EMS-Vision:
Recognition of Intersections on Unmarked Road Networks

M. Lützeler
and *E. D. Dickmanns*

Institut für Systemdynamik und Flugmechanik,
Universität der Bundeswehr München,
D-85577 Neubiberg, Germany

Michael.Luetzeler@unibw-muenchen.de

Abstract

The ability to recognize intersections enables an autonomous vehicle to navigate on road networks for performing complex missions. The paper gives the geometry model for intersections applied and their interaction with active viewing direction control. Quality measures indicate to performance monitoring processes the reliability of the estimation results. The perception module is integrated in the EMS-Vision system. Results from autonomous turn-off maneuvers, conducted on unmarked campus roads are discussed.

Keywords road recognition, dynamic machine vision, autonomous vehicles, road navigation

1 Introduction

Recognizing and handling of intersections is a key ability for an autonomous agent in order to successfully navigate on road networks. This has been initially investigated at Universität der Bundeswehr München (UBM) by N. Müller, see [1] and [2]. His approach was validated by hardware in the loop simulation and autonomous turn-off maneuvers on marked campus roads. Exploiting images from two monochrome cameras, intersections were detected and tracked utilizing an active pan-tilt head (TACC) to direct the focus of attention. Building on these results, vision based intersection navigation has been integrated in the current Expectation-based Multi-focal Saccadic Vision system (EMS-Vision), see [3], [4].

2 Integration in the perceptual framework of EMS-Vision

Perception modules use sensor input to estimate state variables describing relative positions and geometry parameters for objects in 3D space and time. In the EMS-framework, perception modules are implemented as processes, each with areas of expertise. At startup, a process announces via its process node in the knowledge base which object classes it can handle; for an overview of knowledge representation see [3]. An attention control module assigns expected objects, e. g. roads and landmarks, and detection areas, e.g. for vehicles ahead, to perception experts. Obsolete objects are removed from the task list of the perception module, for example after a landmark has left the field of view. For object control-flow see [5].

Figure 1 shows the knowledge represented in the EMS-system to perform a locomotion mission including a turn-off. The right hand column shows the mission plan, decomposed into mission elements, e. g. `Follow Road` and `Turn Left`. For details on mission planning and monitoring see [5] . Mission elements contain expected objects relevant for the current task. These objects are added to the scene tree (second column), the central internal representation for physical objects and their relative positions. Poses are in their most general form described by six degrees of freedom (`6DOF`). The left two columns show the software processes, here two perception modules, and the processing nodes they are active on. The perception modules make use of the camera signals available on the nodes they reside on. The transformation between the cameras and the frame-grabber node (`FG`) is perspective projection and some shift (in general). The `FG` nodes contain pointers to the digitized video images and the information, on which computer the stream is captured. All transformations in the scene tree are described by homogeneous coordinate transformations (HCT). The `Ego` node comprises the ego-vehicle and its sub-objects; `Distant Road` and `Local Road` nodes represent the road currently traveled on. Additionally to the objects needed for the current mission el-

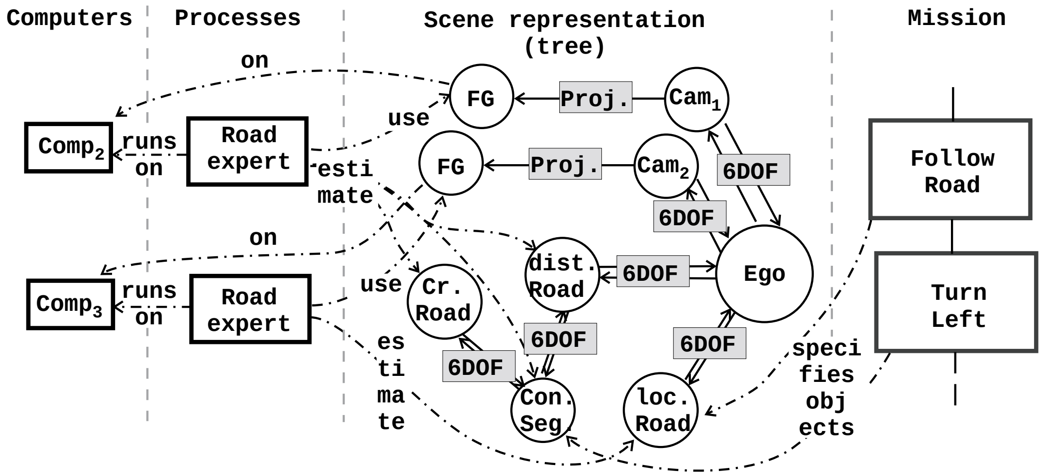

Figure 1: EMS-Vision data bases: navigating turn-offs

ement, the scene tree may contain objects required by the following elements. In the turn-off example the cross-road nodes `Connect Segment` and `Cross-road` must be detected while the `Follow Road` element is still active.

