

## Ansichtenbasierte Erkennung von Fahrzeugen

E. Michaelsen und U. Stilla

FGAN-FOM Forschungsinstitut für Optronik und Mustererkennung  
Gutleuthausstr. 1, 76275 Ettlingen  
{mich,usti}@fom.fgan.de

**Zusammenfassung.** Zur ansichtenbasierten Detektion von Fahrzeugen auf Einzelbildern wird ein modellbasiertes, strukturelles Erkennungsverfahren eingesetzt. Fahrzeuge werden in einer Bestandteilhierarchie durch artikulierte 3D-Polyeder modelliert. Aus den 3D-Modellen wird durch Variation der Kameraparameter und hidden-line Projektion eine umfangreiche Datenbank von 2D-Ansichten generiert. Zur Erkennung werden aus den Bildern L- und T-Strukturen extrahiert und mit den Strukturen in der Ansichtsdatenbank verglichen. Analog zur generalisierten Hough Transformation dient ein Akkumulator zur Evidenzanhäufung. Als Bildbeispiel dient eine Schrägsicht auf ein Fahrzeug mit Anhänger in natürlicher Umgebung.

### 1 Einleitung

Für die 3D-Erkennung von *man-made objects* werden häufig modellbasierte Verfahren vorgeschlagen. Zur Modellierung sind verschiedene Formen der Wissensrepräsentation bekannt [12], wie z. B. Produktionen, Frames oder semantische Netze. In der vorliegenden Arbeit werden zur Fahrzeugerkennung Produktionensysteme [14] verwendet.

Im Gegensatz zu Objekten mit großer Formvariation, wie z. B. Gebäuden, können bei der Modellierung von Fahrzeugen formfeste 3D-Modelle verwendet werden, da die spezielle Geometrie eines Fahrzeugtyps als bekannt vorausgesetzt werden kann. Die besondere Schwierigkeit bei der Erkennung ist durch die unbekannte Ansicht aufgrund der freien Position und Orientierung gegeben.

Während eine Gruppe von Ansätzen zur Erkennung von Fahrzeugen die Verfolgung in Bildfolgen als Zielsetzung hat, wie z. B. in [11], versuchen andere Ansätze eine Detektion des Fahrzeugs im Einzelbild. Hermiston et al. [4] untersuchen zur Detektion Ähnlichkeiten zwischen den Grauwertverläufen im Bild und im gerenderten Modell ('mutual information'). Dagegen stehen konturbasierte Ansätze [2],[15]. Binfort & Levitt [2] verwenden zur Modellierung generalisierte Zylinder und werten die Evidenz mit Baeysschen Netzen aus. Wang [15] verwendet artikulierte Polyedermodelle und Linearkombination charakteristischer Ansichten. Die Verwendung charakteristischer Ansichten erfordert jedoch komplizierte Fallunterscheidungen. Im vorliegenden Beitrag werden für komplexere Fahrzeuge ebenfalls artikulierte Polyedermodelle verwendet, jedoch wird statt der Auswahl charakteristischer Ansichten der Ansichtenraum gleichmässig diskretisiert.

In vorhergehenden Arbeiten wurden Fahrzeuge durch Stereoanalyse detek-

tiert [10]. Dabei konnten die 3D-Modellteile direkt im 3D-Raum verglichen werden. Die Folge von Kamerapositionen hat eine hinreichend stabile Triangulation zugelassen. Das ist nicht möglich für Einzelansichten oder wenn sich die Kamera direkt auf die beobachtete Szene zubewegt. Für Anwendungen, bei denen das der Fall ist, schlagen wir daher eine *ansichtenbasierte* Methode vor.

Abschnitt 2 stellt die hier gewählte ansichtenbasierte Modellierung vor. Die Implementierung des entsprechenden Produktionsnetzes wird in Abschnitt 3 erklärt. Abschnitt 4 stellt damit erzielte vorläufige Ergebnisse vor, bevor mit Abschnitt 5 eine kurze Diskussion den Beitrag abschließt.

