 9: links: Konturbild und das zugehörige Distanzbild (mitte). Je heller ein Bildpunkt, desto weiter ist er von der nächsten Kontur entfernt. Rechts: untersuchte Positionen und gefundene Dreiecke und Kreise.

5. Ampelerkennung

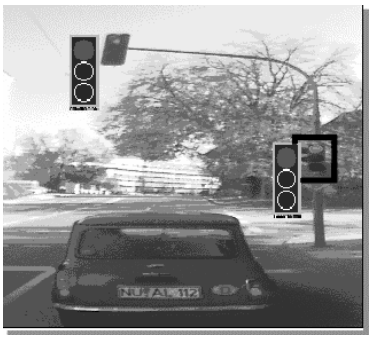
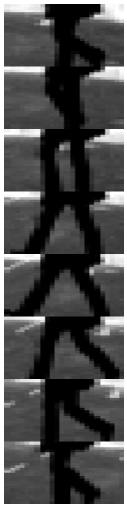


Bild. 10: Im Farbbild erkannte Ampeln

Die Farbkamera wird ebenfalls für die Erkennung der Ampeln eingesetzt. Im Bild gefundene Farbblobs werden detektiert und die umgebenden Bereiche entsprechend der Größe der Blobs ausgeschnitten. Ein nachfolgendes neuronales Netz klassifiziert den Grauwertauszug des jeweils übergebenen Bildausschnitts und unterscheidet damit Ampeln von den ebenfalls detektierten Bremslichtern eines vorausfahrenden Fahrzeugs (s. Bild 10).

6. Erkennung bewegter Fußgänger



Großes Gewicht muß im innerstädtischen Umfeld auf die sichere Erkennung von Fußgängern gelegt werden. Ein spezielles Time-Delay Neural Network mit raum-zeitlichen rezeptiven Feldern [Hei98] [Woe98] ist in der Lage, querende Fußgänger anhand der typischen Beinbewegung zu erkennen. Diesem Modul werden Objekte der Stereoanalyse zugeführt, die aufgrund ihrer Form und ihrer Bewegung potentiell Fußgänger sein können. Dazu werden, wie nebenstehend gezeigt, die Bildbereiche bis zu einer Höhe von 1m ausgeschnitten und auf eine Größe von 24x24 Bildpunkten normiert. Derzeit werden dem neuronalen Netz jeweils die letzten 8 Frames zugeführt. Pro Zeitschritt wird ein Fußgängermuster mit einer Wahrscheinlichkeit von ca. 86% bei einer False-Positive-Rate von etwa 1.5% erkannt. Das bedeutet, daß ein Fußgänger in der Sequenz stabil erkannt wird (Drop-out im Mittel nur etwa alle 7 Frames), sofern er sich auch nur einigermaßen vom Hintergrund abhebt. Dank einer zusätzlichen Rückkopplung erhält man noch stabilere Ergebnisse, auch wenn einzelne Bilder z.B. durch den Scheibenwischer gestört sind.

Als Ergänzung dazu wird an einem Verfahren gearbeitet, das in der Lage ist, Fußgänger im Einzelbild zu erkennen. Dies ist notwendig, um bei Annäherung an einen Fußgängerüberweg gezielt nach Passanten auf dem Gehweg suchen zu können, auch wenn diese sich nicht im Tiefenbild abheben, weil sie z.B. vor einem Baum stehen. Der Ansatz führt zunächst eine Hauptachsentransformation durch und reduziert damit die Vielfalt von Fußgängererscheinungsformen auf eine geringe Zahl von Eigenvektoren (Eigenbildern). Die Detektionsstufe ermittelt anschließend im resultierenden Merkmalsraum den Abstand eines unbekannten Objektes zu zuvor trainierten Beispielen.

