RECOGNITION OF PARTIALLY OCCLUDED MOVING OBJECTS

Michael Schmid

Universität der Bundeswehr München
Fakultät für Luft- und Raumfahrttechnik
Werner-Heisenberg-Weg 39, D-8014 Neubiberg, Germany

Abstract

A computer vision system in an autonomous vehicle guidance application is presented for interpreting image sequences acquired by a camera moving relative to the environment. Objects with different shapes and changing positions as well as motion parameters in the perceived scene have to be recognized even if they occlude each other. The approach described is based on checking hypotheses by a combination of methods from knowledge representation and from control theory, e.g. recursive estimation. Hypothesis verification is done by analysing the estimated motion parameters using methods from statistics. These algorithms have been implemented and tested on synthetic images. Tests using noise corrupted measurements from a CCD-camera are currently performed.

Keywords

Computer vision, 3D-object recognition, occlusion, hypothesis generation and verification, recursive estimation

I. INTRODUCTION

Recognizing shape and position of three-dimensional (3D) rigid objects of a given scene is regarded as one of the main research fields in computer vision. There exist many different techniques to handle this task in moderately complex situations successfully; to get an overview see e.g. [Brady 81], [Chelappa et al. 90] and [Enkelmann 90]. But increasing complexity of the scene observed causes significant problems in identifying and locating the objects of a given situation. In most cases multiple objects with different shape and motion may appear or disappear, and probably they may partially occlude each other. This fact complicates the task of object recognition, but it is an essential feature of a computer vision system to be able to deal with a wide range of everyday situations including partially occluded objects.

Occlusions occur usually in every kind of image processing application by different reasons. By using a CCD-camera the viewing angle onto the environment is restricted. This fact causes a clipping of the observed objects, if they are moving at the verge of the perceiving CCD-chip. Also occlusions may result from the movement of the camera relative to the surrounding environment or from autonomous moving objects, e.g. cars overtaking each other. The research work discussed here deals with occlusions arising from situations of overtaking cars on German motorways. But it should be no problem to adapt the algorithms to different situations. Figure 1 shows a synthetic image of a German standard "Autobahn" scene generated by a graphic-workstation with two cars (similar to trucks) driving in front of the ego-car causing occlusions.

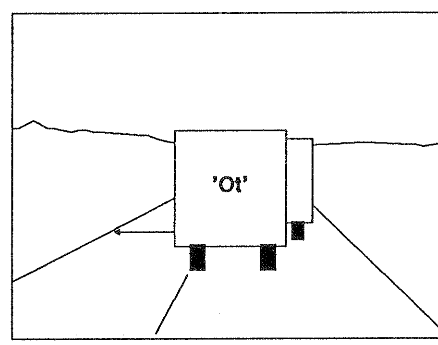

Figure 1 Traffic situation with an occluded object

Section II starts with a short introduction of the machine vision system for autonomous vehicle guidance on motorways developped at the 'Universität der Bundeswehr München' (UniBwM) by the group of Prof. Dickmanns. Section III gives an overview of the prerequisites for the internal adaptive model of the real world inside the image processing system, which is necessary to compare the measurements from the camera with the internal description of the tracked objects for updating the estimated parameters of these objects. Section IV describes the process of initializing an object hypothesis by the assumption of an existing occlusion. A method how to assess the generated hypothesis is represented in Section V. The requirements for the implementation and some practical results are pointed out in Section VI. Finally Section VII summerizes the results and gives an outlook on future research works.

II. SYSTEM OVERVIEW

The structure of the object recognition module will be described now to give an overview (figure 2).

