



*Institut für  
Neuroinformatik*

*Ruhr-Universität  
Bochum*

Internal Report 98-04

## Videobasierte Hinderniserkennung

*by*

Uwe Handmann, Thomas Kalinke und Christos Tzomakas

Ruhr-Universität Bochum  
Institut für Neuroinformatik  
44780 Bochum



IR-INI 98-04  
2. Auflage September 1998  
ISSN 0943-2752

# Videobasierte Hinderniserkennung

Uwe Handmann, Thomas Kalinke und Christos Tzomakas\*  
Institut für Neuroinformatik, Ruhr-Universität Bochum, FRG

## Zusammenfassung

In diesem Artikel wird ein System vorgestellt, welches eine videobasierte Hinderniserkennung zur automatisierten Bildanalyse von Straßenverkehrsszenen durchführt. Eine Unterteilung der Hinderniserkennung in Objektdetektion, Objektverfolgung und Objektklassifikation läßt eine Extraktion und eine Attributierung von Verkehrsteilnehmern zu. Eine Szeneninterpretation ist ableitbar.

## 1 Einführung

Systeme zur automatisierten Bildanalyse, insbesondere zur autonomen oder teilautonomen Führung von Fahrzeugen, lassen sich in den meisten Fällen in drei Subsysteme unterteilen: die Objektdetektion, die Objektverfolgung und die Objektklassifikation. Das Ziel ist es, Verkehrsteilnehmer zu extrahieren und attributieren, um eine Szeneninterpretation und die daraus resultierende Verhaltensgenerierung zu ermöglichen ([10], [2]) und [3].

In diesem Artikel wird ein System vorgestellt, das in der Lage ist, Objekthypothesen aus Bildsequenzen zu generieren, die Objekte über der Zeit zu verfolgen und zu klassifizieren.

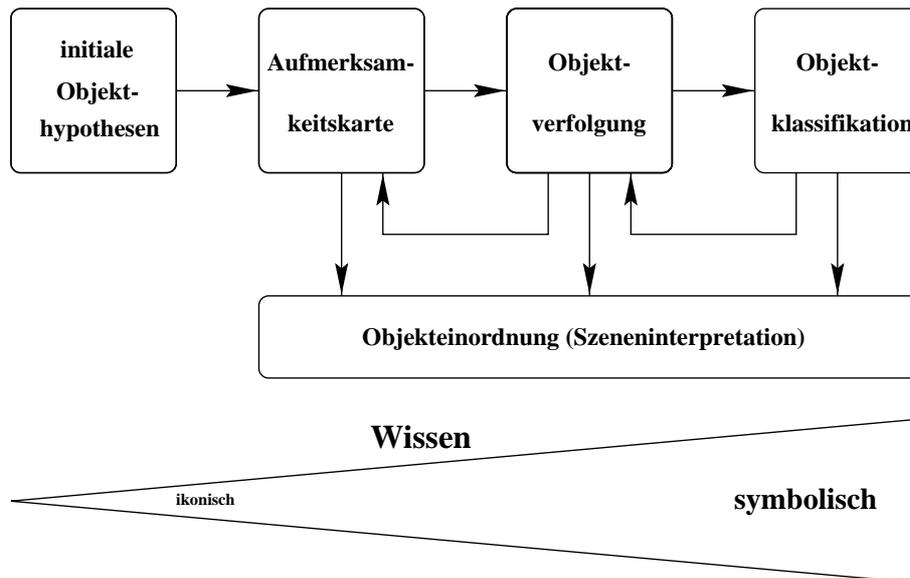


Abbildung 1: Hinderniserkennungs-System

Die gewonnenen Ergebnisse resultieren dann in einer Szeneninterpretation. In Abbildung 1 ist die Funktionalität schematisch dargestellt. Initial detektierte Objekthypothesen werden zur Verifikation und Stabilisierung in eine Repräsentation eingekoppelt. Hierbei wird eine Robustheitserhöhung

\*institut@neuroinformatik.ruhr-uni-bochum.de

durch Erzeugung von Redundanz in Form einer Rückkopplung der Ergebnisse der Objektverfolgung geschaffen. Die aus der Repräsentation resultierenden Objekte werden schließlich über der Zeit verfolgt und klassifiziert. Die gewonnenen Objektattribute wie relative Position (Abstand, time-to-collision), Geschwindigkeit und Objektklasse stützen eine Szeneninterpretation (siehe Abbildung 4). Nachfolgend werden die einzelnen Verfahren grob skizziert.

