Recursive 3-D Road and Relative Ego-State Recognition

Ernst D. Dickmanns and Birger D. Mysliwetz

Abstract—**The general problem of recognizing both horizontal and vertical road curvature parameters while driving along the road has been solved recursively. A differential geometry representation decoupled for the two curvature components has been selected. Based on the planar solution of [7] and its refinements, a simple spatio-temporal model of the driving process allows us to take both spatial and temporal constraints into account effectively. The estimation process determines nine road and vehicle state parameters recursively at 25 Hz (40 ms) using four Intel 80286 and one 386 microprocessor. Results with the test vehicle (VaMoRs), which is a 5-ton van, are given for a hilly country road.**

Index Terms— **Intelligent control, parallel architectures, real-time machine vision, recursive estimation, spatio-temporal modeling, 3-D image sequence processing, visual navigation (of road vehicles).**

I. Introduction

THE SPECIFIC subtask of the general theme of the issue considered in this paper is 3-D road recognition while driving along a road. It is claimed that 3-D scene recognition during motion is easier than with a static camera if some knowledge about the motion behavior of the vehicle carrying the camera, the so-called "ego vehicle," is given. In the present case, it is assumed that the ego vehicle is an ordinary car with frontwheel steering driving on ordinary roads. Taking the known locomotion measured by ordinary odometers or speedometers into account, integration of measurements over time from a single passive monocular 2-D sensor allows motion stereo interpretation in a straightforward and computationally very efficient way.

The capability of road vehicles, which are abundant in our technically developed civilization, to orient themselves and avoid damage by exhibiting safe behavior on their own could considerably reduce the present death toll and economic losses in road traffic in the future. The European project PROMETHEUS aims to develop this technology among other countries. The research on which we report here was one of the pioneering efforts in Germany preceeding PROMETHEUS.

Being oriented toward general road networks, the type of scenes investigated are the human-built infrastructure "road," which is standardized to some extent but is otherwise quasi-natural with respect to environmental conditions like lighting including shadows, weather, and possible objects on the road. Here, we confine ourselves to just road recognition; beginning work on obstacle detection and relative spatial state recognition using the same methods has been reported in [12].

One of the goals of this research effort is to demonstrate that machine vision systems are able to get along with the infrastructure developed for the human driver and, in addition, to take advantage of most or all of the installations put up for the driver. For this purpose, it seems advantageous to endow the system with the capability of reasoning in space and time and to provide it with a basic "feeling" of how to react with respect to its control inputs over time, given the physical object state and, possibly, the context of a more complex situation. This implies the use of a) implicit knowledge about the motion behavior of objects, as has been developed in the engineering disciplines using dynamical models, on some lower reflex-like behavioral levels and b) explicit knowledge on higher levels for mission-specific decision taking. Contrary to other approaches using an "artificial intelligence" framework for basic spatial scene understanding, a conventional estimation scheme is used here to arrive at a spatio-temporal internal representation by analysis through synthesis.

For the planar case, this method developed in [7], [9], and [10] has shown superior performance with modest computational efforts. It rests on a differential geometry representation of the road skeletal line in the near environment of the vehicle combined with a perspective mapping model based on a Cartesian space representation of the same environment. The essential point is that time is an integral part of this internal representation allowing exploitation of temporal continuity constraints derived from the known motion behavior capabilities of the vehicle and its control input.

This method has been extended in [25] to the general case of roads with both horizontal and vertical curvature. As opposed to more theoretical approaches suggested in [5], [26] and [30], which conceive the problem as a merely static 3-D interpretation of line drawings not taking observer dynamics into account, our approach relies on data fusion from conventionally measured distance traveled (odometer data) and has been successfully implemented and tested in real world experiments with our van.

In the next section, the task to be solved is formulated. Then, image sequence analysis is interpreted as a tele-measurement process with no direct link to the object being measured, and, therefore first requiring object identification. To this end, in the following sections, the 3-D road model, the measurement model, and the egomotion model are given, which are then combined in state transition form for the case of time discrete

Manuscript received October 15, 1990; revised February 26, 1991.

E. D. Dickmanns is with the Department of Aerospace Technology, Universität der Bundeswehr Munchen, Neubiberg, Germany.

B. D. Mysliwetz is with Baasel Lasertech, Starnberg, Germany.

IEEE Log Number 9102684.

measurements. The sections on the "4-D approach" with recursive state and parameter estimation using an extended version for image sequence processing of the "extended Kalman filter" and the extraction and aggregation of visual features through "Gestalt" ideas represent the backbone of the method. Experimental results in the simulation loop and with our test vehicle VaMoRs conclude the main body of the paper. In an outlook on concurrent work in 4-D obstacle recognition, the resulting modular processing structure oriented toward physical objects is discussed.

