EXPECTATION-BASED MULTI-FOCAL VISION FOR VEHICLE GUIDANCE

by

Ernst D. Dickmanns
Universitaet der Bundeswehr Muenchen
D-85577 Neubiberg, Germany
Tel.: +89 6004 2077/3583; Fax: +89 6004 2082;
e-mail: Ernst.Dickmanns@unibw-muenchen.de

ABSTRACT
Based on experience with several generations of vision systems for road vehicle guidance a new complex vehicle eye and corresponding control schemes for viewing direction control and feature extraction are proposed allowing new levels of performance with state of the art general purpose processors. Modeling along the time axis is the key to an efficient use of the degrees of freedom gained by saccadic viewing strategies.

1 INTRODUCTION

Over the last decade, research in visual road vehicle guidance has made considerable progress worldwide. In Europe, the EUREKA-project 'Prometheus' has advanced the state of the art of driving on high-speed roads while in the US the (D)ARPA projects on Autonomous Land Vehicles (ALV) and on Unmanned Ground Vehicles (UGV) did concentrate on driving on lower order roads and cross-country. In Japan (late 80ies) and Korea (early 90ies) activities have been picked up with attention focusing on well marked roads.

A good survey on the state of the art may be seen from the conference proceedings of IEEE-'Robotics and Automation', 'SPIE: Mobile Robots' (both since around 1986) and the International Symposia on 'Intelligent Vehicles' [Masaki 92 to 95].

Today, almost any larger car manufacturing company has some activity running in the field of visual road vehicle guidance. Many university groups around the globe are active in this field, very often in association with some car manufacturing company. This joint effort in conjunction with the rapid development of microprocessor technology has led to fast progress; within just one decade the field evolved from essentially non-existence to demonstrated driving capabilities in normal public traffic (under favorable conditions, of course) at normal driving speeds of more than 100 km/h.

However, viewed from a price/performance and general reliability point of view, there is still a long way to go until systems will be on the market. What has been shown is the general feasibility of the approach and the flexibility obtainable in traffic automation. Legal aspects have barely been touched.

2 STATE OF THE ART AT UBM

The Universitaet der Bundeswehr Munich (UBM) has been one of the institutions pioneering 'vehicles with the sense of vision' and dynamic machine vision in general with the test vehicle **VaMoRs** since 1985/6. In 1987, driving fully autonomously (with both longitudinal and lateral control achieved by machine vision) at speeds of close to 100 km/h has been demonstrated on an empty stretch of Autobahn not yet turned over to the public.

In 1988 first results in obstacle recognition and automatic stops in front of a stationary obstacle from speeds up to about 40 km/h have been shown. At the demonstration in connection with the board member meeting of the Prometheus project at Turin in 1991, our industrial partner Daimler-Benz showed the capability of convoy driving behind another vehicle and of 'Stop-and-Go' performance (as needed in a traffic jam) with its test vehicle VITA (a 7.5-ton van) equipped with the UBM vision system.

After this event the switch to transputers as processing elements has been completed which allowed to decrease the volume of the parallel processing system to about one third, increasing at the same time the computing power available. A major change in performance level resulted from simultaneous transition to 'bifocal' vision in which the signals of two cameras with different focal lengths, mounted fixed relative to each other on a pointing platform, are evaluated in conjunction in order to allow both a large viewing angle nearby and good resolution farther away. This arrangement of cameras has been in use for quite some time, however, with separate usage of the video signals for separate purposes: road recognition has been achieved with the wide angle camera while obstacle detection and relative state estimation has been done with the tele-signal.

By combining image evaluation of both cameras for both road recognition and obstacle detection and tracking, with the new computing power available the performance level could be raised considerably.

2.1 Perception system (state 1995)

In the van **VaMoRs** a two-axis pointing system with high bandwidth and inertial stabilization in pitch is being used; saccades of 20 degrees amplitude may be performed in about a tenth of a second. In the test vehicle **VaMP** two miniature single axis platforms are installed; the focal lengths chosen are 7.5 and 24 mm. The front unit is almost hidden behind the rearward looking mirror in the center of the front windshield. Only heading direction is controlled in order to facilitate a large horizontal viewing range.

Both vehicles do have two sets of inertial sensors: three orthogonally mounted accelerometers are located close to the center of gravity while the corresponding set of rate gyros (piezo-electric rate sensors respectively) may be positioned anywhere convenient.

A wheel position encoder on one of the wheels acts as odometer and speedometer by proper signal processing. The steering angle is also measured by a digital position encoder.

In **VaMoRs** a GPS-receiver (for the US-Global Positioning System based on a set of satellites) has been installed for providing coarse initial guesses for the position of the vehicle on the globe.

The transputers as processing elements are now (1996) being replaced by PowerPC's (MPC's) providing about ten times the computing performance (depending on the special task, of course) [Mau 96]. Computing cycle times depend on the signals and their use. Rate information for the stabilization of the viewing direction is processed at 500 Hz (2 ms cycle time) while image processing has been done mainly at 12.5 Hz (80 ms) in the transputer system and at 25 Hz (40 ms) in the MPC-system. Higher level system components like mentally keeping track of vehicles to the side may run as slow as 2 Hz (500 ms).