3 Generic Model for the Geometry of Intersections

In this section an overview is given of the coordinate systems, which are used for describing the position of road segments relative to the vehicle. Subsequently, the geometry model for road segments and intersections are explained. This leads to a description how intersections form fixpoints for the road segment models.

3.1 Coordinate Systems

Building on the calling conventions introduced in [6], coordinate systems for the cross-road, index "$_{cr}$", and the connecting segment, "$_{cs}$" are added. In an ego-centered world description the body fixed coordinate system of the vehicle has its origin in the center of gravity (cg) with the x_f and z_f axes lying in the plane of symmetry of the vehicle. The "surface-oriented" coordinate system has its origin below the cg and its x_s-direction parallel to the local tangent of the road skeleton line; its $x_s - y_s$ plane is parallel to the local road surface. It thus holds the yaw angle, ψ_{s_cg}, between road and vehicle. The position of the cg relative to the road is described by a lateral offset (y_{r_s}) from the "surface" coordinate system. The "connect segment", describing the intersection of two road segments, has a longitudinal offset, x_{cs_r}, to the road-base coordinate system. It decreases as the vehicle approaches the intersection. The "cross road" is linked to the "connect segment" by the turn-off angle ψ_{cr_cs}.

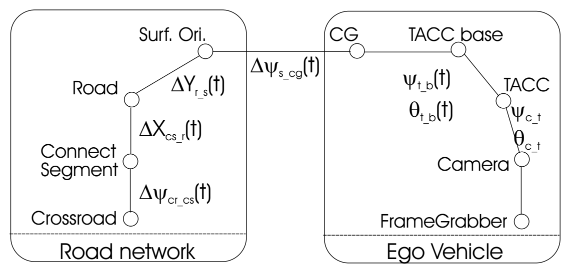

Figure 2: Scene tree for road representation: Dynamic position parameters

The scene tree for this road network is given in figure 2. Parts of the mapping geometry within the ego-vehicle are shown on the right hand side, cg, TACC-base, TACC, an example camera and a framegrabber, representing perspective mapping. The left hand side shows the road network nodes with the surface oriented system, the actual road, the connect segment and the cross-road. Dynamic variables in this scene tree are, besides the relative position vehicle to road, the TACC pan- and tilt-angles, ψ_{t_b}, θ_{t_b}.

3.2 Road Segment Geometry

The generic road model consists of a center-line and the road width perpendicular to this curve. The center-line is given by a horizontal clothoid, [7], (c_{0h}, c_{1h}). Clothoids are the trajectories wheeled vehicles follow, given constant steer-rate at constant speed. Special cases are straight lines ($c_{0h} = c_{1h} = 0$) and circular paths ($c_{0h} \neq 0$ and $c_{1h} = 0$). Allowing a linear change of road width along the path (b_1) results in a lookahead dependent formulation of the actual width. This differential geometric description results in a very compact parameterization, $[c_{0h}, c_{1h}, b_0, b_1, L]^T$, with L

as the maximum lookahead range.

A birds-eye view of an example clothoidal band is given in figure 3. $P_r(l)$ denotes a point on the centerline, $\chi(l)$ the respective heading angle.

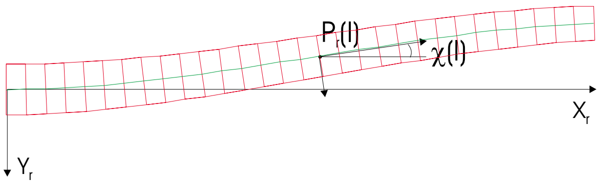

Figure 3: Road model with $c_{0h} = -0.015\text{m}^{-1}$, $c_{1h} = 0.0007\text{m}^{-2}$, $b_0 = 3.5\text{m}$, $b_1 = -0.03$ and $L = 40\text{m}$

Additionally for the near lookahead range, where segment curvature is negligible, a "local road" model is available, consisting of straight lines.