Dynamic vision, as developed here, and the integrated treatment of spatial and temporal aspects lead to a somewhat different conception of the role that artificial intelligence methods may have to play in efficient machine vision systems. A sound combination of well-proven engineering and artificial intelligence methods seems to be in demand.

II. Task Formulation

Since steering control of conventional road vehicles immediately leads to path curvature, this term seems to be well suited to centering the analysis of road vehicle guidance around it. For the same reason, civil engineering has come to define road layouts in curvature terms. In the horizontal plane, the so-called clothoid model is used for piecing together general high-speed roads: curvature C as the inverse of the radius of curvature R

$$C = 1/R \tag{1}$$

is limited to changing linearly over arc length on each road segment:

$$C = C_0 + dC/dl * l = C_0 + C_1 * l. \tag{2}$$

$C_1 = 1/A^2$ is piecewise constant, and A is the so-called "clothoid parameter." This layout leads to the fact that while driving on the road with constant speed and steering wheel turn rates, there should be no deviation from the ideal path on the road. Therefore, an essential task for smooth road vehicle guidance is to recover the coefficients C_0 and C_1 using vision while driving.

Perspective mapping is most easily described in Cartesian space coordinates. In the planar case, for convenience, two of these coordinates are assumed to lie in this plane, say x and y. Now, a transition from the differential geometry description in curvature terms to the Cartesian description has to be found. The first integral of curvature is the heading direction χ

$$\chi = \chi_0 + \int_o^l C(\tau)d\tau \tag{3}$$

which combined with (2) yields

$$\Delta_\chi = \chi - \chi_0 = C_0 l + C_1 l^2/2. \tag{4}$$

According to Fig. 1, for the second integrals the following relations hold:

$$x = x_0 + \int_o^l \cos\chi(\tau)d\tau \tag{5}$$

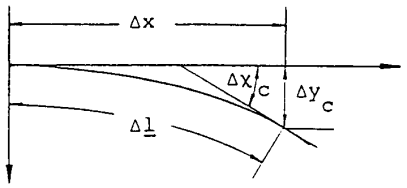


Fig. 1. From differential geometry to Cartesian coordinates.

$$y = y_0 + \int_o^l \sin\chi(\tau)d\tau. \tag{6}$$

For small heading changes $|\chi| < 15°$, the cosine is very close to 1, and the sine is close to its argument. Therefore, together with (4), the following relations are obtained:

$$x - x_0 = l \tag{7}$$

$$y = y_0 \approx C_0 l^2/2 + C_1 l^3/6. \tag{8}$$

Note that this model is not limited to a small class of applications (as has sometimes been claimed), but that by choosing proper pieces with heading changes less than 15°, any type of road can be represented to sufficient accuracy: On straight roads, C_0 and C_1 are 0 (and l is unlimited), on circular arcs C_1 is 0, and on pure transition arcs (as usually used in road design), either C_0 or the curvature at the end of the arc is 0. For the heading changes mentioned, the resulting curve corresponds very well to a spline approximation, which is well known to allow the representation of any type of curve.

Now, the task in the more general 3-D case is to recover the curvature parameters in a plane that is tangential to the road surface at the point where the vehicle just happens to be while there also is a nonzero vertical curvature. No previous knowledge about the specific environment is assumed to be given, but everything has to be derived from sensory information gathered around the point of "here and now" in space and time in conjunction with general knowledge on meaningful road parameter ranges.

Again, for the dynamical aspects, the vertical curvature (i.e., differential geometry) is of importance since additive vertical accelerations result from it. (Throughout, it is assumed that driving in standard situations occurs under $1g$ normal Earth acceleration ($g = 9.81\mathrm{m/s}^2$)). With respect to visual perception, however, Cartesian coordinates again are better suited for describing the relevant relationships. It should be noted that the differential geometry description does not contain integration constants and is therefore well suited for independently representing changing local conditions, whereas orientation in an absolute framework (like the $g$vector) requires knowledge about the integration constants.

As general (implicit) knowledge about the real world, it is taken for granted that road curvature is small: On the standard German "Autobahnen," the curvature is less than 0.001; on cross-country roads, it may go up to, say, 0.1 (dimension: 1/m). This means that depending on the situation, efficient low-pass filtering can be applied to the road parameter estimation

process; high frequency noise from uneven road surface or perturbed road boundaries can be removed by well-known standard methods.

The actual motion is assumed to correspond to standard driving situations (close to tangential to the road and no hazardous vehicle state in one of the limit regions of safe operation); the odometer and/or speedometer readings are assumed to be almost correct and correspond to the driven path arclength. In future versions, self-checking capabilities may be implemented.