## 2 Ansichtenbasierte Objekterkennung

Bei der ansichtenbasierten Objekterkennung findet der Vergleich zwischen Modell und Daten im 2D-Bildraum statt. Dazu müssen 2D-Ansichten der 3D-Modellteile als Modelle verwendet werden. Die Nutzung von Bestandteilhierarchien ist weiter möglich. Zur Konsistenzprüfung, ob die Teile korrekt zu einander im Raum liegen, ist dabei aber jeweils eine Rückprojektion erforderlich.

Eine Menge von 2D-Linien, die durch eine perspektivische hidden-line Projektion eines Polyeders entstanden ist, heißt im folgenden Ansicht des Polyeders. Unter einem Aspekt versteht man hingegen einen Liniengraphen. Änderungen in der Ansicht, die die Topologie nicht ändern, liefern keinen neuen Aspekt [15].

### 2.1 Der Raum der Ansichten

Die Anzahl der zu verwendenden Ansichten pro Modellteil hängt von den Einschränkungen ab, die für Translation und Rotation der Objekte gemacht werden können. In der Regel kann man z. B. bei Fahrzeugdetektion die Entfernung auf ein Intervall und die räumliche Rotation des Modells auf den Azimut beschränken. Das heißt, man geht davon aus, dass das Fahrzeug eben steht und nicht geneigt ist. Translationen des Modells können je nach Brennweite am Bildrand zu geometrischen Verzerrungen führen. Da bei den hier verwendeten längeren Brennweiten dieser Effekt vernachlässigt werden kann, kann dasselbe Ansichtenmodell im gesamten Bildbereich verwendet werden. Damit kann ein Modell also so vor der Kamera positioniert werden, dass es mit seinem Referenzpunkt im Bildhauptpunkt erscheint. Dort wird es mit geeigneter Schrittweite im Azimut und in der Entfernung variiert. Die Schrittweite hängt z. B. von der Bildauflösung und der Modellstruktur ab.

Der sich ergebende zweidimensionale Modellansichtenraum umfasst typischerweise einige hundert Ansichten pro Modell. Er hat die Topologie einer Zylinderfläche. Entsprechend der gewählten Schrittweite sind zwölf benachbarte Ansichten eines Fahrzeugmodells in Abb. 1 dargestellt, die sich nur geringfügig unterscheiden. Waagrecht variiert der Azimut und senkrecht die Entfernung.

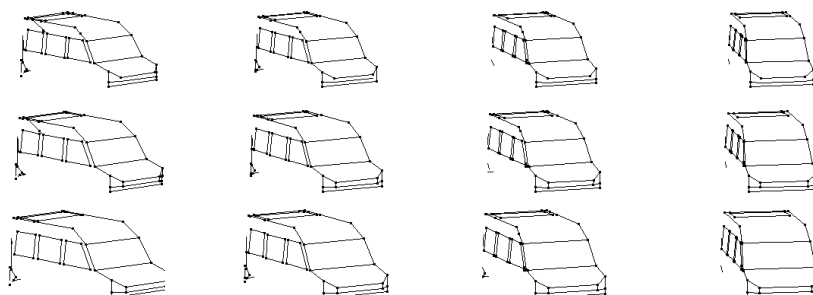


Abb. 1: Einige benachbarte 2D-Ansichten eines 3D-Fahrzeugmodells

## 2.2 Der Vergleich zwischen Bild und Ansichten

Aus dem digitalen Grauwertbild wird durch Gradientenfilter, morphologische Operationen und Approximation der Konturen durch Linien eine Menge von Primitivobjekten (Linieninstanzen) gewonnen (siehe Abb. 4). Ein darauf aufsetzendes Linienverlängerungsverfahren (Produktion p1 in Abb. 4) dient vor allen Dingen der genaueren Erfassung von Orientierungen der Objektkonturen. Die entstehende Menge verlängerter Linien ist mit den Ansichten zu vergleichen. Dafür sind verschiedene Verfahren bekannt wie z. B. alignment [3], geometric hashing [16] oder die generalisierte Hough-Transformation [1]. Das hier vorgestellte Verfahren orientiert sich an der generalisierten Hough-Transformation