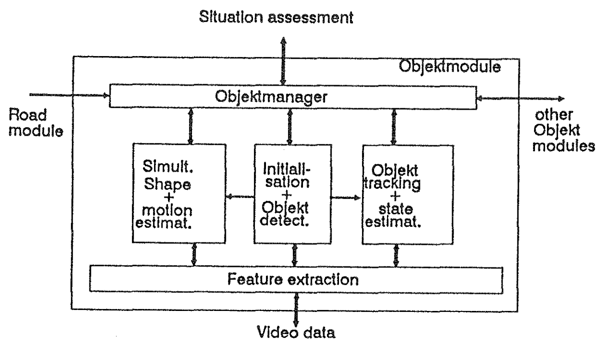

Figure 2 Structure of the object recognition module

An already existing software module performs the detection and the tracking in the image plane of an single not occluded object 'Ot' driving in front of the own vehicle [Solder, Graefe 90], [Regensburger, Graefe 90] and [Thomanek, Dickmanns 92]. Algorithms on a parallel processor system exploiting high level spatio-temporal object models estimate recursively in real-time the relative state of the object using Kalman filtering techniques for state estimation [Dickmanns, Christians 89] and [Dickmanns et al. 90]. Yet another module determines the shape of the tracked object 'Ot' while the aspect conditions change during motion [Schick, Dickmanns 91]. These informations are the input to an extended object recognition module treating more complex situations where occluded objects may appear, for example, while lane changing. If an additional, partially occluded, object has been detected and verified, a module like the one mentioned above is initialized to track the new object. The approach presented in this paper allows the handling of occluded 3D rigid objects on curved roads in real-time image sequences by combining knowledge based methods for feature matching and motion classification with techniques from system theory for motion estimation.

This vision system uses a-priori knowledge about the expected objects concerning possible shape, position, and motion to generate a hypothesis of an object in the scene. The knowledge base contains the modeling information about different object shapes (e.g. car types) with their characteristic features and about constraints in motion. An internal representation in the computer system of the world around the camera, installed in the vehicle, is required in the analysis by synthesis'-method selected because the system compares a generic internal model with the real situation outside. Therefore, different coordinate systems have to be introduced (see Section III). The first hypothesis instantiated by the real-time object tracking module is a 3D-rectangular parallelepiped (wire frame model) encasing the object. Another, more sophisticated process estimates the shape in more details, but up to now not under real-time conditions. The tracking process evaluates a set of state variables containing the object position and motion by using a 4D-model. A further process estimating the curvature of the road communicates its results to the object tracking module, which is then able to determine the relative position of object 'Ot' on the road.

III. MODELING OF THE PERCEIVED SCENE

The internal representation of the real world consists mainly of two parts. First it is necessary to introduce a different coordinate system $(x,y,z,\psi,\theta,\varphi)$ for each object in the real world, e.g. ego-car, camera, road, other cars, traffic signs, ..., to allow modeling of independent movements between the different objects in the scene (figure 3). This implies for the object recognition task in an autonomous road vehicle guidance application an extra coordinate system P for the two axis platform, which allows to control the pan and the tilt angle of the camera in order to track an object in the environment. The other coordinates (x,y,z,φ) are constant relative to the ego-car. The main coordinate system E belongs to the own vehicle, which is adjusted every system cycle of the image processing system, because every object location or motion is only relevant relative to the own car's position. A separate road and ego-state module determines the state variables of the ego relative to the road and of the curvature describing the road in front of the car (in the future also in the back) in the viewing range. A further coordinate system R has to be introduced for the description of the road, because the estimated position of the other cars moving on the road are only relevant relative to the road. To describe the curvature of the road a clothoidial representation is applied. Finally, there exists a specific coordinate system O for each object driving on the road. The distance of other cars to the own is approximated by the Pythagoras equation of the road coordinates at the estimated location of the tracked vehicle. This technique is well suited for slight curvatures. That way the motion of an tracked object can be transformed in camera fixed coordinates K by calculating

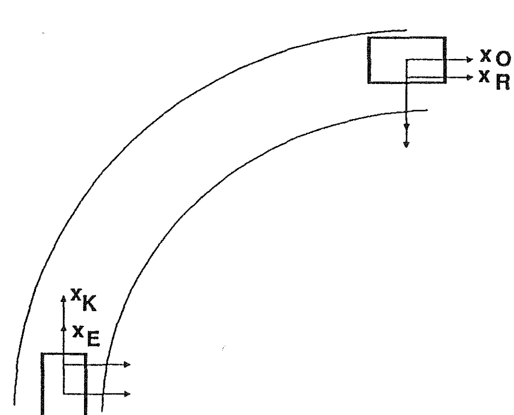

Figure 3 Coordinate systems for modeling

$$S_{\mathrm{K}} = R(S_{\mathrm{O}} - S_{\mathrm{O_o}})\, R(S_{\mathrm{R}} - S_{\mathrm{R_o}})\, R(S_{\mathrm{E}} - S_{\mathrm{E_o}})\, R(S_{\mathrm{P}} - S_{\mathrm{P_o}}) \qquad (1)$$

with

$$S = (x, y, z, \psi, \theta) \qquad (2)$$

$$R = \begin{pmatrix} \cos\psi & \sin\psi & 0 \\ -\sin\psi & \cos\psi & 0 \\ 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} \cos\theta & 0 & -\sin\theta \\ 0 & 1 & 0 \\ \sin\theta & 0 & \cos\theta \end{pmatrix} \qquad (3)$$