Since the ideal skeletal line of the road is not visible in reality, the road parameters have to be recovered by observing the well-visible road boundaries or lane markings, if they are available. In the case of shadows on the road from trees, from installations along the road, or from buildings, the resulting intensity edges in the image may be more pronounced than those from the road boundaries. In order to be able to efficiently deal with ambiguities in such situations, it is assumed that the road boundaries are approximately parallel, which under approximately known viewing conditions allows reasoning with a "Gestalt" like appearance of the road in the perspective image. Combining this with temporal continuity in egomotion, both the road parameters and the lateral relative state of the ego vehicle to the road have to be determined in parallel.

III. Image Sequence Analysis as a Tele Measurement Process

For measuring state components of objects, usually well-designed special equipment, which delivers data in a fixed temporal pattern to a collection device with well defined data input, is being used. In vision, there is no direct interconnection between the object measured and the data processing; instead, both the object and its general state have to be inferred from the wealth of visual data that is remotely sensed. Knowledge about the visual appearance of objects under perspective projection is an essential ingredient to this process. For this reason, vision has long been considered as a domain of artificial intelligence. In an effort to solve easier problems first, as is always recommendable, one has started out with recognizing static objects under snapshot conditions (single image).

When the next difficult step of motion recognition was ready to be tackled, the methods developed for static digital image processing were carried over. Contrary to this generally chosen approach, there were a few groups with experience in systems dynamics who tried to tackle the visual motion task by resorting to the use of dynamical models for the physical motion of the objects observed [15], [23]. Note that this differs from the widely spread (mis)use of recursive estimation techniques for smoothing noisy data in the measurement domain without referring to laws of physics as knowledge representation for the motion of massive bodies. In the original sense as developed in [19], the dynamical model represented knowledge about the motion behavior of specific objects (or object classes). Instead of doing batch processing with a sufficiently large set of data for noise smoothing in the least squares sense exploiting a parameterized solution curve for the body motion (as Gauss had done in his original work on orbit determination about two centuries ago), Kalman [19] succeeded in obtaining an optimal (least squares) solution recursively by using the generic differential equation description of the solution curves together with known noise statistics. An essential achievement of Kalman's famous filter solution was the fact that by using the dynamical model, all state variables could be recovered, even though only a subset of output variables could be measured. This situation of incomplete measurements is inevitably given in vision since images are 2-D and the process to be recovered is, in general, 3-D.

Recursive estimation for real-time vision as pioneered in [15] and our group has become more popular recently (see [3], [29], [32], [17], [18], [20], [21], and [28]). Note, however, that the integrated 4-D approach as used here exploits the dynamical model for fusion of conventionally measured data (speed V and arc length l from an odometer) and for control determination as well as for the prediction of the effects of this control output on the evolution of the trajectory and on corresponding changes in the perspective image over time. The dynamical model is *the* tool for very high-level knowledge representation about the real world and its evolving processes over time. The usage of this knowledge leads to efficient motion stereo vision for the general case (here, motion along a surface) in a very natural way. The most demanding application in real-time visual guidance of vehicles known to the authors is the landing approach, including flare until touchdown of a jet aircraft that has been demonstrated by hardware-in-the-loop simulation in [14], [10] and [31]. It is, by far, not just the noise smoothing aspect (to which most authors confine the application) that makes recursive estimation with dynamical models so valuable for moving robots.

In the general vision task, both the imaging camera and some objects in the image may be moving. Therefore, the approach chosen was to describe both the motion of the camera mounted on the ego vehicle and of other moving objects depicted in the image sequence by separate dynamical models for their spatial motion in the real world instead of in the image plane. This invariably requires space *and* time as the basic variables, and it includes the effects of possible control input on the motion behavior. By this choice, the internal representation is spatio temporal and contains knowledge about the motion behavior of objects over time, which, as explicit knowledge, is a rather late achievement of human science (only less than three centuries old). Before this, humans had implicit knowledge of motion behavior of objects (and subjects), but this capability was taken for granted and was considered to be more like a skill or a talent (like in sports or in fighting) than knowledge. Most artificial intelligence approaches to robotics disregard this knowledge at the expense of computing power needed.

The laws of motion in physics are formulated for an ideal point: the center of gravity (cg). General motion of rigid bodies, to which our analysis has been confined up to now, is the translation along three independent spatial axes and the rotation around those. For each degree of freedom, a second-order differential equation can be formulated according to Newton's law. Working with sampled data in digital processing in a fixed cycle time T, difference equations, which may

be written using transition and control effect matrices, result. In image sequence processing based on conventional video signals, the basic cycle time is 16.66 ms for 60 Hz and 20 ms for a 50-Hz power supply.

Measurable through vision is not the ideal cg position, but a distribution of oriented visual features around the cg is the ideal cg position. In addition, it is not clear, from the beginning, which of the visible features belong to the object of interest and which do not. Very often, this question cannot be decided by analyzing a single frame. Only by exploiting knowledge about both the appearance of objects and their behavior or systematic change over time can this be solved. This is even more true for flexible or deformable objects in the future; for flexible (nonrigid) objects, the approach chosen may also be advantageous since it allows nice integration of elastic eigenmodes.

