Intelligent Vehicles '94
October 24-26, 1994, Paris

THE SEEING PASSENGER CAR 'VaMoRs-P'

E.D. Dickmanns[*], **R. Behringer**[*], **D. Dickmanns**[**], **T. Hildebrandt**[*],
M. Maurer[*], **F. Thomanek**[*], **J. Schiehlen**[*]

Universität der Bundeswehr München, D-85577 Neubiberg, Germany
[*]*Fakultät Luft- u. Raumfahrttechnik, Institut f. Systemdynamik,* [**]*Fakultät f. Informatik, Institut 2,*

Abstract: A passenger car Mercedes 500 SEL has been equipped with the sense of vision in the framework of the EUREKA-project **'Prometheus III'**. Road and object recognition is performed both in a look-ahead and in a look-back region; this allows an internal servo-maintained representation of the entire situation around the vehicle using the 4-D approach to dynamic machine vision. Obstacles are detected and tracked both in the forward and in the backward viewing range up to about 100 meters distance; depending on the computing power available for this purpose up to 4 or 5 objects may be tracked in parallel in each hemisphere. A fixation type viewing direction control with the capability of saccadic shifts of viewing direction for attention focussing has been developed.
The overall system comprises about 60 transputers T-222 (16-bit, for image processing and communication) and T-800 (32-bit, for number crunching and knowledge processing). Beside a PC as transputer host all other processors in **VaMoRs-P** are transputers. A description of the parallel processing architecture is given; system integration follows the well proven paradigm of orientation towards 4D physical objects and expectations with prediction error feedback. This allows frequent data driven bottom-up and model driven top-down integration steps for efficient and robust object tracking.

Keywords: Autonomous mobile systems, machine vision, machine perception, data fusion, parallel computing

INTRODUCTION

Based on experience from equipping two vans, a bus and an all-terrain vehicle with vision systems, in cooperation with Daimler-Benz Research a passenger car Mercedes 500 SEL has been equipped in the framework of the EUREKA-project **'Prometheus III'** with a range of sensors for autonomous navigation comprising the sense of vision and inertial sensors for accelerations and angular rates. Eight years of development and testing with the 5-ton van **VaMoRs** of UniBwM have led to a powerful yet small vision system in the new test vehicle **'VaMoRs-P'** based on transputers; it consists of two sets of cameras fixed relative to each other on a platform for viewing direction control both behind the front and the rear windshield. At least two miniature CCD-cameras will be used on each platform exploiting the bi- or multifocal mode of vision.

Active vision has been expanded to the capability of saccadic vision in order to realize fast attention focussing and high resolution image processing in areas of special interest as, for example, traffic signs and other objects. The 4-D approach allows for intelligent combinations of feedforward and feedback viewing direction control taking egomotion effects into account.

In this paper, a survey is given only; road recognition according to the latest standard is discussed in [Behringer 94], obstacle detection, tracking and relative state estimation in [Thomanek *et al.* 94], viewing direction control in [Schiehlen *et al.* 94] (all in this volume).

After a brief discussion of the overall system architecture based on spatio-temporal (4-D) processes with orientation towards physical objects and temporal events, the main components will be reviewed: Image acquisition and distribution via the technical eyes and the Transputer Image Processing system (TIP), feature extraction with the software system KRONOS running on 16-bit T-222 transputers as the lowest level of the object processor groups (OPG), the specific OPG's for the object classes 'roads' and 'obstacles', the dynamic data base (DDB) as the means for object-oriented data exchange between the system components, and finally, the component for situation assessment and vehicle control.

SYSTEM ARCHITECTURE

Figure 1 shows the overall system architecture based on the 4-D approach to dynamic machine vision [Dickmanns 92]; for the first time it encompasses two 'technical eyes' for both forward and rearward perception of the environment. Both eyes will have at least bifocal vision in their final stage of development, very probably even trifocal vision with focal length ratios of three to four; the platforms specifically developed for this application have several modes of operation like inertial stabilization (horizontally), smooth visual pursuit, saccadic shift of attention, and search. Up to four 1/2-inch miniature CCD-cameras may be packed on one platform; the cameras deliver both color and intensity signals which will be used according to the special needs.