Second an internal model of the object to be tracked is needed including a rough description of the object shape and the dynamics of object motion. The dynamical model of the vehicle is approximated by assuming constant velocity for the object motion parameters. The object shape is modeled as proposed in [Schick 92] by using a polygone model with 12 planes, 26 edges and 48 nodes describing the objects surface (figure 4). This generic shape model is independent of the aspect angle of the viewer and allows a flexible modeling by changing the form parameters. Because of the symmetry the number of independent parameters can be reduced to 12.

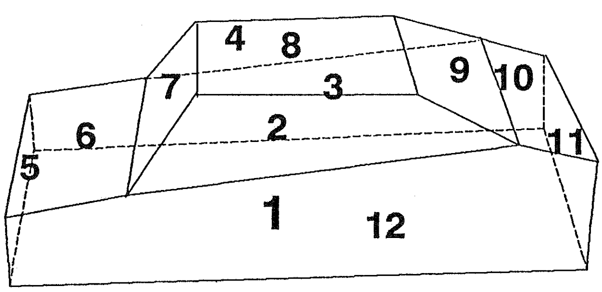

Figure 4 Generic 3D shape model [Schick 92]

By utilizing these variable parameters nearly every kind of car type, like limousines, coupes, pickups or trucks can be coarsely modeled by introducing different metric proportions for each parameter. For the first hypothesis of an object it is often sufficient to apply a simple 3D-rectangular parallelepiped model enveloping the tracked object. This can be modeled by restricting some form parameters of the generic shape model (figure 5). Thus the generic model is easy to adjust to different car types which are analysed by a separate module for shape estimation.

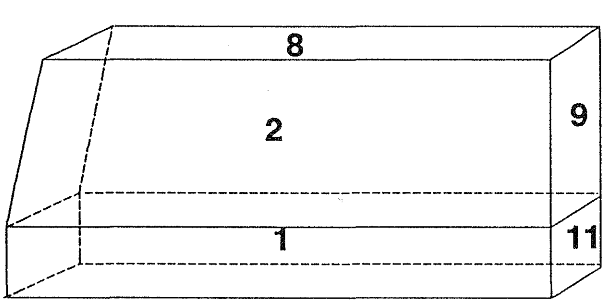

Figure 5 Simplified shape model for a truck

IV. HYPOTHESIS GENERATION

Usually, the object detection module searches for obstacles in a certain area of the image in front of the ego car depending on the actual curvature of the road, which is estimated and described by the road module. This technique of initializing the object tracker works only for the detection of single and not occluded objects. In the case of occluded moving objects a more sophisticated method for generating an initial hypothesis of an object is required. Figure 6 demonstrates the main components of the extension of the object recognition module for handling situations with occluding objects. A hypothesis consists of the supposed shape of the object and an assumed location in the scene. Recognizing only part of shape of an object in the scene may produce a valid hypothesis, because we assume that the tracked object 'Ot' occludes the other region of the second object.

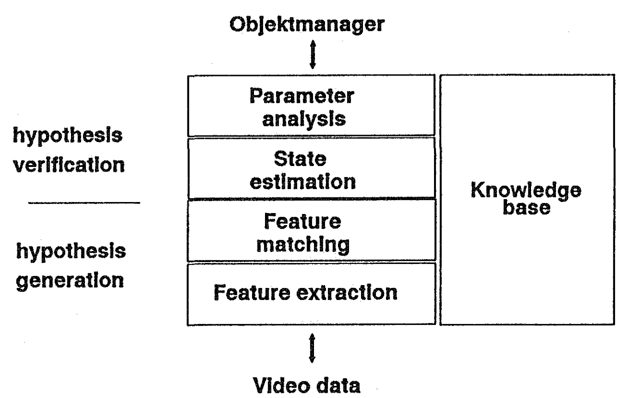

Figure 6 Components for recognizing occluded objects

Usually, occluded objects appear in front of the tracked object 'Ot' at its left or right side. Therefore, the extended object recognition module for occlusions searches near the left and the right boundary of object 'Ot' for some characteristic features like corner points or edges, which can be grouped to a partially occluded box. These features are generated by exploiting the information of the estimated position of object 'Ot'and the estimated road curvature. Thus the aspect conditions of the assumed occluded object can be taken into consideration an analysis can be performed in order to determine the visible and measureable features, which are to be extracted from the video image. Matching of the extracted features to an internal model of the object is performed by a set of rules from the knowledge base. Because of the occlusion it is not possible to constrain the number of features by symmetry. Thus a hypothesis is generated if a set of features appears which may belong to an occluded object. In the next section a method will be discussed for verifying these generated object hypotheses.