Zunächst werden Paare von geeignet liegenden verlängerten Linien zu L-Strukturen oder T-Strukturen zusammengefasst (durch die Produktionen p2 und p3 in Abb. 4). Die gleiche Prozedur wird auch - vorab - auf jeder Ansicht durchgeführt. Als Schlüssel werden die beiden Orientierungen einer solchen Struktur verwendet. Abb. 2 skizziert das Prinzip. Eine Struktur im Bild (z. B. Abb. 2a links) spricht für einen Teil einer Ansicht (z. B. Abb. 2b), wenn beide Orientierungen hinreichend ähnlich sind. Die Lage des Referenzpunktes der Ansicht erhält man, indem man die Position des Ansichtenteils relativ zu ihrem Bildhauptpunkt von der Bildkoordinate der Struktur abzieht (Abb. 2c).

Da nur die Orientierung Verwendung findet, spielt die Länge der Linien keine Rolle. Gegen partielle Verdeckungen ist also eine gewisse Stabilität gegeben. Beim Zusammenfügen von Modellteilen aus verschiedenen Bildbereichen ist zu beachten, dass der einer Ansicht assoziierte Azimutwinkel relativ zum Sehstrahl des Referenzpunktes transformiert werden muss in einen Winkel relativ zur optischen Achse oder absolut im Raum.

## 2.3 Stabilität durch Akkumulation

Häufig kann man in Bildern von realen Szenen nicht alle modellierten Strukturen zuordnen. Man muss also versuchen, möglichst viele konsistente Zuordnun-

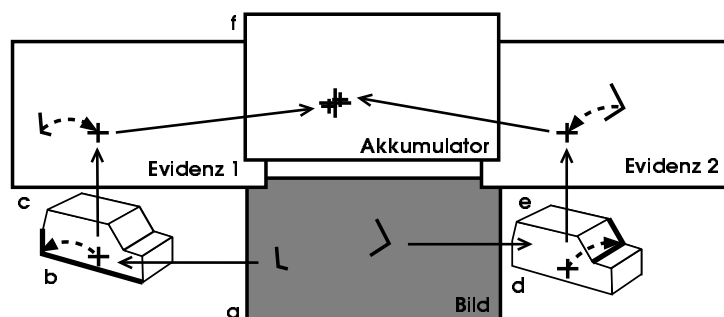


Abb. 2: 2D Vergleich zwischen Bildstrukturen und Teilen von Modellansichten

gen zwischen Modellstrukturen und Bildstrukturen zu vereinigen. Die Detektion und Lokalisation eines Fahrzeuges stützt sich auf viele Hinweise, die für sich alleine jedoch vage sein können.

Da davon auszugehen ist, dass die im Bild bestimmten Orientierungen von Linien fehlerbehaftet sind, sind für 'ähnliche' Orientierungen entsprechende Toleranzen zu berücksichtigen. Bei den hier gewählten Toleranzen können einer einzigen L-Struktur im Bild häufig dutzende oder gar hunderte von Ansichtenstrukturen entsprechen. Diese Hinweise werden alle an ihren entsprechenden Referenzpunktpositionen in den Akkumulator eingetragen. Durch die Überlagerung von Hinweisen auf verschiedene Modellteile für dieselbe Ansicht an derselben Stelle sollte sich eine Anhäufung dort ergeben, wo sich das Objekt befindet. Beispielsweise zeigen Abb. 2.b,c und Abb. 2.d,e zwei Hinweise von verschiedenen Modellteilen auf die gleiche Ansicht. Zur Detektion sucht man also im Akkumulator signifikante Anhäufungen. Aufgrund von Störungen und der Rasterung des Ansichtenraumes sind benachbarte Ansichten und Bildpositionen als kompatibel anzusehen. Zu entscheiden ist dann, welche Hinweise zu einem Cluster gehören. Abb. 2.f) zeigt die Zusammenfassung und Mittelung der beiden Evidenzen.