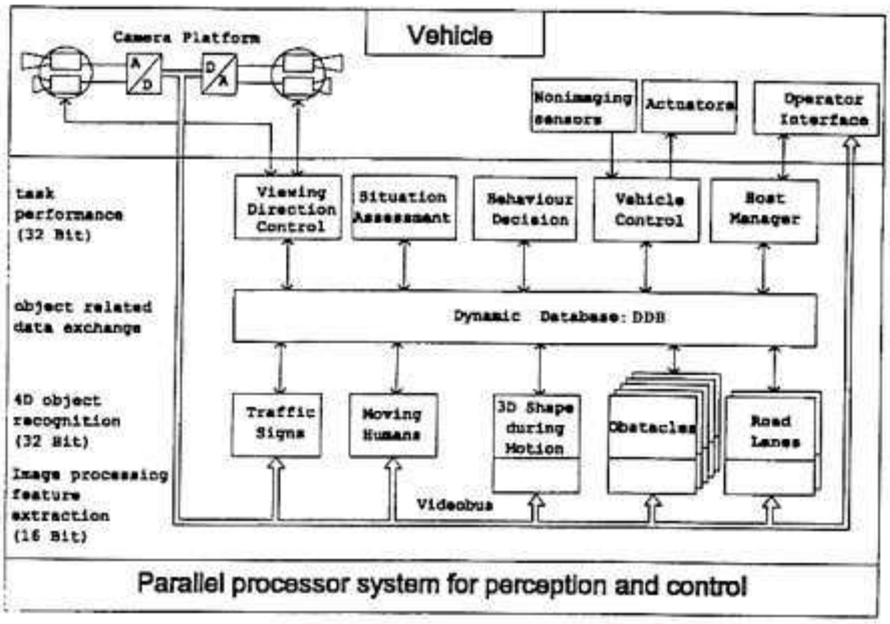


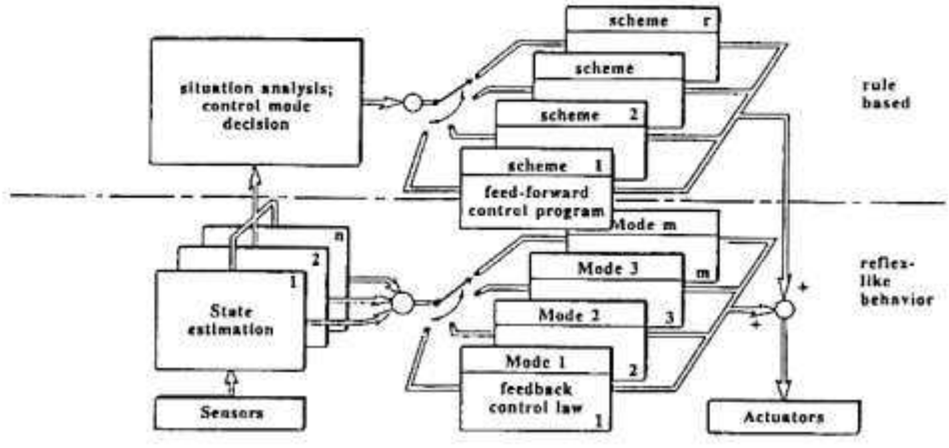
Figure 1. Overall object - oriented system architecture

The video data are A/D-converted in the framegrabber stage of the TIP (top left); two mono-framegrabbers (MFG) and one color framegrabber are in use at present, each capable of handling up to 4 monochrome and up to 2 color video signals (by time multiplex). These digital data streams are then distributed to the 'Versatile Processing Units' (VPU) taking their needs for object-oriented data interpretation into account; for example, the 'road detection and tracking'-OPG (RDT) receives data from the wide angle camera for road interpretation nearby and from the tele-camera for more distant road segments (see fig. 3 below). Similarly, the 'obstacle detection and tracking'-OPG's receive both images and make use of them according to the actual situation.
Each OPG (lower blocks in fig.1) consists of a VPU for data distribution according to the request of the '4-D object processor' (OP, see below), optionally, the T-222's for image feature extraction with the KRONOS-software, and the upper level 32-bit T-805 object-processors (OP) which implement the recursive state estimation algorithms and the functions of process management for interpretation and object perception (see [Behringer 94; Thomanek *et al.* 94]).