The number of processors in the new MPC-system is now down to about one dozen (plus some transputers kept for tasks where their performance is sufficient).

1. Obstacle detection, tracking and relative state estimation has been extended to up to six objects (three in each video data stream) in a viewing range of up to 120 m [Tho 94].

2. Detection of cross-roads, perception of the road junction parameters by cross-road fixation and active viewing direction control while approaching, and turning off onto the cross-road has become possible [Dic95b].

3. In the new high-speed test vehicles VaMoRs-P, in short **VaMP** of UBM, and **VITA_2** of our industrial partner Daimler-Benz, two passenger cars 500- SEL, a rearward looking bifocal set of miniature TV-cameras has been installed in addition to the front set in order to be able to monitor the traffic around the vehicle in plus/minus 120 m range.

This allowed to track up to 12 other vehicles around the own one in parallel. By keeping mentally track of passing vehicles leaving the rearward field of view until they reappeared in the front field, also the number of vehicles invisible to the side could be kept track of (on a two-lane road only, of course).

In October 1994, for the final demonstration of the Prometheus project on the Autoroute A1 north of Paris, the two S-class test vehicles **VITA_2** and **VaMP** were the only ones to drive fully autonomously with machine vision in normal dense three-lane traffic with speeds up to 130 km/h. Convoy-driving at all speeds with distance to the vehicle in front as a function of the speed actually driven has been extensively demonstrated with guests on board.

For the first time ever, the vehicle also gave a notice when it thought it would be safe to perform a lane change for going at higher speed; this included that no other vehicle was approaching from the rear with high excess speed, that there was no vehicle to the side, and that the neighboring lane was free in a certain look-ahead range in order to pass the vehicle in front. The safety driver, always in charge of monitoring safety aspects, had to give a sign of acceptance before the vehicle was allowed to really perform the lane change maneuver autonomously.

This system has been realized in **VaMP** with the aid of about 5 dozen transputers in total; about 40% of these were 16-bit processors devoted to image feature extraction and communication. The rest were T-805's with a floating point unit on the chip; the structure of the processor system has been discussed in [Maurer et al. 94]. The Daimler-Benz test vehicle **VITA_2** had 14 more CCD-cameras on board and used additional computing power from a set of C-40's [Ulm 94].

2.2 Perceptual capabilities

The most essential facts about the environment on a highway which have to be perceived precisely are the road, other vehicles on the road and possibly cross-roads for turn-off:

2.2.1 Road recognition

This has to be performed in conjunction with the estimation of the ego-state relative to the road since the mapping conditions of road and lane boundaries depend on the position of the vehicle on the road. It is based on a viewing range of up to 100 m; several lanes may be tracked in parallel [Beh 94].

The following state variables and road parameters are being estimated 25 times a second: 1. Horizontal curvature parameters c_{h0} and c_{h1} of the skeletal line of the road (lane); 2. vertical curvature parameters c_{v0} and c_{v1}; 3. lane width and number of lanes in the look-ahead range. 4. Relative vehicle state: lateral offset from center (lane or road); lateral speed (or heading angle rel. to the lane), angular rate of vehicle around vertical axis, slip angle and steering angle. It has turned out to be advantageous to also have the vehicle pitching angle estimated; this is being done with a high frequency component and a low frequency component (for bias estimation) separately.

2.2.2 Other vehicles

The relative state of up to ten (12) other objects on the road is continuously being estimated with the following variables: 1. Range, range rate and range acceleration, 2. lateral position in the lane, lateral speed and 3. bearing as well as 4. the width of the object.

2.2.3 Cross-roads

For cross-road recognition both the intersection angle and the width of the cross-road are being estimated (**VaMoRs** only).

2.3 Control system and behavioral capabilities

Both systems use the steer angle, the throttle and the braking system as controls. Reflex-like behavior in response to perturbations is achieved through state variable feedback with proper control laws depending on the situation and driving speed.

Triggered by certain events, feed-forward control time histories may be activated as for performing a lane change when running up to a slower vehicle in front or for making a turn-off onto a cross-road.

Certain parameters in the overall control system may also be set depending on environmental parameters determined: If the lateral acceleration shall be limited (to say 0.2 Earth-gravity, ~ 2 m/s^2), from the horizontal curvature the corresponding speed maximally allowed can be computed and set as one of the behavioral parameters; higher values of this parameter lead to a more sportive driving style.

Convoy driving at a speed-dependent distance behind another vehicle has become a standard behavior by now [Brue 94]. In this case the determining factor turning speed into 'relative distance to be maintained' is an adjustable factor up to now set by the human operator.

2.4 Deficiencies remaining

Even though the feasibility of autonomous vehicle guidance by dynamic vision has been demonstrated, there are many points left open for improvement in performance. Above all, robustness under adverse lighting conditions has to be improved; area based image processing, in addition to the edge-based approach favored up to now in order to achieve real-time capabilities, has to be added taking color and texture information into account. The computing power for this is becoming available now.