## 2 Objekthypothesen

Als erster Schritt in der Bildverarbeitungskette ist die initiale Detektion potentieller Hindernisse auf der Fahrbahn zu leisten. Hierbei stehen bereits etablierte Verfahren am Institut für Neuroinformatik zur Verfügung [1], [6] und [8]. In der hier vorgestellten Anwendung beschränkt man sich auf Grund der zur Verfügung stehenden Rechenzeiten auf die Detektion des einfachen, aber dennoch signifikanten Merkmals des Schattenwurfs der Fahrzeuge. Es hat sich gezeigt, daß dieses Merkmal stabil und robust ist [9] und [7]. Da dieser ikonisch orientierte Verarbeitungsschritt am Beginn der Verarbeitungskette steht und mit wenig Daten- bzw. Modellwissen operiert, darf die Reduktion der gesamten Bildinformation auf Aufmerksamkeitsbereiche nur eingeschränkt vollzogen werden. Vielmehr ist eine weitere Fokussierung der Szeneninformation durch die nachfolgend operierenden Module zu leisten.

## 3 Repräsentation

Zur Bewertung einzelner Bildbereiche hinsichtlich der Signifikanz für die Aufgabenstellung wird eine pixelorientierte Bildrepräsentation eingeführt. In dieser Repräsentation, welche als Aufmerksamkeitskarte organisiert ist, werden einkoppelnde Bildbereiche über lokale Aktivitäten durch aufgabenbezogene geeignete Funktionale bewertet und deren Signifikanz verifiziert (Abbildung 2).

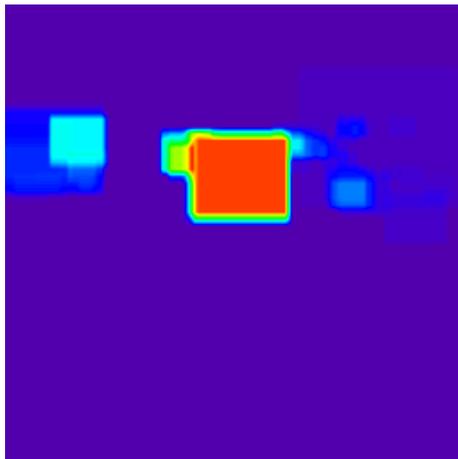


Abbildung 2: Repräsentation am Beispiel der Objektverfolgung

Eine Integration von differierenden Modulen in das Gesamtsystem wird durchgeführt. Module zur Fusion [4], [5] können auf dieser Verarbeitungsstufe ebenfalls integriert werden. Um eine zeitliche Stabilisierung und Robustheit zu erreichen, führt die Repräsentation die Zeit als dritte Achse mit. Die Repräsentation stellt nachfolgenden Verarbeitungsschritten eine Liste von Bildbereichen bezüglich ihrer aufgabenbezogenen Hindernissignifikanz zur Verfügung. Eine Rückkopplung in die Objektdetektion ist prinzipiell denkbar. Eine Einkopplung in Objektverfolgung und -klassifikation ist realisiert.