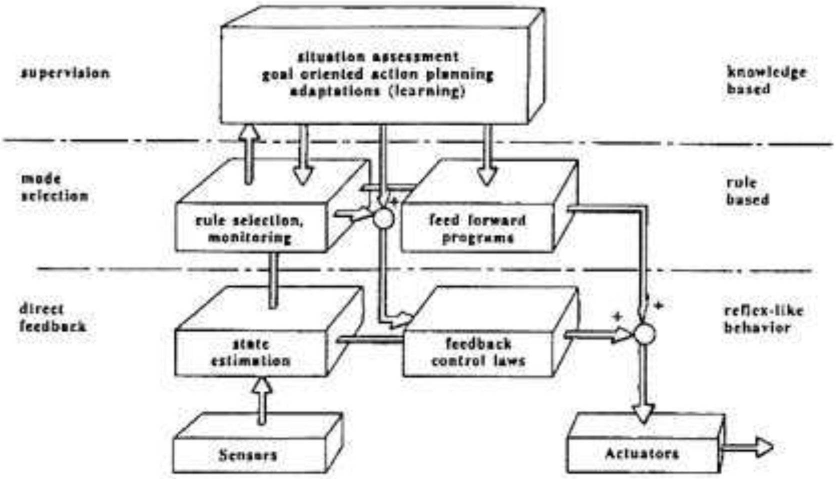
The results of these OPG's running at 12.5 or 25 Hz processing rate are sent to the DDB (central horizontal bar in fig.1) for further distribution according to a predefined schedule. The DDB is the exchange platform defined as a standard for implementation-independent data communication; in the Daimler-Benz vehicle VITA_2, also equipped with the core of our transputer system, it is used for coupling other vision subsystems into the overall system; these other sytems may run on different hardware and with different software.

On the higher levels in figure 1 the block 'vehicle control' (VC) is the important component in the autonomous system responsible for generating the correct control outputs; in a more advanced stage it will closely cooperate with the module for 'behavior decision' (BD) basing its output on the results of a module for 'situation assessment' (SA) still under development for VaMoRs-P by our Computer Science partners at UniBwM. This architecture has proven successful with the elder VaMoRs- and VITA-systems.

Figure 2 shows the three-layered approach developed for fast and flexible control behavior several years ago at UniBwM [Dickmanns 91]. On the lowest layer, reflex-like behavioral capabilities are implemented based on state vector feedback known to be optimal for linear systems and quadratic goal functions. The 4-D approach to vision yields directly the state variables required for this type of implementation: lane keeping and convoy driving are typical behavioral capabilities realized on this level. This type of implementation yields very fast reaction times and



a) Selectable continous state feedback superimposed by event-triggered feed-forward



b) Three-layered scheme based on a)

Figure 2. Situation dependent, intelligent control with knowledge based mode switching

elimimates completely the requirement for action planning for these behaviors. Care has to be taken in order to detect situations in which these simple feedback control laws are no longer applicable. For example, when an obstacle blocks one lane or a forking of the road appears, higher levels capable of understanding the new situation or of coming up with the right decision in the mission context have to override the mode of control running.

In the case of an obstacle in the own lane, and a neighboring lane being free of obstacles, a feedforward generic control time history may be called up with a proper set of parameters which is known to steer the vehicle safely into the neighboring lane; this maneuver may be achieved by applying simple rules to a data set composed of the relative states of several other objects. The upper layer in figure 2a and the middle layer in figure 2b represent this rule based behavior triggered by special events recognized through vision. The transition from free-lane cruising to convoy driving when running up to a slower vehicle in front is another typical event-triggered behavioral component.

When a lane change is desirable, it first has to be checked whether this lane change is actually possible without endangering another car; this complex perception task can now be tackled with the capability of rearward viewing. With respect to the task of recognizing objects beside the ego-vehicle in the neighboring lane the two systems being developed jointly with our industrial partner Daimler-Benz (DB) and another university team are different. While DB in their vehicle VITA_2 rely on another two sets of six cameras each on each side of the vehicle for stereo vision [Ulmer 94] we prefer saccadic viewing direction control for two sets of cameras on a platform for active vision covering the forward and the rear hemisphere in connection with the 4-D approach in order to accumulate over time the information needed (like humans do by moving their head and eyes). Knowledge about motion processes of bodies with inertia is being exploited systematically in order to make up for the drastically reduced set of data collected. We will get along with at most 8 cameras in total (actually 4) on two platforms in VaMoRs-P while VITA_2 actually has 12 fixed ones just for looking to the sides; in total, this will lead to at least twice the number of cameras (actually 18), however, with no need for large amplitudes in panning control.