#### 2.4 Bestandteilhierarchien und artikulierte Modelle

Einige Autoren schlagen vor, die zu erkennenden Objekte bei der Modellierung hierarchisch, d. h. stufenweise aber ohne Zyklen, in geeignete Bestandteile zu zerlegen, so dass nicht das ganze Modell, sondern nur die primitivsten Bestandteile mit den Daten verglichen werden [12,7]. Der Erkennungsprozess kehrt diese Zerlegung dann um und setzt das Objekt aus den Teilzuordnungen schrittweise zusammen, wobei jeweils auf geometrische Konsistenz der Teile zueinander abgeprüft werden kann. Einen solchen Ansatz haben wir für die 3D-Erkennung von 3D-Fahrzeugmodellen in [10] beschrieben. Er lässt sich auf das hier vorgestellte ansichtenbasierte Verfahren übertragen.

Nicht alle Fahrzeuge können adäquat durch starre Modelle erfasst werden.

So können Bestandteile eines Fahrzeuges durch Gelenke oder Lager miteinander verbunden sein oder Züge aus Fahrzeug und Anhänger bestehen. Dabei sind unterschiedliche Freiheitsgrade zu berücksichtigen.

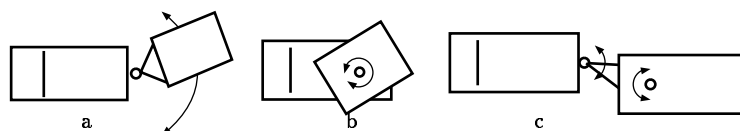


Abb. 3: Artikulierte Fahrzeugmodelle

Abb. 3 zeigt drei Beispiele solcher Konstruktionen, die wir angelehnt an die Literatur [15] als artikulierte Modelle bezeichnen. Das Zugmodell aus Zugfahrzeug und Einachsanhänger in Abb. 3a lässt den Azimut des Anhängers relativ zum Zugfahrzeug (um die Kupplung) in einem weiten Intervall frei. Abb. 3b zeigt ein Fahrzeugmodell mit drehbarem Aufsatz, wie beispielsweise bei einem Autokran. Hier ist der Azimut um das entsprechende Lager völlig frei. Abb. 3c zeigt einen Zug aus Zugfahrzeug und Zweiachsanhänger. Hier gibt es eine tiefere Hierarchie, weil der Anhänger in Drehgestell und Auflieger zerfällt.

Die sich daraus ergebenden Zwangsbedingungen können zur Steigerung der Erkennungssicherheit genutzt werden. Dazu werden in der ansichtenbasierten Erkennung die Teilmodelle in die Szene zurückprojiziert, um die Konsistenzprüfung vornehmen zu können. Es wird geprüft, ob die Translation und Rotation, die die Koordinatensysteme der Teile ineinander überführt, im zulässigen Bereich liegt.

### 3 Implementierung der Fahrzeugerkennung

Zur Modellierung von Fahrzeugen werden Produktionsnetze verwendet [8,10], die auch bereits in anderen Arbeiten, wie z. B. zur Luftbildanalyse, verwendet wurden [13,14]. Produktionsnetze dienen zur semantischen Modellierung und zeigen den generellen Zusammenhang von Produktionen und Objekttypen (Konzepten). Die Implementierung der Produktionsnetze erfolgte in einer Blackboard-Architektur in der Systemumgebung BPI [6].

Jede einzelne Produktion ist durch ein Verarbeitungsmodul realisiert, das den Bedingungsteil der Produktion auf einer Konfiguration von Objektinstanzen prüft, und bei Erfüllung eine neue Instanz generiert. Das System arbeitet akkumulierend. Einmal erzeugte Teilergebnisse stehen so zur Verfolgung verschiedener Hypothesen während der gesamten Analyse zur Verfügung. Die Datenbasis wird in einem Assoziativspeicher gehalten, der einen effizienten Zugriff auf Mengen von Instanzen erlaubt [14,9].