Conceiving vision as a measurement process, therefore, first requires identification of objects by their overall visual appearance, i.e., feature grouping, and then requires determination of its state by detailed analysis of the feature distribution and its change over time. In the case of road recognition, the problem is a little more simple since the object road is static in space, and only the ego vehicle is moving. In order to link the internal state representation of the vehicle carrying the camera with the visual appearance of the road boundaries in the image, a measurement model for perspective projection has to be coupled to the ego-motion model. The differential geometry model for road representation now allows interpretation of the spatial continuity conditions for the road as temporal continuity constraints in the form of difference equations for the estimation process while the vehicle moves along the road. By this choice, the recursive 3-D road parameter and relative ego-state estimation task can be transformed into a conventional estimation task with three cooperating dynamical submodels. Since the basic equations are nonlinear, a linearization around the present state is performed continuously. The individual models are discussed next.

IV. The 3-D Road Model

As a simplifying assumption for the 3-D case, it is assumed that horizontal and vertical curvature relative to the visual look-ahead range (to be adapted correspondingly) are so small that the effect of both curvatures can be treated as a linear superposition of the decoupled individual ones. This may not be true for hairneedle curves of mountain roads but seems to be a reasonable assumption for the general case of normal cross-country roads (and certainly for highways). This decoupling alleviates the modeling considerably; the planar solution remains practically unchanged except for the influence of vertical curvature on the look-ahead range. Vertical curvature is modeled without the $C1$ term differential equation constraint in the horizontal model.

Fig. 2 shows the individual influences of the lateral vehicle state and of the horizontal and vertical curvature parameters on the visual appearance of the road under standard driving conditions. The elevation of the camera above the road surface H_K is the major parameter influencing these images. Fig. 3

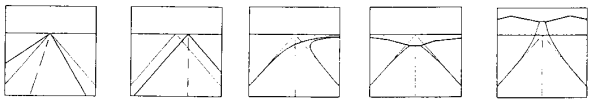


Fig. 2. Individual influences of the lateral vehicle state and of road parameters on the road image: (a) Lateral offset y_v; (b) heading offset ψ; (c) horizontal curvature right; (d) downward vertical curvature; (e) upward curvature.

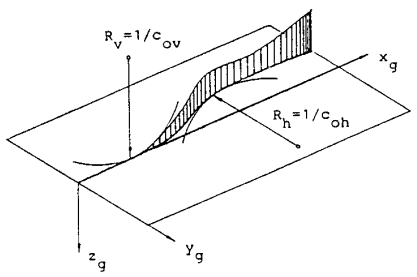


Fig. 3. Spatial road segment with horizontal and vertical curvature.

shows a perspective view on a spatial road segment with both vertical and horizontal curvature. For the horizontal curvature, the clothoid model as given in (2) is valid. When the vehicle drives along the road, the temporal curvature change becomes

$$dC/dt = dC/dl \cdot dl/dt = C_1 \cdot V \tag{9}$$

where V is the vehicle speed. The time derivative of (2) yields, for a road with piecewise linearly changing curvature (C_1 is a step function over arc length, see Fig. 4(b)), an additional term containing the time derivative of C_1, which is a sequence of Dirac impulses (see Fig. 4(d)).

Since the position of these impulses (corresponding to the stepwise jumps in C_1) cannot be determined reliably in noise corrupted measurements, a locally varying "averaged" clothoid model has been proposed in [9] and evaluated in [25]. It takes the structure of the clothoid step function into account for modeling the lateral offset y; however, it substitutes for the discrete step model for each measurement at a certain look-ahead range L a corresponding "averaged" model (index m) according to (2), yielding the same lateral offset y. This finally leads to the dynamical model (with a noise term $n_{c1h}(t)$ driving the "constant" C_{1h} in a deq. formulation)

$$\begin{aligned} \dot{C}_{0hm} &= C_{1hm} * V \\ \dot{C}_{1hm} &= -3V/L * C_{1hm} + 3V/L * C_{1h} \\ \dot{C}_{1h} &= n_{c1h}(t) \end{aligned} \tag{10}$$

which, in matrix-vector notation, may be written as

$$\dot{\mathbf{x}}_{ch} = \mathbf{A}_{ch}\mathbf{x}_{ch} + \mathbf{n}_{ch} = \begin{bmatrix} 0 & V & 0 \\ 0 & -3V/L & 3V/L \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} C_{ohm} \\ C_{1hm} \\ C_{1h} \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ n_{c1h} \end{bmatrix} \tag{10a}$$

where $n_{c1h}(t)$ results from the actual measurements.