V. HYPOTHESIS VERIFICATION

The hypothesis verification algorithm checks if the assumed hypothesis of a second occluded object becomes consistent over some images by using a linearized Kalman filtering technique to estimate the state variables of the partially occluded object. Otherwise the hypothesis of a second occluded object will be canceled and the generation mechanism goes on searching at the left and right boundary of the tracked object 'Ot'. The results are analysed by fuzzy sets in a knowledge base exploiting the possible constraints in motion parameters (figure 6). The state variables of each object tracked are evaluated by an recursive estimation algorithm. The Kalman filter for the optimal estimate of x(k) is divided into two steps [Brammer, Siffling 77] and [Maybeck 79]:

1. prediction (extrapolation) of $\hat{x}(k)$
2. innovation (correction) by measurement update.

In order to improve the performance and the handling of the recursive estimation process some additional features as described in the sequel had been added. The order of the system matrix modeling the dynamical model is reduced from 4 (6) state variables (distance, velocity, lateral offset, lateral velocity, ev. yaw angle and yaw velocity) to two (three) matrices with the order of two by estimating position and velocity of each degree of freedom separately. That way the efficiency could be improved without loosing much performance. For the estimation of distance the width respectively the height of the object measured in pixel is the input to the estimation process. In this case the evaluation of the observation matrix C(k) which has to be done every system cycle becomes rather complex if analytically done, but by using numerical differencing techniques this task can be solved in an easier manner. In order to improve the initialisation phase of the estimation process the error of the system model is represented by an exponentially decreasing function Q(k) (figure 7). So the system variance Q is about an order of magnitude higher in the beginning of each estimation than later on. The dynamical model for all estimated state variables and parameters is

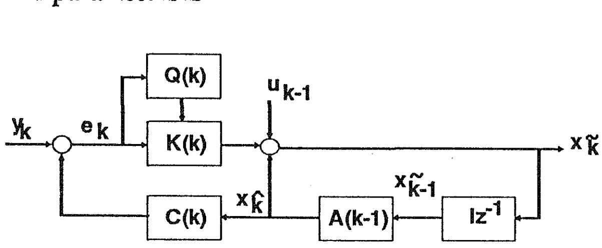

Figure 7 Recursive estimation by a Kalman filter scheme

$$\begin{bmatrix} x_s \\ \dot{x}_s \end{bmatrix}_{k+1} = \begin{bmatrix} 1 & T \\ 0 & 1 \end{bmatrix}_k + \begin{bmatrix} q_{x_s} \\ q_{\dot{x}_s} \end{bmatrix}_k \tag{4}$$

The state vector can be substituted by the desired value to be estimated

$$\begin{bmatrix} x_s \\ \dot{x}_s \end{bmatrix}_k = \begin{bmatrix} x_R \\ \dot{x}_R \end{bmatrix}, \begin{bmatrix} y_0 \\ \dot{y}_0 \end{bmatrix}, \begin{bmatrix} \psi_0 \\ \dot{\psi}_0 \end{bmatrix}, \ldots \tag{5}$$

Q: covariance matrix of system error
A: system transition matrix
C: observation matrix
K: Kalman gain matrix

For the partially occluded object an estimation process is instantiated also. But the results are worse in general because the measurement vector is reduced due to occluded features. Nevertheless, the state variables can be estimated but the estimation error has increased (figure 8a and 8b). The standard deviation for the estimated distance of the occluded object is about two times larger than for the not occluded object. In the case that the estimation process for the second occluded object becomes consistent the results were analysed by comparing them with those of the original tracked object 'Ot'.