Ähnlich wie bei dem in [10] vorgestellten Netz werden zur Konstruktion parametrisierte Produktionen verwendet. Die Clusterung erfolgt jedoch hier im 2D-Raum. Ein Schema zum Datenfluss ist in Abb. 4 dargestellt.

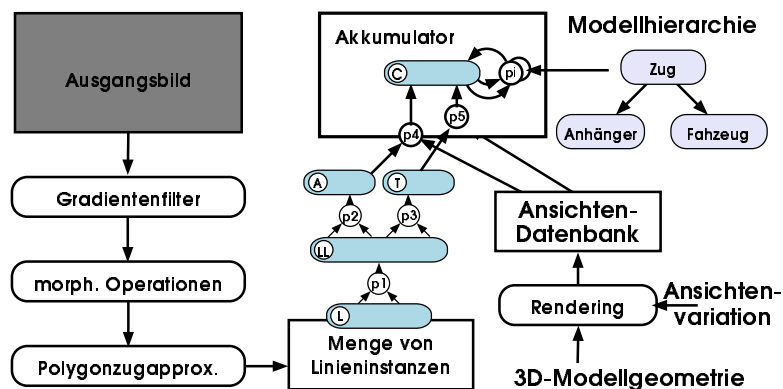


Abb. 4: Verfahrenübersicht und Produktionsnetz

#### 4 Beispiel: Kombi mit Anhänger

Das Verfahren wird derzeit im sichtbaren Spektralbereich anhand verschiedener Fahrzeugmodelle erprobt. Abb. 5 zeigt einen Bildausschnitt, in dem sich unter anderem ein Kombi vom Typ Renault 'Kangoo' mit kleinem Anhänger befindet. In der Mitte der Abbildung sind die gefundene Instanz *Zug* und ihre Teilinstanzen *Kangoo* und *Anhängen* mit einigen Attributen angegeben. Die zu den Teilinstanzen gehörenden Ansichten aus dem gerasterten Modellansichtenraum sind jeweils daneben graphisch dargestellt. Oben in Abb. 5 ist die

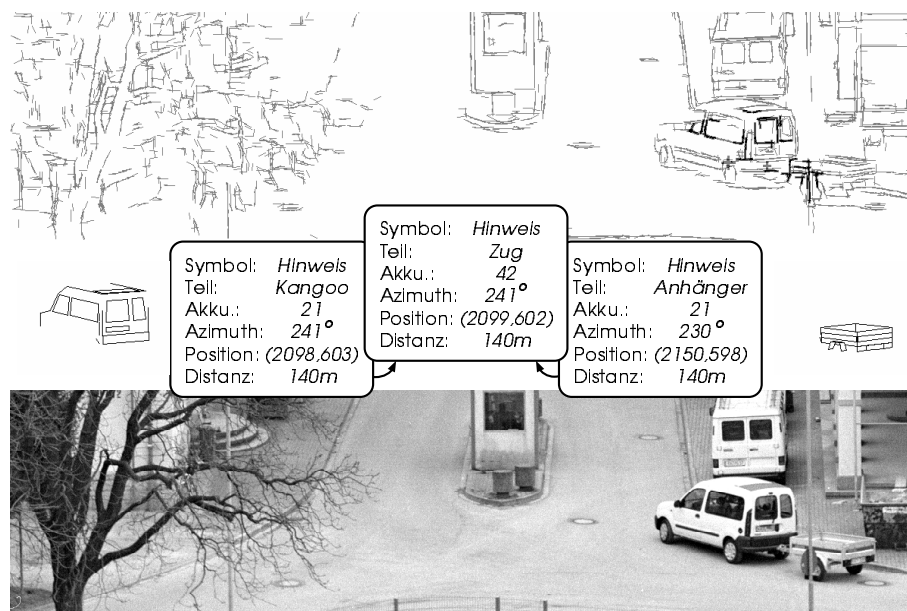


Abb. 5: Analyseergebnis für das Modell *Kombi mit Anhänger*

Menge aller Linieninstanzen des Bildausschnittes gezeigt. Die in die Ableitung des Endergebnis eingehenden Linieninstanzen sind fett hervorgehoben. Man sieht, dass die Anzahl der beteiligten Instanzen recht klein ist. Der Vergleich zwischen Bild und den automatisch gefundenen Ansichten zeigt, dass die gesuchten Objekte richtig lokalisiert und zugeordnet werden konnten.