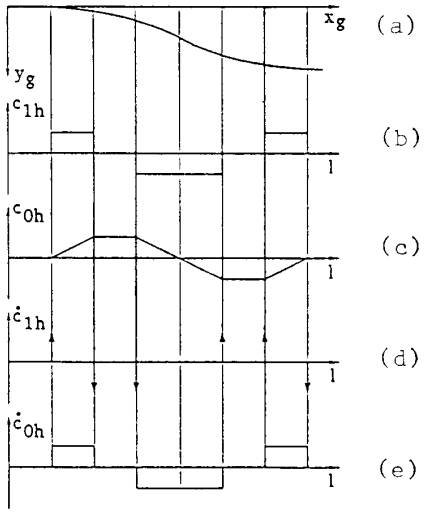


Fig. 4. Road track and horizontal curvature terms: (a) Road track; (b) curvature rate $c_{1h}(\underline{1})$; (c) curvature $c_{0h}(\underline{1})$; (d) time derivative of (b); (e) time derivative of (c).

The vertical curvature model does not have the $C1$ =const. constraint like the horizontal one (since there is no direct relation to steering control but only to vertical additive acceleration onset). In practical road building, the step in curvature corresponding to the transition from planar to circularly curved vertical cross section is smoothed by the construction process. The following dynamical model has been chosen as a reduced-order model for the vertical plane:

$$\begin{aligned} \dot{C}_{0vm} &= C_{1vm} \cdot V \\ \dot{C}_{1vm} &= n_{c1v}(t), \qquad n_{c1v}(t) = (\text{driving noise term}) \end{aligned} \tag{11}$$

or in matrix-vector notation

$$\dot{\mathbf{x}}_{cv} = \mathbf{A}_{cv} + \mathbf{n}_{c1v} = \begin{bmatrix} 0 & V \\ 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} c_{ovm} \\ c_{1vm} \end{bmatrix} + \begin{bmatrix} 0 \\ n_{c1v} \end{bmatrix}. \tag{11a}$$

V. The Measurement Model

As can be seen from Fig. 2, both the lateral vehicle states and the curvature parameters affect the appearance of the road in the image plane. The horizontal and the vertical mapping models will be discussed separately.

A. Horizontal Mapping Geometry

Fig. 5 shows the horizontal road and mapping geometry. The horizontal image coordinate y_B of a road boundary edge element at the look-ahead distance L from the projection center PZ is determined by the following road curvature parameters, mapping parameters, and relative state terms:

f	focal length [mm]
k_y	horizontal scaling factor for the camera [pel/mm]
L	look-ahead distance [m]

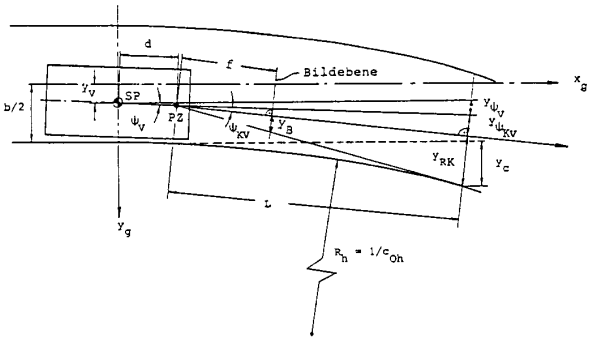


Fig. 5. Horizontal road and mapping geometry; point SP is the vehicle center of gravity, PZ is the camera projection center; the following index conventions are valid: V is the vehicle, K is the camera, R is the road, B is the image coordinate, and g is the relative to geodetic coordinate system.

d	distance from projection center to vehicle cg [m]
ψ_{KV}	camera heading angle relative to vehicle reference axis [rad]
ψ_V	vehicle heading angle relative to road tangent [rad]
y_V	lateral offset of vehicle cg from road center [m]
b	width of road [m]
C_{0hm}	average horizontal road curvature [1/m]
C_{1hm}	average horizontal road curvature rate [1/m^2] .

Because of the small angles involved, the sine is approximated by its argument and the cosine by 1. Since $f \ll L$, the position of the projection center and the rotation axis for the camera are assumed to have the same distance to the point at range L mapped. It is seen from Fig. 5 that at the look-ahead distance L, the following relation approximately holds (cos = 1):

$$y_c + b/2 = y_{RK} + y_{\psi KV} + y_{\psi V} + y_V \tag{12}$$

where

$$y_c = (L+d)^2/2 \cdot C_{0hm} + (L+d)^3/6 \cdot C_{1hm} \tag{13}$$

according to (8), and

$$y_{\psi KV} = L \cdot \psi_{KV}, \qquad y_{\psi V} = (L+d) \cdot \psi_V. \tag{14}$$

From (12), there follows

$$y_{RK} = b/2 + y_c - y_V - y_{\psi KV} - y_{\psi V}$$

which is being mapped into y_B in the image plane by the law of perspective mapping as

$$y_B = (f \cdot k_y)/l \cdot y_{RK} \tag{15}$$

[in picture element (pel) units].