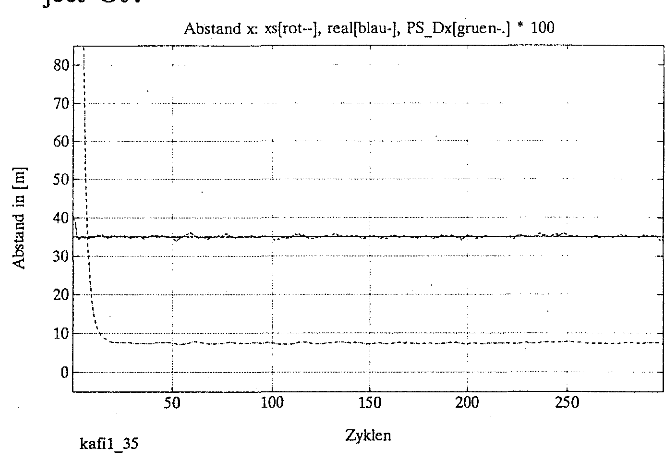

Figure 8a Estimated distance with a complete measurement vector

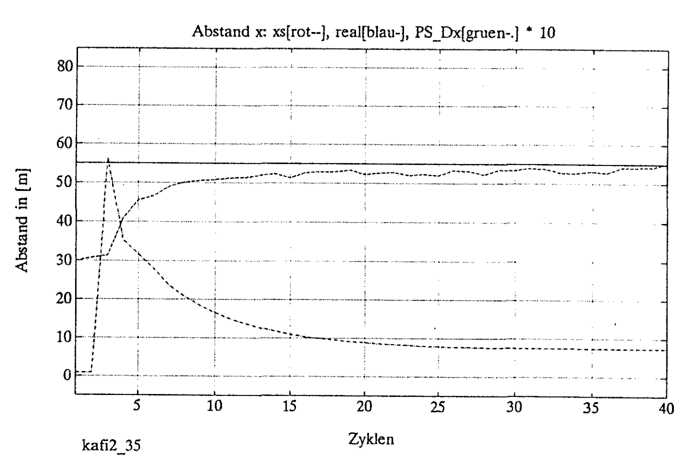

Figure 8b Estimated distance with a reduced measurement vector due to occlusion

In the "Autobahn" driving application the estimated motion and location parameters of the object 'Ot' and the occluded object should be similar, because both of them are driving in the same direction on the same road. But the position state components may not be identical, for this case the two objects would occupy the same space. They may be close to each other, e.g. a truck linked with its trailer, and the assumed hypothesis of two independently moving objects may have been wrong. With this method it is also possible to distinguish objects standing or parking beside the road, being uncovered, when the object 'Ot' continues to move in front of the own vehicle. By this way it is verified that the generated hypothesis of an occluded object is either a casually existing object in the surrounding like bridges or stakes nor a part of the originally tracked object 'Ot', because of a not exactly matching shape model. Therefore, this module is also able to provide information to the situation assessment module about objects located near by the road boundary.

In this approach decisions are made by analyzing the estimated motion and position parameters statistically over a period of time with respect to the existence of a second partially occluded object. If the hypothesis of an occluded object was verified the algorithm tracks the object until it disappears.

VI. IMPLEMENTATION AND EXPERIMENTAL RESULTS

To verify the practicality of the proposed approach the extended object recognition module has been implemented. All the experimental work is done in a closed loop simulation consisting of a graphics workstation for image generation and a parallel-processor system for image processing. Two possible modes of operation are implied. In the first one, only the original synthetic images of the workstation are used to test the implemented algorithms with a cycle time of 80 ms. In the second one, a real CCD-camera takes the images from the graphic screen, and all the problems using noisy measurement data under different viewing and illumination conditions were covered. All implementations have been done in C to achieve the real-time demands. Up to now, all research work was performed in the software simulation environment without using real images from a CCD-camera. The results achieved show, that the analysis of a hypothesis takes about 20 video cycles to make a decision about the assumption of an partially occluded object.

The following figures show the estimated state variables for the object location (distance, lateral offset) of two cars one overtaking the other by using simulated noisy measurements. Figure 9 illustrates the results produced by the verification algorithm.