## 5 Diskussion

Bei der klassischen, generalisierten Hough-Transformation wird der Raum des Akkumulators in Zellen (buckets) diskretisiert. Die Auswertung erfolgt durch Auszählen der Hinweise in den Zellen. Die Dimensionierung der Zellgrößen stellt ein Problem dar, da bei zu kleinen Zellen die Belegung zu dünn wird und bei zu großen Zellen inkompatible Hinweise zusammengefasst werden. Der erreichte Wert einer Zelle hängt auch von der Lage eines Clusters kompatibler Hinweise bzgl. der Zellengrenzen ab. Der vorgestellte Ansatz vermeidet diese Problematik.

Das Verfahren stützt sich ausschließlich auf Konturinformation und ist damit auf Bildmaterial von hoher Qualität bzgl. Auflösung, Schärfe und Kontrast angewiesen. Bei geringer Auflösung und kleinen Objekten lassen sich die Orientierungen nur ungenügend messen. Bei unscharfen oder kontrastarmen Aufnahmen liefert die Merkmalsextraktion nicht genügend Objektkonturen.

Eine Verbesserung wäre zu erwarten, wenn man sich sowohl bei der Modellierung als auch bei der Erkennung nicht nur auf Konturen stützt, sondern Grauwertinformationen einbezieht. Das erfordert auch eine Modellierung der Beleuchtung wie sie z. B. von [4] verwendet wird.

Bei dem vorgestellten Ansatz ist der Beitrag unterschiedlicher Objektstrukturen gleich gewichtet, obwohl gelegentlich bestimmte Objektstrukturen mehr Evidenz liefern sollten als andere. Eine Berücksichtigung dieser Gewichtungen, wie in statistischen Ansätzen (siehe z. B. [5]), lässt eine Steigerung der Robustheit erwarten. Geringe Einschränkungen bzgl. der Lage oder Rotation ergeben eine große Anzahl von Ansichten. Diese führen in Bildern mit stark strukturiertem Hintergrund zu Verwechslungen. Durch Einbeziehung von Szenenkontext, wie z. B. den Straßenverlauf, lassen sich die Detektionssicherheit erhöhen und der Aufwand verringern.

Abb. 6 zeigt ein Beispiel zu dem artikulierten Fahrzeugmodell aus Abb. 3c. Die Container wurden in je zwei gleichartige Containerendstücke zerlegt, die gegeneinander um  $180^\circ$  verdreht sind. Das Zugfahrzeug bestehend aus Führer-

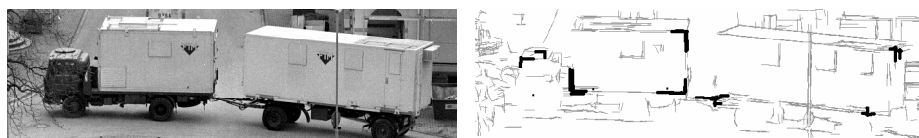


Abb. 6: Beispiel Zugfahrzeug mit Zweiachsanhänger

haus und Container fand eine korrekte Zuordnung, die sich jedoch auf nur sehr wenige Linieninstanzen stützt. Aufgrund des geringen Kontrastes konnte das vordere Containerendstück des Anhängers und somit auch der Anhänger nicht zugeordnet werden. Der gesamte Zug wurde nicht gefunden, da der Anhänger fehlte.