3.3 Connect Segment Geometry

The connecting segment is defined by the intersection of two road segments, e.g. the own road and a cross-road. The intercept point of the skeleton lines of the own road and the cross-road is the origin of the connecting segment. Its x_{cs}-axis is oriented parallel to the own road, the z_{cs}-axis perpendicular to the local surface and the y_{cs}-axis completes a right hand system. The geometry of the connecting segment is described by the road width of the two branching segments, their relative positions, the turn-off angle ψ_{cr_cs} and the inner radius of the curve r_{cr_cs} as geometry parameters. For an example intersection geometry see fig. 4.

Using the intersection geometry parameters as given above, the distance of the end-point of the own road P_{e_cs} to the connecting segment is given by:

$$d_{gr\_cs} = \frac{1}{\sin(\psi_{cr\_cs})} \cdot (r_{cr\_cs} + \frac{b_{cr}}{2}) - \frac{1}{\tan(\psi_{cr\_cs})} \cdot (r_{cr\_cs} + \frac{b_r}{2}), \quad (1)$$

which can easily be deduced from

$$d_{gr\_cs} = \tan(\beta) \cdot (r_{cr\_cs} + \frac{b_r}{2}) \quad (2)$$

where $\tan(\beta)$ is given by the joint hypotenuse of the two right-angled triangles α and β and $\psi_{cr\_cs} = \alpha + \beta$.

In equation 1, b_{cr} and b_r denote the respective road width and ψ_{cr_cs} the turn-off angle. The coordinates of the end point are d_{gr_cs}:

$$P_{e\_cs} = \begin{bmatrix} x \\ y \\ z \end{bmatrix} = |d_{gr\_cs}| \cdot \begin{bmatrix} -1 \\ 0 \\ 0 \end{bmatrix} \quad (3)$$

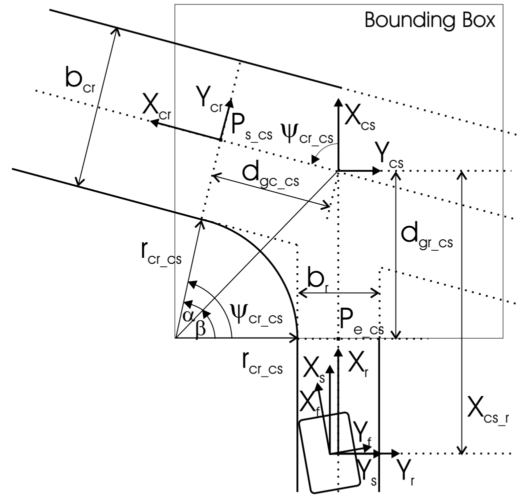

Figure 4: Intersection geometry, with $\psi_{cr\_cs} = -75°$ and $b_{cr} \approx 2 \cdot b_r$

Similarly the base point of the forking road, P_{s_cs}, is given by:

$$P_{s\_cs} = |d_{gc\_cs}| \cdot \begin{bmatrix} \cos(\psi_{cr\_cs}) \\ \sin(\psi_{cr\_cs}) \\ 0 \end{bmatrix} \quad (4)$$

A bounding box, centered around the origin of the connect segment, is used as rough approximation for the intersection's geometry. Its size is given by $2 \cdot d_{gr\_cs}$ and $2 \cdot (r_{cr\_cs} + \frac{b_r}{2})$.

Commonly road segments extend beyond the lookahead distance covered by the sensor system. Current information updates can only be acquired about the portion within this lookahead distance. This fact leads to a road representation moving with the vehicle cg. Intersections are a case where the lookahead distance can reach beyond the length of a road segment, e.g. at a T-junction. Besides the definite end of a road segment at a T-junction, the continuity assumptions underlying a road description with a single set of parameters can be violated at a crossing.

For these reasons intersections represent break points in road geometry description. If the centerline of a road segment reaches into the area occupied by an intersection, the lookahead distance has to be adjusted appropriately.

3.4 Fixation point

An approach to active viewing direction control is that each object to be perceived specifies a fixation point in 3D. Adding optimal viewing direction to the properties of an object gives the perception expert the ability to dynamically control viewing direction for performing his perception task.

If the vision system comprises multiple cameras, the desired camera must be specified as well. The actual viewing direction can then be computed by the "behaviour generation for gaze and attention" process in an optimal manner for all perception tasks, see [8].