Cycle	results from motion analysis
29	generating Hypothesis of an second occluded object
29	Initialisation Hypothesis with identical motion
42	second occluded object moving in front (right)
108	two different motion types
137	changing Hypothesis
202	second occluded object moving in front (left)
234	canceling Hypothesis of an occluded object

Figure 9 Analysis of a motorway situation with an occluded object

Exploiting the estimated velocities becomes very difficult, because the estimation error has the same magnitude as the estimated state variable. Therefore it seems to be sufficient to analyse the position parameters. In the case, that velocities are to be analysed additionally it is recommended to smooth the state variables in a succeeding low-pass-filter (figure 13). Figures 10a and b show the estimated distance of the two objects. The second object hypothesis is initialized at cycle 29 with the initial assumption that the occluded object belonges to the already tracked object 'Ot'. At cycle 42 an occluded object was verified and the motion analysis results with the statement that an other object is moving

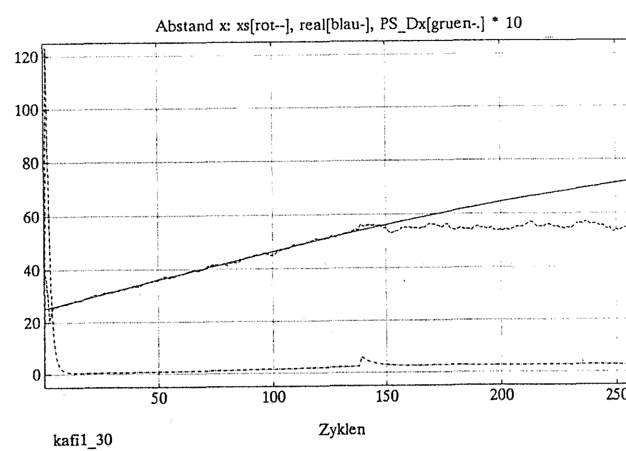

Figure 10a Estimated distance of object Ot

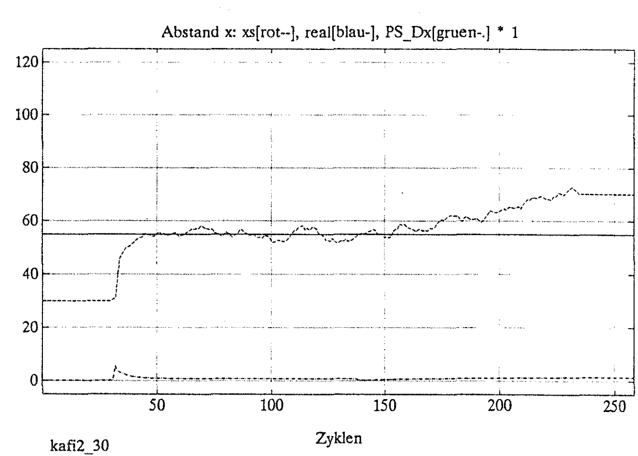

Figure 10b Estimated distance of the occluded object

in front of the first one. When both objects are nearly at the same distance and the occlusion has disappeared the analysis states two different separated moving objects at cycle 108 . Then, the object 'Ot'is passing the other slower object and starting to get occluded by the overtaken object on the right lane. The hypothesis for the occluded and not occluded object has to be exchanged (cycle 137). At cycle 234 the hypothesis of an occluded object has been canceled, because the overtaking object changed back to the right lane and therefore disappeared. The standard deviation for the distance increases with distance as supposed, but stays below 2% of the real distance. Distinctly one can recognize the changing of the hypothesis in figures 10-13 at cycle 137.

Figure 11a and b illustrate the same situation for the lateral offset of the two objects. The standard deviation stays below 1% of the real lateral offset.