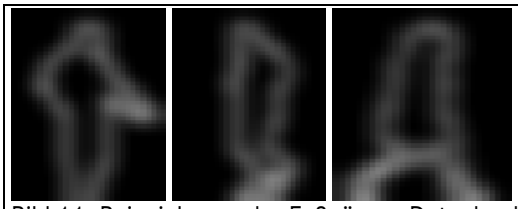


Bild 11: Beispiele aus der Fußgänger-Datenbank

Ausgangspunkt des Ansatzes ist das Gradienten-Betragsbild des zu untersuchenden Bildbereiches. Bild 11 zeigt Beispiele aus unserer Fußgängerdatenbank, anhand deren die Hauptachsentransfor-

mation generiert wurde. Einige Eigenvektoren sind in Abbildung 12 der Größe der Eigenwerte nach sortiert wiedergegeben.

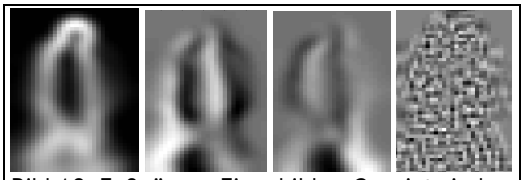


Bild 12: Fußgänger-Eigenbilder. Gezeigt sind die Eigenvektoren 0 (Mittelwert), 1, 2 und 25.

7. Integration

Die skizzierten Module sind einzeln entwickelt, getestet und auf einem Parallelrechnersystem bestehend aus derzeit 3 PowerPCs MPC 604e zu einem Gesamtsystem integriert worden. Dabei teilen sich die farbbasierten Methoden Ampel- und Verkehrszeichenerkennung, die Spurerkennung und Pfeilerkennung sowie die Stereo-Objektdetektion- und Verfolgung inklusive der Fußgängererkennung jeweils einen Prozessor.

Eine intelligente Ablaufsteuerung hat die Aufgabe, die einzelnen Tasks situationsabhängig zu aktivieren und zu parametrisieren. Beispielsweise wird die Verkehrszeichenerkennung deaktiviert, wenn das Fahrzeug vor einer roten Ampel steht oder die Pfeilerkennung ausgeschaltet, falls der Abstand zum Vordermann zu klein wird. Damit besteht die Möglichkeit, die verfügbare Rechenleistung im Sinne einer Aufmerksamkeitssteuerung auf aktuell wichtige Objekte zu konzentrieren.

Um der Steuerung des Sehsystems möglichst viele Freiheiten zu geben, streben wir an, für eine Aufgabe mehrere verschieden aufwendige Lösungen zur Verfügung zu haben. Beispielsweise stellt die Spurverfolgung, wie erwähnt, ein sehr effizientes Trackingverfahren als Alternative zur Vollbildanalyse zur Verfügung. Die Stereo-Objektverfolgung kann zu jedem Zeitpunkt auf ein korrelationsbasiertes Verfahren umgeschaltet werden, das Abstand und Relativgeschwindigkeit in 4-8 msec liefert. Weiterhin kann die knappe Resource Rechenleistung geschont werden, indem die Bildbereiche, auf denen einzelne Methoden wie beispielsweise die Ampelerkennung aktiviert werden, anhand der aktuellen Situation (Spurverlauf, Position des vorausfahrenden Fahrzeugs etc.) eingeschränkt werden.

Dazu wird eine Karte der lokalen Umgebung aufgebaut, die mit dem fahrenden Fahrzeug mitgeführt wird. Ausgehend von der Bewertung der aktuellen Situation werden unter Berücksichtigung der verfügbaren Resourcen Tasks auf den entsprechenden Prozessoren angestoßen.

Die intelligente Ablaufsteuerung wird durch ein Multi-Agentensystem durchgeführt. In einem Multi-Agentensystem arbeiten mehrere Problemlöser (Agenten) zusammen, um ein komplexes Gesamtproblem zu lösen. Für den Stop&Go Betrieb steuert das „Agent NeTwork System“ (ANTS) die beschriebenen Bildverarbeitungsprozesse sowie Fahrzeugsteuerungs- und Fahrerschnittstellenprozesse. Die flexible Software-Architektur von ANTS erlaubt eine einfache Erweiterung des Systems. Es können sowohl weitere Agenten hinzugefügt als auch neue Aufgaben entworfen werden, wie z.B. ein videobasierter intelligenter Tempomat oder eine optische Spurhaltung auf Autobahnen. Dabei können sich die neuen Aufgaben in einer Art Baukastensystem der vorhandenen Agenten zur Lösung des neuen Problems bedienen (siehe [Goe98]).