The fixation point for a road segment is determined dependent on the distance this camera has to the origin of the segment (d_{c_r}). The fixation point is given on the centerline of the road at an arclength (l_{fp}).

$$l_{fp} = \frac{L}{2} \cdot \begin{cases} 0 & : \quad d_{c\_r} > d_{max} \\ \frac{d_{max}-d_{c\_r}}{d_{max}-d_{min}} & : \quad d_{min} < d_{c\_r} < d_{max} \\ 1 & : \quad d_{c\_r} < d_{min} \end{cases} \tag{5}$$

With the parameters set to $d_{max} = 35$m and $d_{min} = 1.5$m.

4 The Vision System for Road and Intersection Detection and Tracking

The MARVEYE camera configuration (figure 5) consists of up to four cameras. A monochrome pair of cameras is equipped with wide-angle lenses, 4.8*mm* or 6*mm*. Their optical axes lie in one plane and may diverge by an angle of $\sim 40°$ for a large field of view ($> 100°$) with a small overlapping stereo region, or be aligned in parallel allowing full conventional stereo processing, see [9]. A 3-chip color camera is centered between these two cameras and has a pitch offset of 10° to the wide-angle cameras. This allows to pitch the camera arrangement 17° towards the ground plane and to have 10 % (3-chip) and 12 % (monochrome) of the image above the horizon.

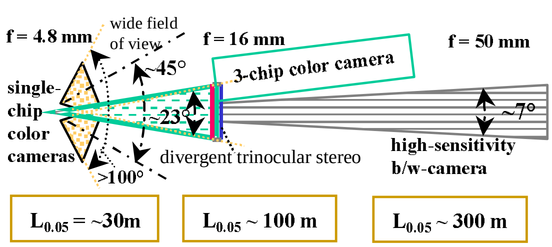

Figure 5: MARVEYE camera configuration

The images from the wide-angle cameras are exploited to update the parameters of a local road model. For a lookahead distance of up to 15m a straight line approximation for the road geometry is commonly sufficient. The yaw angle ψ_{s_cg}, and the horizontal offset, y_{r_s}, from the road centerline are taken from this local representation, see fig. 2. The color camera is used for road curvature estimation and cross-road detection.

Figure 6 shows three images taken at a single timestep from the two wide-angle and the mild tele camera. The TACC is panning $\sim 5°$ to the left to facilitate detection of a cross-road to make a left turn. The local road segment extends 13m from the vehicle cg. The cross-road is tracked at a distance of 52m.

5 Measurement of performance quality

State variables describing relative positions and geometry parameters represent one level of information for an intelligent autonomous agent. Additionally, the system requires information on how secure the perception modules are about the states supplied, see [8] for attention control issues and [10] for requirements for locomotion. For state estimation the well known technique of Extended Kalman Filtering is applied. The quality of estimation for each state is given by the variance, taken from the covariance matrices P of the filters, see Biermann [11].

Besides this information about the state of the filter, reliability information on the sensor input is required. This indicator of sensor performance is dependent of the sensor type. Measurement reliability can be static, e.g. noise level for angular rate sensors, or dynamic, e.g. performance of image processing techniques. Two key quality indicators for model based image processing are the ratio of expected to matched features and the sum of the absolute values of the differences between expected feature positions in the image and measured positions (residuals). With the index of expected features i ranging from $1 \ldots m$ and the index of matched features j within $[1,n]$ the matched feature ratio is:

$$feat_{rat} = \frac{n}{m} \leq 1 \tag{6}$$

Using the predicted feature position in 3-D camera-coordinates $Q^*(j)$ and projecting it onto the image plane gives $P^*(j)$. With the extracted position $P(j)$, the average residual is given by:

$$res_{sum} = \frac{1}{n} \sum_{j=1}^{n} |P^*(j) - P(j)| \tag{7}$$

Low values of $feat_{rat}$ (eq. 6) indicate poor image quality, unsuitable parameterization of feature extraction or large differences between model assumptions and scene observed. Large values of the residual sum (eq. 7) hint at model errors in geometry or system dynamics or at mismatches between features.

6 Experimental results

Autonomous turn-off maneuvers were conducted with the experimental vehicle VAMORS on unmarked campus roads. The intersection recognition was fully integrated

Figure 6: Intersection in MARVEYE camera configuration

into the mission context, controlling the vehicle's locomotion and perception tasks. The viewing direction of the active pan-tilt head was controlled dynamically by the gaze control process, based on fixation points specified by each object.