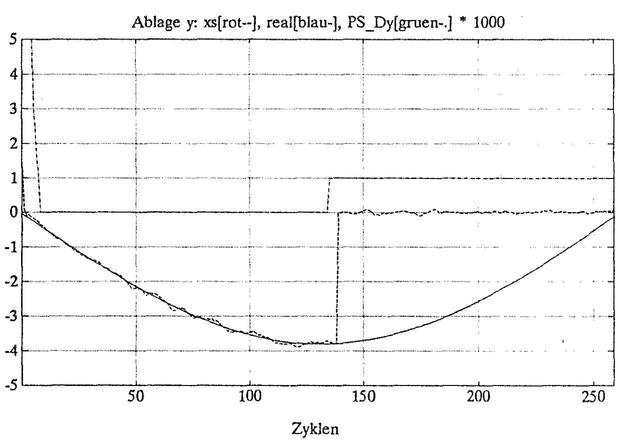

Figure 11a Estimated lateral offset of object Ot

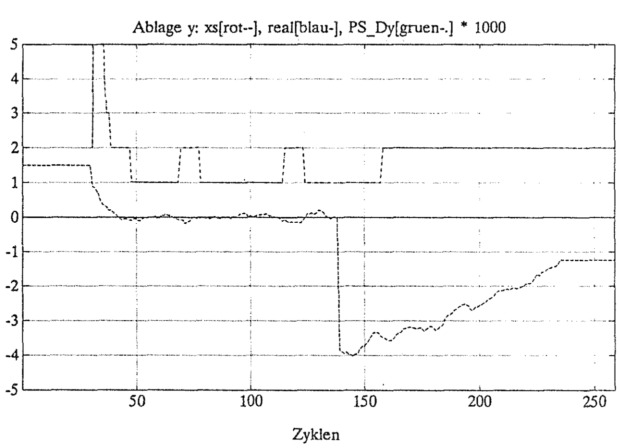

Figure 11b Estimated lateral offset of the occluded object

Figures 12a and b show the difference between the two moving objects concerning distance and lateral offset which is used for the motion analysis. The solid curve is the exact difference of the estimated state variables in comparison with the dotted curve which is the low pass filtered value, but for the position parameters it is not really necessary to filter contrary to the analysis of velocities (figure 13).

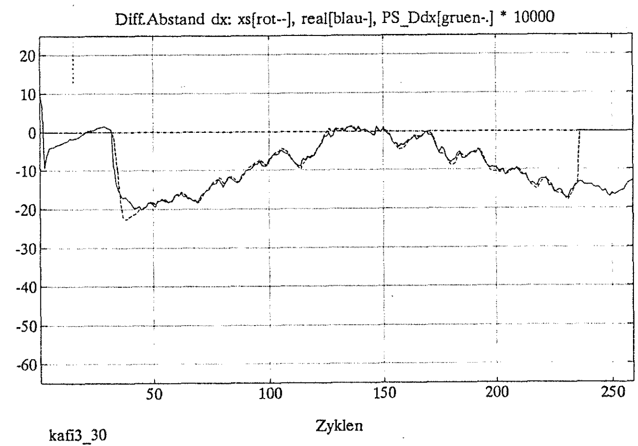

Figure 12a Difference in distance

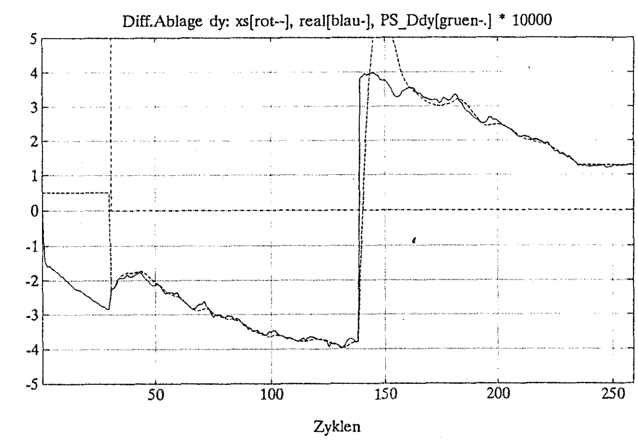

Figure 12b Difference in lateral offset

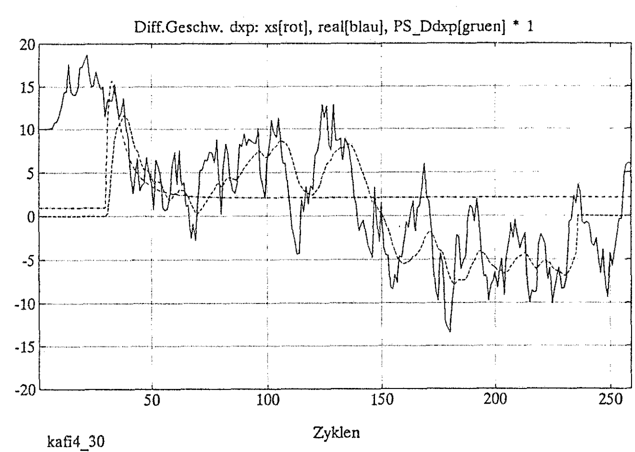

Figure 13 Difference in longitudinal velocity

Currently experimental work is being conducted using images of real scenes on an image processing system consisting of a PC-based cluster of 7 Transputers (figure 14); practical results are expected for the near future.