The decision for saccadic vision will affect the overall architecture of the visual perception system deepply; the advantage of a drastically reduced data stream from the vision sensors combined with the need for intelligent control of the perception system over time puts much more emphasis on temporal representations. However, this is not considered to be a disadvantage since intelligence per se needs powerful representations along the time axis which has been overlooked by the AI-approach to vision once before. The engineering sciences have sufficiently powerful tools available for handling the problems encountered.

Our expectation is that the need for temporal smoothing and both inter- and extrapolation over saccadic periods will lead to further developments of the 4-D approach considered mandatory for achieving performance levels approaching the human one in the long run.

ACTUAL REALISATION ON TRANSPUTER HARDWARE

Figure 3 shows that part of the system architecture realized in VaMoRs-P which is directly associated with visual perception; an additional transputer subsystem exists for conventional sensor data processing and actuator control. In total, the system is implemented on a network of about 5 dozen transputers about 2 dozen of which are 16-bit T-222, the rest being 32-bit T-805 waiting to be replaced by next generation components.

The Transputer Image Processing system (TIP) may be characterized roughly by the following numbers in our implementation:
The video bus data transfer rate of 67 MB/s is used to distribute the images of two cameras connected to each TIP-bus subsystem. At a cycle time of 40 ms, both the tele- and the wide-angle images of size 320 by 256 pixel are transferred to all VPUs connected. Double buffering allows for convenient image access; additional frames for drawing facilitate visualisation of the results on the color graphics displays (CGD, see lower right corner).

With the software package KRONOS each T222 is capable per video cycle of searching about 400 pixel

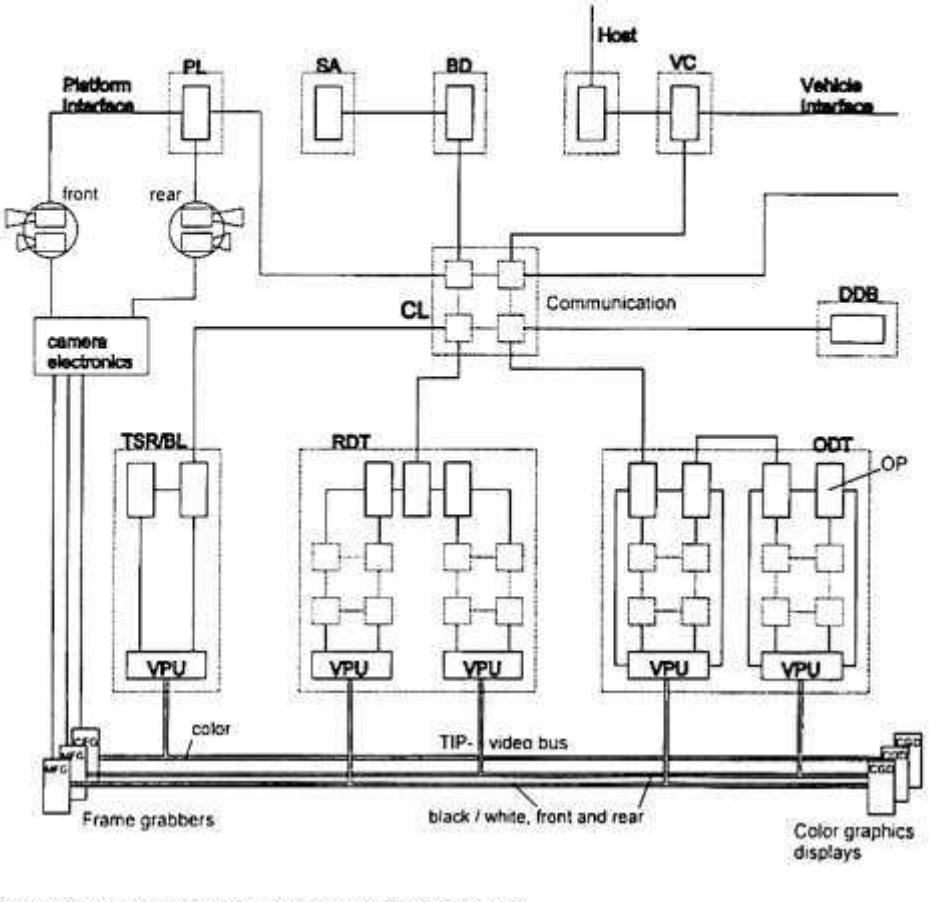


Figure 3. Transputer system architecture for visual perception and control;
⬚ = T222, ▭ = T805 (▯)

path length in intelligently assigned windows for edges, depending on the mask size; about 20 features usually result which are communicated to the object processor OP directly via links.