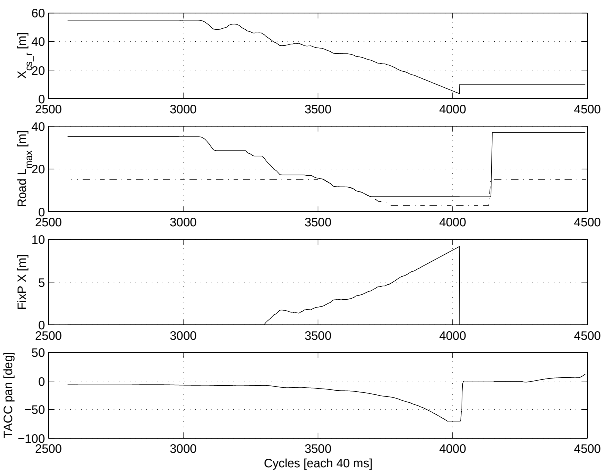

Figure 7: Approaching a T-junction (from top to bottom): a) estimated distance, b) lookahead for road and local-road segments, c) fixation point on cross-road d) resulting TACC pan-angle.

Figures 7 and 8 shows data plots while approaching a T-junction. The top-most plot shows x_{cs_r}, the distance of the connect segment to the own vehicle. The estimated value decreases as the vehicle approaches the intersection. At cycle 4025 mission monitoring removes all road objects and reinserts a new own-road with an initial value for the relative position taken from the last values for the cross-road. The second row gives the lookahead, L_{max} range for the `Own Road` and the `Local Road` segments (dashed); they are trimmed according to the rules given in section 3.3. The third and fourth subplot show the fixation point as it moves along the centerline of the cross-road segment and the resulting pan-angle of the gaze control platform, ranging from $-7°$ to $-70°$.

Figure 8 shows the quality measures for the same experiment. In the initial phase of the approach, when the cross-road is still too small in the image to be tracked robustly, no estimation is conducted and the variance, depicted in the second subplot from the top is not valid. During cross-road tracking the variance decreases. In the final phase, the position is updated by prediction only, resulting in a cubic increase in variance. Image processing in the initial phase is characterized by occasional total loss of matched features and a high residual sum. In the tracking phase, the matched feature ratio remains above 75%.

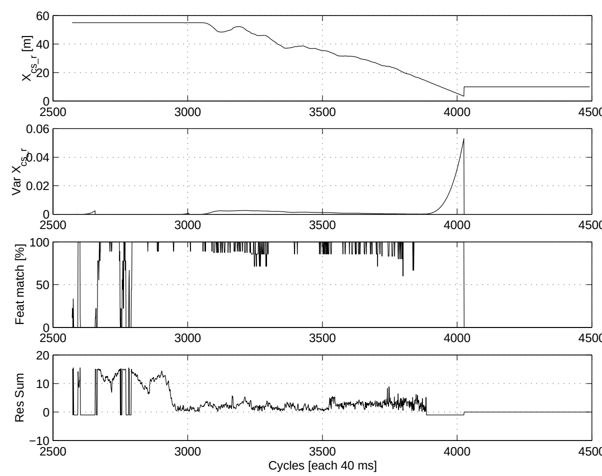

Figure 8: Quality measures at a T-junction (from top to bottom): a) estimated distance, b) distance variance c) matched feature ratio, d) residual sum

Experiments for the detection of intersections have been conducted using the mild-tele and the right wide-angle camera of the MARVEYE configuration with conventional stereo alignment of the wide-angle cameras. The top-most row, labeled (a), in figure 9 shows the initial phase of the approach, the left column contains images from the mild-tele, the right column from the right wide-angle camera.