## 4 Objektverfolgung

Bei der Objektverfolgung ist der geometrische Vergleich von Formen die zu lösende Hauptaufgabe. In dem gewählten Ansatz [2] wird die Abweichung einer Menge von Modellmerkmalen, in der Regel die Objektkonturen, zu den Bildmerkmalen mit Hilfe der *Hausdorff Distanz* [10] gemessen. Hierbei wird zum Zeitpunkt  $t = 0$  ein Modell des Objektes akquiriert und über der Zeit unter den Freiheitsgraden Translation und Skalierung in der Bildsequenz wiedergefunden. Um ein möglichst hohes Maß an Robustheit und Flexibilität zu gewährleisten, wird das Modell an die gefundenen Bildmerkmale zeitlich adaptiert. Hierbei ist ein Verfahren entwickelt worden, das Modell geeignet anzupassen, so daß nicht etwa der Hintergrund, sondern die Objektveränderungen selbst gelernt werden. Der Algorithmus ist erfolgreich auf einer Vielzahl von Bildsequenzen unterschiedlicher Geometrie getestet worden (Abbildung 3).

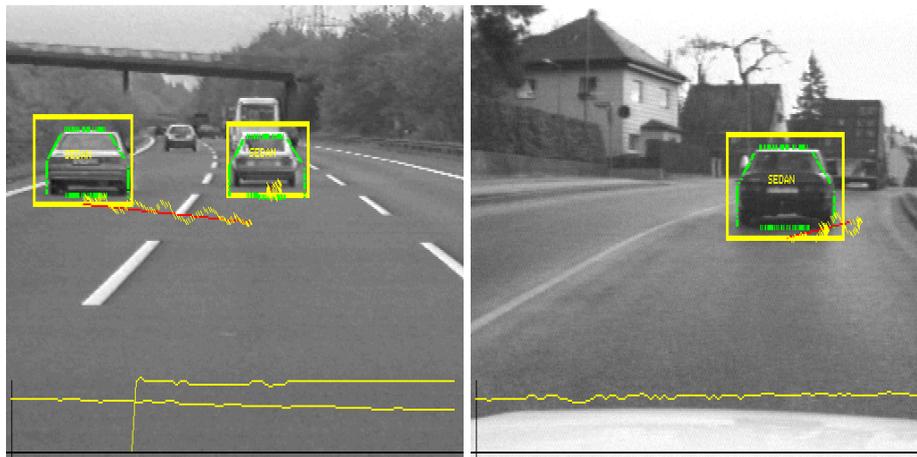


Abbildung 3: Objektdetektion, -klassifikation und Objektverfolgung auf Autobahnen und Landstraßen

## 5 Objektklassifikation

Um eine Szenendeutung durchzuführen, ist es notwendig, extrahierten Bildbereichen eine Objektklasse zuzuordnen. Am Institut für Neuroinformatik werden zur Lösung dieser Aufgabe hauptsächlich neuronale Klassifikatoren eingesetzt [7], [10]. Grundsätzlich lassen sich die Klassifikatoren in merkmalsbasierte und modellbasierte unterteilen. Der schnelle merkmalsbasierte neuronale Klassifikator führt mit Hilfe der lokalen Orientierungskodierung [10] eine Initialklassifikation von Fahrzeugen durch. Histogrammbasierte Eingangsdaten werden einem neuronalen MLP-Netz (Multilayer Perzeptron, universeller Approximator kontinuierlicher Funktionen) übergeben. Der Netzausgang beschreibt die Klasse der übergebenen Regionen. Im konkreten Fall wurden vier Klassen (Pkw, Lkw, Transporter und Hintergrundstruktur) gelernt. Der modellbasierte neuronale Klassifikator ordnet Bildregionen mit Hilfe der *Hausdorff Distanz* konkrete Modelle zu (in diesem Fall: Pkw und Lkw). Ein neuronales Netz bewertet hierbei die Ähnlichkeit zwischen den Modellen und der gegebenen Hypothese. Beide Typen der Klassifikatoren sind auf einer Datenbank mit 1500 verschiedenen Bildern unterschiedlicher Kameratypen und -geometrien trainiert worden.