In the planar case, the look-ahead distance L depends on the camera pitch angle θ_K relative to the plane and on the image line z_{Bi} that is evaluated; if the road is vertically curved, this directly affects the look-ahead distance, and the relations become more nonlinear, as will be discussed next.

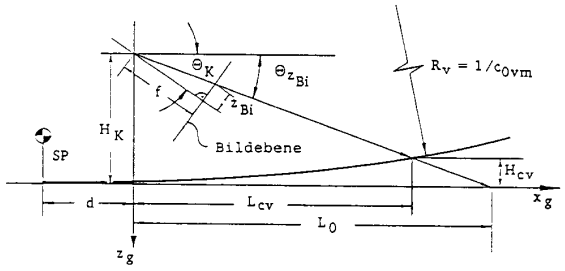


Fig. 6. Vertical road and mapping geometry.

B. Vertical Mapping Geometry

According to Fig. 6, the vertical mapping geometry is mainly determined by the camera elevation H_K above the local tangential plane and the pitch angle. The longitudinal axis of the vehicle is assumed to be always tangential to the road at the vehicle cg, which means that high-frequency pitch disturbances are neglected. This has proven to be realistic for stationary driving states on "normal," i.e., flat, well-kept roads. However, this is not valid for uneven roads or under powerful acceleration or deceleration maneuvers that cause heavy pitching motions of the vehicle. For handling the latter cases, the additional estimation of the pitch angle remains to be implemented, which is, however, straightforward in the approach chosen. The additional terms used in the vertical mapping geometry are collected in the following list:

k_z	camera scaling factor, vertical [pel/mm]
H_K	elevation of the camera above the tangential plane at cg [m]
θ_K	camera pitch angle relative to vehicle pitch axis [rad]
z_B	vertical image coordinate in pel-units [pel]
L_0	look-ahead distance for planar case [m]
L_{cv}	look-ahead distance with vertical curvature [m]
H_{cv}	elevation change due to vertical curvature [m]
C_{0vm}	average vertical curvature of road [1/m]
C_{1vm}	average vertical curvature rate of road [1/m2].

To each image scan line at z_{Bi}, there corresponds a pitch angle relative to the local tangential plane of

$$\theta_{zBi} = \theta_K + \arctan\left(z_{Bi}/(f \cdot k_z)\right). \tag{16}$$

From this, the planar look-ahead distance corresponding to z_{Bi} is obtained as

$$L_{0i} = H_K / \tan\left(\theta_{zBi}\right). \tag{17}$$

Analogous to (8), the elevation change due to the vertical curvature terms at the distance $L_{cv} + d$ relative to the vehicle cg is

$$H_{cv} = C_{0vm} \cdot (L_{cv} + d)^2/2 + C_{1vm} \cdot (L_{cv} + d)^3/6. \tag{18}$$

From Fig. 6, the following relationship can be read immediately:

$$H_{cv} = H_K - L_{cv} \cdot \tan\left(\Theta_2 B_i\right) \tag{19}$$

which combined with (18) yields the following third-order polynomial for determining the look-ahead distance L_{cv} with vertical curvature included:

$$a_3 L_{cv}^3 + a_2 L_{cv}^2 + a_1 L_{cv} + a_0 = 0 \tag{20}$$

where

$$\begin{aligned} a_3 &= C_{1vm}/6 \\ a_2 &= (C_{0vm} + d \cdot C_{1vm})/2 \\ a_1 &= d \cdot (C_{0vm} + d \cdot C_{1vm}/2) + \tan\left(\theta_{zBi}\right) \\ a_0 &= d^2 \cdot C_{0vm}/2 + d^3 \cdot C_{1vm}/6 - H_K. \end{aligned}$$

This equation is solved numerically with the nominal curvature parameters of the last cycle via a Newton iteration; taking as a starting value for L_{cv} the solution of the previous iteration or the planar solution according to (17), the iteration typically converges in two or three steps, which means at small computational expense.

Neglecting the a_3 term in (20) or the influence of C_{1vm} on the look-ahead range entirely would lead to a second-order equation that is easily solvable analytically. Disregarding the C_{1vm}term resulted in errors in the look-ahead range when entering a segment with a change in vertical curvature and lead to wrong predictions in road width. The lateral tracking behavior of the feature extraction windows (see below) with respect to the road width change resulting from vertical curvature could be improved considerably by explicitly taking the C_{1vm} term into account. (There is, of course, an analytical solution to a third-order equation that is also available; however, the iteration is more efficient computationally since there is little change from k to $k+1$. In addition, this avoids the need for selecting one out of three solutions of the third-order equation).