8. Zusammenfassung

Nachdem unsere Versuchsfahrzeuge VITA und OSCAR sich sicher auf Autobahnen autonom bewegen können, hat UTA die ersten Schritte in der Innenstadt erfolgreich absolviert. Möglich wurde diese Entwicklung vor allem durch den Einsatz der Stereobildverarbeitung zur direkten Gewinnung von 3D-Information und schneller Objekterkennungsverfahren.

Stereosehen im Fahrzeug ist realzeitfähig und gestattet eine komfortable Abstands- und Positionsregelung. Obwohl das Stereoverfahren in UTA für den Innenstadtbereich entworfen wurde, wird es auch auf Autobahnen im dichteren Verkehr als Abstandsregeltempomat eingesetzt.

Für viele Erkennungsaufgaben kristallisiert sich mehr und mehr eine einheitliche Vorgehensweise heraus: ein Detektionsverfahren, das auf einfachen Merkmalen wie Tiefe, Farbe/Helligkeit oder Form beruht, liefert Objektkandidaten, die nach einer Größennormierung von einem Klassifikator ohne weitere Vorverarbeitung bewertet werden. Im Rahmen der beschriebenen Arbeiten kommen RBFs und Neuronale Netze zum Einsatz. Als sinnvoll erweist sich in vielen Fällen eine vorgeschaltete Hauptachsentransformation, mit Hilfe derer die Dimension des Merkmalsvektors signifikant reduziert werden kann.

Literatur

[Enk95] W.Enkelmann, G.Struck, J.Geisler: „ROMA – A System for Model-Based Analysis of Road Markings“, Intelligent Vehicles ’95, S.356-360

[Fra96] U.Franke, I.Kutzbach: “Fast Stereo based Object Detection for Stop&Go Traffic“, Intelligent Vehicles ‘96, Tokyo, S.339-344

[Fra97] U.Franke, S.Görzig, F.Lindner, D.Mehren, F.Paetzold: „Steps towards an intelligent Vision System for Driver Assistance in Urban Traffic“, Intelligent Vehicles ’97, 9.-12.Nov 1997, Boston, S.601-606

[Gav98] D. Gavrila: „Multi-feature Hierarchical Template Matching Using Distance Transforms'', International Conference on Pattern Recognition, S.439-444, August 1998, Brisbane.

[Goe98] S.Görzig, U.Franke: „ANTS – Intelligent Vision in Urban Traffic”, in IEEE Conference on Intelligent Transportation Systems, October 1998, Stuttgart

[Hei98] B. Heisele, C. Wöhler, „Motion-based Recognition of Pedestrians“, International Conference on Pattern Recognition, S. 1325-1330, August 1998, Brisbane.

[Jan93] R.Janssen, W.Ritter, F.Stein and S.Ott: „Hybrid Approach for Traffic Sign Recognition'', Proc. of Intelligent Vehicles Conference, 1993.

[Pae98] F.Paetzold, U.Franke; “Road Recognition in Urban Environment”, IEEE Conference on Intelligent Transportation Systems, October 1998, Stuttgart

[Ulm94] B.Ulmer: „VITA II - Active collision avoidance in real traffic“, Intelligent Vehicles ‘94, Paris, 24.-26. Oct.. 1994, S.1-6

[Woe98] C. Wöhler, J. K. Anlauf, T. Pörtner, U. Franke, „A Time Delay Neural Network Algorithm for Real-Time Pedestrian Recognition“, Intelligent Vehicles ’98