This would also allow to recognize the color coding in white and yellow lines at construction sites where the system has difficulties right now; most of the time when the human safety operator has to take over in normal Autobahn traffic is at construction sites and when passing through tunnels.

Picking up passing vehicles early and obtaining good relative state estimation soon is another problem area. By increasing the simultaneous field of view by more than a factor of two (to about 120 to 135 degrees) would be desirable; the additional capability of binocular stereo in the near range would allow to achieve good distance estimates from single features.

For responsible driving at high speeds the viewing range has to be increased to several hundred meters. In order not to be flooded with data, this high resolution capability may be confined to a small area which has to be directed corresponding to the region of special interest.

Up to now it has been assumed that the ground to be driven upon is smooth and only weakly curved; a general vision system should be capable of verifying this assumption and of determining the vertical surface structure in case it should not be flat. Work in this direction based on trinocular stereo has been started [Bat 95].

3. EMS-VISION

In order to remove most, if not all, of the deficiencies mentioned above, a new - third generation - dynamic vision system called 'Expectation-based, Multi-focal, Saccadic' (EMS) vision system has been designed and is being realized. It encompasses a compound technical camera arrangement with up to four separate imaging sensors on a high bandwidth pointing platform called 'MarVEye' [Dic 95].

The simultaneous field of view covered by two wide-angle cameras is around 130 degrees in the horizontal direction, so that it is possible to look both forward and sideways at the same time; there is a central region of overlap covered additionally by a high-resolution color camera with a mild tele-lens which allows for bifocal trinocular divergent stereo in an angular range of around 15 degrees [Rie 96]. An optional fourth imaging sensor with a strong tele-lens provides the capability of very high resolution in a small field of view; this camera may be black-and-white with high sensitivity for lower light levels.

By exploiting viewing direction control and data management along the time axis as developed with the previous systems [Sch 95], the equivalent of vertebrate vision may be achievable. The 4-D approach as refined in [DDi 96] lends itself to this type of dynamic vision since it generalizes the method by taking full advantage of object-orientation.

4. CONCLUSIONS

Experience with driving in normal freeway environments has led to the definition of a new compound technical eye with a multi-focal camera arrangement and vertebrate-like viewing strategies which, hopefully, in conjunction with the new generation of processors and the refined 4-D approach to dynamic machine vision will eliminate most - if not all - of the problem areas encountered.

This new system is called 'Expectation-based, Multi-focal Saccadic' vision (in short: EMS-vision).

5. REFERENCES

[Bat 95] Baten S, Mueller N: Navigation with an autonomous vehicle on arbitrary roads. In 3rd International Workshop High Precision, (Stuttgart), April 1995.

[Beh 94] Behringer R: Road recognition from multifocal vision. Proceedings of Int. Symp. on 'Intelligent Vehicles', (Paris), Oct. 1994.

[Brue 94] Bruedigam C: Intelligente Fahrmanoever sehender autonomer Fahrzeuge in autobahnaehnlicher Umgebung. PhD thesis, UBM, Institut fuer Systemdynamik und Flugmechanik, 1994.

[Dic 95a] Dickmanns ED: Road vehicle eyes for high precision navigation. In 3rd International Workshop High Precision, (Stuttgart), April 1995.

[Dic 95b] Dickmanns ED, Mueller N: Scene recognition and landmark navigation for road vehicle guidance. In 2nd IFAC Conference on Intelligent Autonomous Vehicles, (Helsinki), June 1995.

[DDi 95] Dickmanns D: Knowledge based real-time vision. In 2nd IFAC Conference on Intelligent Autonomous Vehicles, (Helsinki), June 1995.

[Mas 92 - 95] Masaki I: Proceedings of Int. Symp. on 'Intelligent Vehicles': Detroit 1992, Tokyo 1993, Paris 1994, Detroit 1995

[Mau 94] Maurer M, Behringer B, Dickmanns D., Hildebrandt T, Thomanek F, Schiehlen J, and Dickmanns ED: VaMoRs-P An advanced platform for visual autonomous road vehicle guidance, in Mobile Robots IX, (Boston), November 1994.

[Mau 96] Maurer M, Behringer R, Fuerst S, Thomanek F, Dickmanns ED: A compact vision system for road vehicle guidance. Submitted to 13th Int. Conf. on Pattern Recognition, Wien, August, 1996.

[Rie 96] Rieder A: Trinocular divergent stereo vision. Submitted to 13th Int. Conf. on Pattern Recognition, Wien, August, 1996.

[Sch 95] Schielen J: Kameraplattformen fuer aktiv sehende Fahrzeuge, PhD thesis, UBM, Institut fuer Systemdynamik und Flugmechanik, 1995.

[Tho 94] Thomanek F, Dickmanns ED, and Dickmanns D: Autonomous road vehicle guidance in normal traffic. Proceedings of Int. Symp. on 'Intelligent Vehicles', Paris, Oct. 1994.