Die Lagebeschränkungen zwischen Zugfahrzeug und Anhänger werden durch die Deichsel bestimmt. In diesem Beispiel wurde zwar die Deichsel richtig zugeordnet, jedoch wird sie in vielen anderen Ansichten nicht sichtbar sein. Dies erfordert, dass solche Objekte nicht nur als Bestandteil, sondern auch als spezielle geometrische Beschränkungen zu modellieren sind.

## Literaturverzeichnis

- [1] Ballard D. H.; Brown C. M.: *Computer Vision*. Prentice Hall, Engelwood Cliffs, New Jersey, 1982.
- [2] Binfort T. O.; Levitt T. S.: *Model-based Recognition of Objects in Complex Scenes*. In: ARPA (Ed.): *Image Understanding Workshop 1994 (Montery)* Morgan Kaufman, San Franzisco, 1994, pp 149-155.
- [3] Grimson L.; Eric W.: *Object Recognition by Computer: The Role of Geometric Constraints*, MIT Press, Cambridge, Mass., 1990.
- [4] Hermiston K. J.; Booth D. M.; Foulkes S. B.; Reno A. L.: *Pose Estimation and Recognition of Ground Vehicles in Aerial Reconnaissance Imagery*. In: Jain K.; Venkatesh S.; Lovell B. C. (Eds.): *ICPR'98*, IEEE, Los Alamitos, 1998, pp. 578-582.
- [5] Hornegger J. M.: *Statistische Modellierung, Klassifikation und Lokalisation von Objekten*, Diss., Univ. Erlangen-Nürnberg, Technische Fakultät, Erlangen, 1996.
- [6] Lütjen K.: *BPI: Ein Blackboard-basiertes Produktionssystem für die automatische Bildauswertung*. In: Hartmann G. (Hrsg.): *Mustererkennung 1986*, (DAGM 86), Informatik Fachberichte Nr.125, Springer, Berlin, 1986, pp. 164-168.
- [7] Matsuyama T.; Hwang V. S.: *Sigma A Knowledge-Based Aerial Image Understanding System*, Plenum Press, New York, 1990.
- [8] Michaelsen E.: *Über Koordinaten Grammatiken zur Bildverarbeitung und Szenenanalyse*, Diss., Techn. Fak., Univ. Erlangen, 1998.
- [9] Michaelsen E.; Lütjen K.; Stilla U.: *Associative Access and Special Hardware for Production Nets*, German Russian Workshop on Pattern Recognition GRWS-98, Infix, 1999, pp. 290-297.
- [10] Michaelsen E.; Wankmüller U.; Stilla U.: *Wissenserwerb für Produktionsnetze zur 3D-Erkennung von Fahrzeugen*, Mustererkennung 1998, Berlin, Springer, 1998, pp 507-514.
- [11] Nagel H.-H.; Schwarz Th.; Leuk H.; Haag M.: *Tracking Turning Trucks with Trailors*. In: Maybank S.; Tan, T. (Eds.): *IEEE Workshop on Visual Surveillance*, IEEE Comp. Soc., Los Alamitos, 1998, pp 65-72.
- [12] Niemann H.: *Pattern Analysis and Understanding*, Springer, Berlin, 1990.
- [13] Stilla U.: *Map-aided Structural Analysis of Aerial Images*. ISPRS Journal of Photogrammetry and Remote Sensing, Vol 50, 1995, pp 3-10.
- [14] Stilla U.; Michaelsen E.; Lütjen K.: *Structural 3D-Analysis of Aerial Images with a Blackboard-based Production System*. In: Gruen A.; Kuebler O.; Agouris P.: *Automatic Extraction of Man-Made Objects from Aerial and Space Images*, (Ascona Workshop der ETH), Birkhäuser, Basel, 1995, pp 53-62.
- [15] Wang P. S. P.: *Parallel Matching of 3D Articulated Object Recognition*. Int. Journ. of Pattern Recognition and Artificial Intelligence, Vol 13, 1999, pp 431-444.
- [16] Wolfson H. J.: *Model Based Object Recognition by Geometric Hashing*. In: Faugeras O.: *Computer Vision - ECCV90*, Springer, Berlin, pp 526-536.