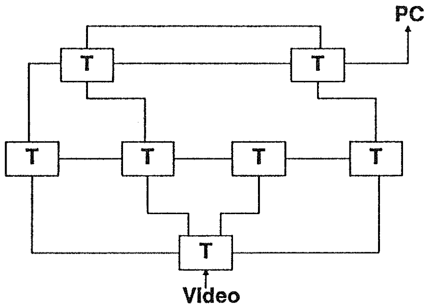

Figure 14 Hardware architecture of the image processing system

VII. CONCLUSIONS

A model based dynamic vision system has been presented for recognition of partially occluded three-dimensional rigid objects in an application for autonomous road vehicle guidance on German standard "Autobahnen". Only image sequences are used as input data for the vision system. On the first processing level a hypothesis is generated using knowledge about the possible location of the appearance of an occluded object with an expected shape by matching measured features with the internal generic model. In the next step tracking and motion estimation of the hypothetical object is solved by a recursive estimation algorithm resulting in the state of the object relative to the road observed. Finally, the verification of the hypothesis is performed exploiting knowledge about the possible range of the estimated motion parameters and consistency over time. The validity of the approach presented was tested with synthetic image sequences, and promising results have been obtained indicating that the application to noisy measurement data from real world scenes should lead to useful results.

The main objective of this paper was to introduce a systematic method of dealing with partial occlusion of objects. For confirmation of the hypothesis it is conceivable that this approach may be extended to checking the consistency of an hypothesized model shape; additionally, it may become more robust by using information from the images such as color, texture, segmentation, etc.

REFERENCES

[Brady 81] M. Brady: "Computer Vision", Noth-Holland Publishing Company, Amsterdam, 1981

[Chellappa et al. 90] R. Chellappa, T.J. Broida: "Recursive 3D-Motion Estimation from a Monocular Image Sequence", IEEE Aerospace and Electronic Systems 26(4), pp 639-656, 1990

[Enkelmann 90] W. Enkelmann: "Obstacle Detection by Evaluation of Optical Flow Fields from Image Sequences, Proc. of the Computer Vision ECCV 90, pp 134-138, 1990

[Brammer, Siffling 75] K. Brammer, G. Siffling: "Kalman-Bucy-Filter, Deterministische Beobachtung und stochastische Filterung, Oldenbourg Verlag , 1975

[Dickmanns, Christians 89] E.D. Dickmanns, T. Christians: "Relative State Estimation for Autonomous Visual Guidance of Road Vehicles", IAS-2, in Kanade et al.(ed), Amsterdam, pp 683-693, 1989

[Dickmanns et al. 90] E.D. Dickmanns, B. Mysliwetz, T. Christians: " Spatio-Temporal Giudance of Autonomous Vehicles by Computer Vision", IEEE Transactiona on System, Man and Cybernetics 20 (6), Special Issue on unmanned Vehicles and Intelligent Robotic Systems, pp 1273-1284 ,1990

[Dickmanns 91] E.D. Dickmanns: "Dynamic vision for locomotion control - An evolutionary path to intelligence", CCG-Lehrgang SE 3.02, 1991

[Maybeck 79] P.S. Maybeck: "Stochastic Models, Estimation and Control Volume 1+2, Academic Press, 1979, 1982

[Regensburger, Graefe 90] U. Regensburger, V. Graefe: "Object Classification for Obstacle Avoidence", Proc. of the SPIE Symposium on Advances in Intelligent Systems, Boston, pp 112-119, 1990

[Schick, Dickmanns 91] J. Schick, E.D. Dickmanns: "Simultaneous Estimation of 3D Shape and Motion of Objects by Computer Vision",IEEE Workshop on Visual Motion, Princeton, N.J., 1991

[Schick 92] J. Schick: "Gleichzeitige Erkennung von Form und Bewegung durch Rechnersehen", Dissertation UniBwM, Fakultät für Luft- und Raumfahrttechnik, 1992

[Solder, Graefe 90] U. Solder, V. Graefe: "Object Detection in Real Time", Proc. of the SPIE Symposium on Advances in Intelligent Systems, Boston, pp 104-111, 1990

[Thomanek, Dickmanns 92] F. Thomanek, E.D. Dickmanns, "Obstacle detection, Tracking and State Estimation for Autonomous Vehicle Guidance", IEEE/RSJ Int. Conf. on Intelligent Robots and Systems, Raleigh N.C., 1992