Beyond a certain negative vertical curvature, it may happen that the road image lies below the image scan line z_{Bi} chosen for evaluation; this means that there is no extractable road boundary element in this line. The curvature for this limiting case, in which the ray through z_{Bi} is tangential to the road surface at the distance L_{cv}, can approximately be determined by the second-order polynomial neglecting the C_{1vm} influence as mentioned above. In addition, neglecting the $d \cdot C_{0vm}$ terms, the approximate solution for L_{cv} becomes

$$L_{cv} = \frac{\tan\Theta_{zBi}}{-C_{0vm}}\left[1 - \sqrt{1 + 2\frac{H_K C_{0vm}}{\tan^2\Theta_{zBi}}}\right]. \tag{21}$$

The limiting tangent case for maximal negative curvature is reached when the radicant becomes zero, yielding

$$C_{ovm,\lim}(z_{Bi}) = -\tan^2\left(\Theta_{zBi}\right)/(2 \cdot H_K). \tag{22}$$

Because of the neglected terms, a small "safety margin" ΔC may be added. If the actually estimated vertical curvature C_{0vm} is smaller than the limiting case corresponding to (22) (including the safety margin of, say, 0.0005), no look-ahead distance will be computed, and the corresponding features will be eliminated from the measurement vector.

In the measurement equations (12)–(15), two vehicle state components, the lateral offset on the road y_V, and the heading angle ψ_V enter. Instead of determining these from a sufficiently large set of measurement data for each image anew from scratch, as is usually done in quasistatic image

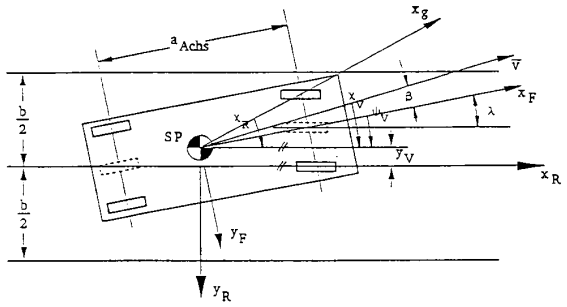


Fig. 7. Substitute model for lateral vehicle motion.

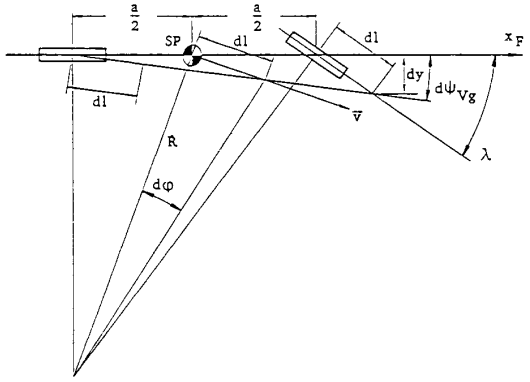


Fig. 8. Steering kinematics.

processing, the fact is exploited here that a dynamical model for the vehicle motion is known. A simple but sufficiently good model will be discussed next.

VI. Ego–Motion Model

Fig. 7 shows the state and control variables for catching the essential lateral motion constraints of a vehicle with frontwheel steering. The heading angle ψ_V describes the orientation of the longitudinal axis relative to the road tangent at the vehicle cg. The road tangent turn rate $\dot{\chi}_v$ relative to inertial space is the product horizontal curvature times speed:

$$\dot{\chi}_v = C_{0h} \cdot V. \tag{23}$$

The C_{1h} term may be neglected because its local influence on heading over arc length traveled during one cycle is small ($dt * V_{\max} = 40$ ms $* 30$ m/s $= 1.2$ m).The vehicle trajectory is controlled by the steering angle λ, which fixes the turn radius R of the vehicle (see Fig. 8). For an infinitesimal move dl, the following approximate relationship holds for $a \gg dl$:

$$dl \cdot \sin(\lambda) = a \cdot \sin[d(\psi_{vg})]. \tag{24}$$

For small angles λ and ψ, the time derivative of this relation yields

$$\dot{\psi}_{vg} = V \cdot \lambda / a \tag{25}$$

where ψ_{vg} is the heading angle with respect to inertial space. In order to obtain the heading change ψ_V relative to the road tangent at the present location of the vehicle cg along the road, the road heading turn rate has to be subtracted:

$$\dot{\psi}_V = V \cdot (\lambda / a - c_{oh}). \tag{26}$$

Again, the C_{1h} term has been neglected due to its smallness. The vehicle heading angle ψ_v is linked to the path azimuth angle χ_V of the vehicle by the slip angle β:

$$\chi_V = \psi_V + \beta. \tag{27}$$

This discrepancy between trajectory and vehicle heading occurs because of tire softness and of slipping between the tire and the ground. The lateral speed relative to the road then is

$$\dot{y}_V = V \cdot \sin(\chi_V) = V \cdot \chi_V. \tag{28}$$

The slip angle β also is constrained by a differential equation to be found in the automotive dynamics literature (see [25])

$$\dot{\beta} = -2 \cdot K \cdot \beta + (V/a - K) \cdot \lambda \tag{29}$$

where $K = k_r/(m \cdot V), k_r$ is the tire lateral force coefficient (150 kN/rad), m is the vehicle mass (4000 kg), V is the actual vehicle speed, and a is the distance between front and rear axis (3.5 m) (numbers in brackets refer to our test vehicle **VaMoRs**, which is a Daimler-Benz van L 508 D).