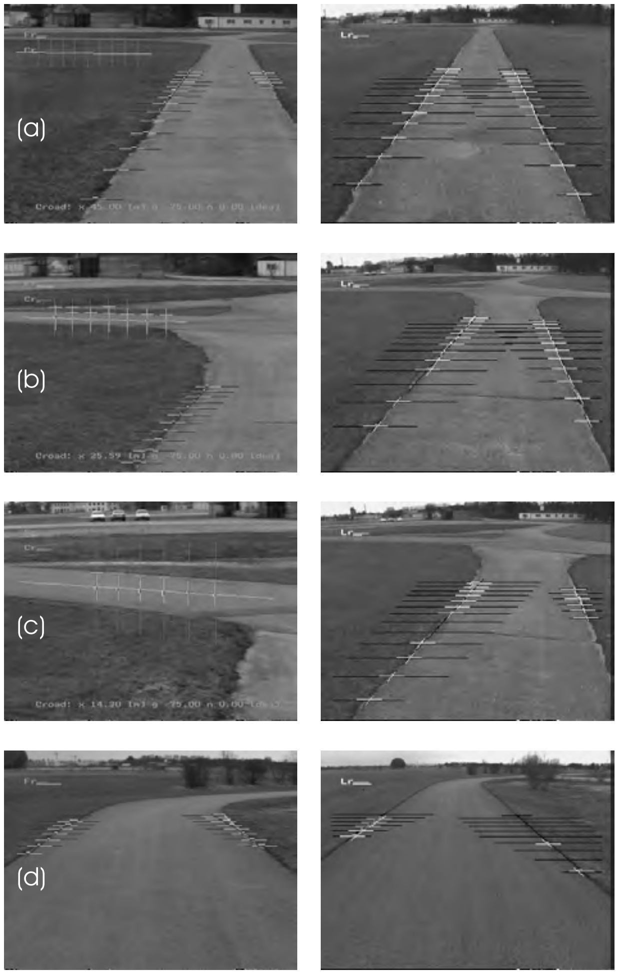

Figure 9: Approaching an intersection, mild-tele camera (left) and right wide-angle camera (right); (d) after turn-off onto cross-road

Vertical search windows are applied at the expected position of the cross-road in the image. The `Local Road` segment is tracked in the wide-angle camera. The pan-tilt head is panning $\sim 5°$ to the left, facilitating cross-road detection; approximately $2/3$ of the image at the expected cross-road position is devoted to the left of the own-road. This focusing has little effect on road detection near by, ((b) & (c)). The shortening of the lookahead range can be seen in the images of row (c), where no feature extraction windows are positioned on the inner radius of the curve. The last pair of images shows the new road being tracked after the turn-off maneuver has been completed.

7 Conclusions

Building on previous work, recognition of intersections has been integrated into the EMS-Vision framework currently under developement at UBM. Intersections represent discontinuities for road segments, resulting in a locally fixed discription of road segments at intersections. Active viewing direction control is achieved by dynamically specifying fixation points. The aim of future work is to navigate on unpaved road networks with intersections with multiple branches.

References

[1] N. Müller. Feedforward control for curve steering for an autonomous road vehicle. In *Proc. IEEE Int. Conference on Robotics and Automation*, Nice, May 1992.

[2] N. Müller and S. Baten. Image processing based navigation with an autonomous car. In *Proc. Int. Conf. on Intelligent Autonomous Systems*, Karlsruhe, March 1995.

[3] R. Gregor, M. Lützeler, M. Pellkofer, K. H. Siedersberger, and E. D. Dickmanns. EMS-Vision: A perceptual system for autonomous vehicles. In this volume.

[4] E. D. Dickmanns and H. J. Wünsche. Dynamic vision for perception and control of motion. In B. Jaehne, H. Haußenecker, and P. Geißler, editors, *Handbook on Computer Vision and Applications*, volume 3. Academic Press, 1999.

[5] R. Gregor and E. D. Dickmanns. EMS-Vision: Mission performance on road networks. In this volume.

[6] M. Lützeler and E. D. Dickmanns. Robust road recognition with MarVEye. In Ichiro Masaki, editor, *IEEE conference on Intelligent Vehicles*, Stuttgart, Germany, october 1998. IEEE Industrial Electronics Society.

[7] E. D. Dickmanns. Dynamic computer vision for mobile robot control. In *Proc. 19th Int. Symp. and Expos. on Robots*, Sydney (Australia), November 1988.

[8] M. Pellkofer and E. D. Dickmanns. EMS-Vision: Gaze control in autonomous vehicles. In this volume.

[9] S. Baten, R. Mandelbaum, M. Lützeler, P. Burt, and E. D. Dickmanns. Techniques for autonomous, off-road navigation. *IEEE Intelligent Systems*, 13(6):57–65, 1998.

[10] K. H. Siedersberger and E. D. Dickmanns. EMS-Vision: Enhanced abilities for locomotion. In this volume.

[11] C. L. Thornton and G. J. Bierman. UDU^T-covariance factorization for Kalman filtering. In C. T. Leondes, editor, *Control and Dynamic Systems: Advances in Theory and Applications*. Academic Press, Inc., New York, USA, 1980.