Object detection and tracking (ODT)

In 4 video cycles (80 ms) up to 5 objects may be tracked with two T222 and one T805. Object detection in ODT is achieved in 26 vertical search windows covering about 1/3 of the image every 80 ms with gradient masks of size 9 by 3. Up to 180 edges may result which are grouped into up to 20 contours; from these, up to 5 objects are extracted, each characterized by 18 attributes. At 12,5 Hz evaluation rate this results in 900 Bytes/second (B/s) data rate per object, i.e. a maximum of 4.5 KB/s output data rate per OP as compared to the image input data rate of 1.024 MB/s per image, yielding a data rate reduction of 228, hopefully with little loss of information with respect to objects in the scene depicted. More details are given in [Thomanek *et al.* 94].

Road detection and tracking (RDT)

With forward and, for the first time, also rearward recognition of the road, VaMoRs-P will be capable of deriving its relative position on the road and in the lane from interpolation of measurement data; in addition, because of the higher velocities envisaged the viewing distance has to be increased. About 200 m viewing range in the direction of the road are desirable and may be achieved with active trifocal vision. In the environmental range then covered (~400m) the road may no longer be modelable sufficiently well by just one or two clothoidal arcs moved along the road fixed to the vehicle center.

At high speeds on a straight stretch of road it is very essential to recognize an upcoming curve early and to be able to estimate the true radius of curvature locally correct; under these conditions the simple averaging curvature model of [Dickmanns, Mysliwetz 92] is no more sufficient. For this reason, a locally fixed (stationary) piecewise curvature model according to the rules used for construction of high-speed roads is being substituted for the simple (time-varying) sliding model used up to now; this will also simplify the handling of changing numbers of landes as they occur at entries and exits of highways. This more general model can, lateron, be easily expanded for road forkings and general road crossings.

Another extension that has become necessary with the capability of tracking several objects in each visual hemisphere in parallel and with driving in dense public traffic situations on Autobahnen is to keep track of those image areas where lane and road boundaries are no more visible due to occlusion by these objects. (More than 3000 km have been driven autonomously on public roads since 1992 with the two vans VaMoRs and VITA at speeds up to 80 km/h.)

The boundaries of these regions of occlusion may disturb the lane recognition process by simple edge feature machting; therefore, once an object has been detected and recognized, the corresponding predicted image area is blanked out for road recognition.

In addition, it is hoped that the next generation of processors will allow region based feature components for improving robustness of the recognition processes.

Details on the module RDT for road detection and tracking may be found in [Behringer 94].

CAMERA PLATFORM

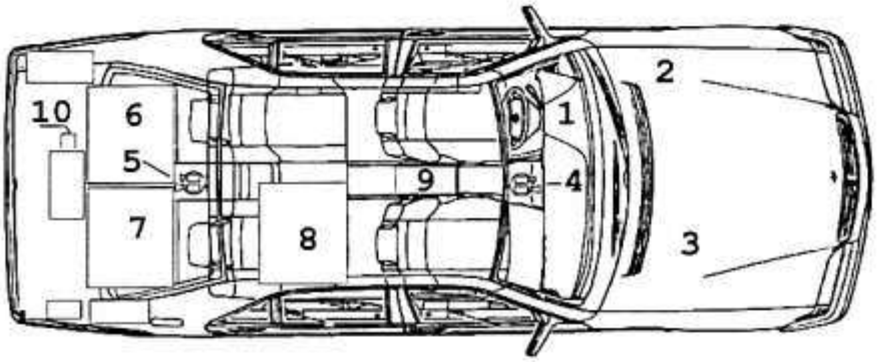
Each platform performs pan movements. It is driven direct by a brushless motor. The angle is measured by a resolver with a theoretical resolution of 0.005 degree. In addition, piezoelectric vibrating gyroscopes are used to measure the inertial angular rate. Two cameras (with 1/2" chip) are used, one with a tele lens (focal length 24 mm) and one with a wide angle lens (focal length 7.5 mm). This bifocal system has the advantage of a large field of view and, simultaneously, at least in one subarea high resolution. Brightness changes are compensated by shutter control acting like an electronic automatic iris. With recent miniature CCD cameras small size and weight of the platform is achieved.