## 6 Diskussion

Im vorliegenden Artikel ist ein System beschrieben worden, welches stabil auf unterschiedlichen Szenen (Autobahn, Landstraße) eine robuste Objektdetektion, Objektklassifikation und Objektverfolgung durchführt. In den Abbildungen 3 und 4 sind Ergebnisbilder des vorgestellten Systems



Abbildung 4: Objektdetektion, Objektklassifikation und Objektverfolgung an einer Sequenz von 1000 Bildern mit unterschiedlichen Beleuchtungsverhältnissen

bei unterschiedlichen Verkehrssituationen und unter diversen Beleuchtungsverhältnissen (Brückendurchfahrt, zeitweiliges Gegenlicht) dargestellt.

Bildbereiche, die über die Objektdetektion und die Repräsentation extrahiert wurden, werden dem Modul der Objektverfolgung übergeben und daraufhin verfolgt (gelber Rahmen). Die Klassifikatoren ordnen den Regionen Objektklassen zu (grünes Pkw-Model). Die Historie beschreiben die im unteren Bildabschnitt eingezeichneten Graphen der Skalierung und die im Bild mitgeführten Koordinaten der einzelnen Bildbereiche (gelb: Position, rot: Bewegungsrichtung über der Zeit). Eine stabile Objektverfolgung ist über mehrere 1000 Bilder durchgeführt worden.

Der Einsatz einer geeigneten Repräsentation gewährleistet das flexible Zusammenspiel der einzelnen Module. Eine Szeneninterpretation kann durch den Einsatz aufgabenspezifischer Repräsentationen, beispielsweise für die Verfolgung von Fahrzeugen oder die Beachtung des Gegenverkehrs, gestützt werden.

Eine mit dem Videotakt schritthaltende Implementierung in Kooperation mit einem Partner aus der Automobilindustrie wird derzeit realisiert.

## Literatur

- [1] 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).
- [2] U. Handmann, T. Kalinke, C. Tzomakas, M. Werner und W. v. Seelen. Computer vision for driver assistance systems. In *Proceedings of SPIE Vol. 3364*, Orlando, 1998. SPIE. Session Enhanced and Synthetic Vision 1998.
- [3] U. Handmann, T. Kalinke, C. Tzomakas, M. Werner und W. v. Seelen. An image processing system for driver assistance. In *IV'98, IEEE International Conference on Intelligent Vehicles 1998*, Stuttgart, 1998. IEEE.
- [4] U. Handmann, G. Lorenz, T. Schnitger und W. v. Seelen. Fusion of different sensors and algorithms for segmentation. In *IV'98, IEEE International Conference on Intelligent Vehicles 1998*, Stuttgart, 1998. IEEE.
- [5] U. Handmann, G. Lorenz und W. von Seelen. Fusion von Basisalgorithmen zur Segmentierung von Straßenverkehrsszenen. In *Mustererkennung 1998*, Heidelberg, 1998. Springer-Verlag.
- [6] T. Kalinke und C. Tzomakas. Objekthypothesen in Verkehrsszenen unter Nutzung der Kamerageometrie. Internal Report IRINI 97-07, Institut für Neuroinformatik, Ruhr-Universität Bochum, D-44780 Bochum, Germany, 1997.
- [7] T. Kalinke, C. Tzomakas, M. Werner und W. v. Seelen. A Texture-based Object Detection and an adaptive Model-based Classification. In *IV'98, IEEE International Conference on Intelligent Vehicles 1998*, Stuttgart, 1998. IEEE.
- [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] F. Thomanek, E.D. Dickmanns und D. Dickmanns. Multiple Object Recognition and Scene Interpretation for Autonomous Road Vehicle Guidance. In *Proceedings of the Intelligent Vehicles '94 Symposium, Paris, France*, Seite 231–236, 1994.
- [10] 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. Technische Bildverarbeitung dynamischer Szenen. Internal Report IRINI 97-13, Institut für Neuroinformatik, Ruhr-Universität Bochum, D-44780 Bochum, Germany, Juli 1997.