In the normal mode of operation the wide angle camera is used to obtain reliable information about the road and obstacles nearby for vehicle control; the tele camera is used to detect obstacles further away. The goal of viewing direction control here is to keep the road in the middle of the tele image. The platform receives information on the center of the road in the tele image from the road detection modul RDT and uses a proportional controller to obtain the angular rate command for the motor controller. Due to the wide curves on highways only slow and robust tracking performance is needed.

If there are traffic signs beside the road that have to be read, this is alleviated if the gaze is directed to them. In order to avoid disturbances of the normal operations this has to be done with short and fast gaze shifts. The implementation of such a mode is given in detail in [Schiehlen *et al.* 94].

SYSTEM INTEGRATION

The installation of the components in the vehicle is shown schematically in figure 4; the trunk of the vehicle is filled with additional electronics in two 19-inch racks (6,7). In order to allow generous equipping of the vehicle with test-, recording- and program development facilities, one of the rear seats has also been taken for installing additional electronic gear (8).



1 electrical steering motor
2 electrical brake control
3 electronic throttle
4 front pointing platform for CCD-cameras
5 rear pointing platform
6 Transputer Image Processing system
7 platform and vehicle controllers
8 electronics rack, human interface
9 accelerometers (3orthogonal)
10 inertial rate sensors

Figure 4. Top view of VaMoRs-P, components for autonomous driving

The platforms carrying the cameras are mounted from the top behind the center of the front (4) and rear windshield (5); the inertial sensors are located close to the center of gravity between the two front seats (9). A PC with interface to the transputer system serves as host systems and man/machine interface.

System integration has been performed in summer 1994 in parallel to the companion system VITA_2 of Daimler-Benz. The latter one is the official Common European Demonstrator vehicle 'Obstacle Avoidance' (CED-3) for the final Demo of the EUREKA-project 'Prometheus' in October 1994.

CONCLUSIONS

With **VaMoRs-P** of UniBwM and the twin vehicle **VITA_2** of Daimler-Benz new test vehicles of advanced performance levels with slightly different perception systems are available which allow to further explore the promising new technology of dynamic machine vision for autonomous road vehicle guidance. The systems are brand new and have, by far, not yet been tested to their performance limits.

The new generation of microprocessors becoming available over the next months will allow to further increase system capabilities over the next years while simultaneously shrinking both volume and power consumption of the system. Both the perception subsystems and the overall system architecture are becoming more and more stable and powerful, which is a good indication of the systems becoming more mature. However, it should not be overlooked that in order to achieve robust performance approaching the performance level expected from human drivers there is still a long way to go.

This research has been supported
by BMFT and Daimler-Benz AG
under the EUREKA-Projekt
Prometheus III
(Grant ITW 9301A).

LITERATURE

Behringer, R.: 'Road Recognition from Multifocal Vision', (this volume) 1994

Dickmanns, E.D.: '4-D Dynamic Vision for Intelligent Motion Control'. In C. Harris (ed.): *Special issue of the Int. Journal for Engineering Applications of AI (IJEAAI) on 'Intelligent Autonomous Vehicles Research'*, 1994.

Dickmanns, E.D.; Mysliwetz, B.: 'Recursive 3D Road and Relative Ego-State Recognition'. *IEEE-Trans. PAMI, Special Issue on 'Interpretation of 3D Scenes'*, Febr. 1992

Dickmanns, E.D.: 'Machine Perception Exploiting High-Level Spatio-Temporal Models'. *AGARD-Lecture Series 185*, Hampton, USA; Munich; Madrid (available through AGARD distribution centres as chapter of AGARD-LS-185 'Machine Perception'), 1992

Schiehlen, J.; Dickmanns, E.D.: 'A Camera Platform for Intelligent Vehicles', (this volume) 1994

Thomanek, F.; Dickmanns, E.D.: 'Multiple Object Recognition and Scene Interpretation for Autonomous Road Vehicle Guidance', (this volume) 1994

Ulmer, B: 'VITA II - Active Collsion Avoidance in Real Traiffic' (this volume) 1994