The steer angle is set by a stepping motor, the dynamic behavior of which is roughly modeled as an integrator

$$\dot{\lambda} = k_\lambda \cdot U. \tag{30}$$

Collecting these relations results in a linear velocity-dependent fourth-order state model

$$\begin{aligned}
\dot{\lambda} &= k_\lambda \cdot U \\
\dot{\beta} &= -2K \cdot \beta + (V/a - K) \cdot \lambda \\
\dot{y}_v &= V \cdot (\psi_v + \beta) \\
\dot{\psi} &= V/a \cdot \lambda - V \cdot C_{0h}
\end{aligned}$$

or in matrix-vector notation

$$\dot{\mathbf{x}}_v = \mathbf{A}_v \cdot \mathbf{x}_v + \mathbf{b}_v \cdot U + \mathbf{B}_c \cdot C_{0h} \tag{31}$$

where

$$A_v = \begin{bmatrix} 0 & 0 & 0 & 0 \\ b_F & a_F & 0 & 0 \\ 0 & V & 0 & V \\ c_F & 0 & 0 & 0 \end{bmatrix};$$

$$x_v = \begin{bmatrix} \lambda \\ \beta \\ y_v \\ \psi_v \end{bmatrix}; b_v = \begin{bmatrix} k_\lambda \\ 0 \\ 0 \\ 0 \end{bmatrix}; b_c = \begin{bmatrix} 0 \\ 0 \\ 0 \\ -V \end{bmatrix}$$

with the elements

$$c_F = V/a, \quad b_F = c_F - K$$

and

$$a_F = -2 \cdot K.$$

These equations represent knowledge about the lateral motion behavior of a road vehicle under normal driving conditions. The first term on the right-hand side of (31) gives the state transition under zero control and curvature input; the second term indicates that control action just turns the steering wheel, which then, through the first column in the matrix A, affects the slip angle β and the heading angle ψ after integration. The third term gives the influence of road curvature on the heading change relative to the road.

In the automotive literature, much more elaborate motion models can be found, leading to system descriptions of more then 20th order; they contain vehicle pitch, yaw, and roll and the wheel suspension dynamics individually for each wheel. In our applications, we are more interested in the basic trajectory dynamics in order to 1) reduce the search space for the evaluation of the next image, given the results of the previous evaluations, and 2) to reduce the combinatorial explosion of feature aggregation by exploiting basic temporal continuity conditions. Depending on the special task to be performed, some expansions of the model, like the pitching motion mentioned above, may be advantageous in the future.

VII. The Combined Dynamical Model for 3-D Road Recognition

The overall dynamical model for 3-D road and relative lateral state recognition consists of the three subsystems: 1) for the lateral dynamics of the vehicle, 2) for the horizontal, and 3) for the vertical curvature dynamics. The following first-order system with nine state components describes the differential equation constraints for guiding the 3-D interpretion exploiting knowledge about temporal processes in the real world. It can be seen that the three subsystems in the continuous form are almost completely decoupled; only the horizontal road curvature affects the lateral vehicle dynamics.

For sampled data systems in digital control with the sampling period T, the corresponding state transition matrices and the control effect coefficient vector shown at the bottom of this page in (32) can be obtained by the number of standard methods in systems dynamics theory (details are given in [25]). It should be noted that most of the coefficients depend on the vehicle speed; if vertical curvature is nonzero, the look-ahead range L depends on the curvature parameters, and the system becomes more nonlinear. Since the recursive estimation technique requires linear models, the equations are locally linearized around the actual reference point.

VIII. The 4-D Approach to Real-Time Vision

The dynamical models link time to spatial motion, in general. The shape models contain the spatial distribution of visual features that allow recognition and tracking of objects. In order to exploit both types of models at the same time, the prediction error feedback scheme for recursive state estimation developed by Kalman and his successors has been extended to image sequence processing by our group [32]. There are so many publications on this approach that only a short summary will be given here (see, e.g., the survey article [11]). . 9 shows the resulting overall block diagram of the vision system based on these principles. To the left, the real world is shown by a block; control inputs to the vehicle may lead to changes in the visual appearance of the world either by changing the viewing direction or through egomotion. The continuous changes of objects and their relative position in the world over time are sensed by CCD-sensor arrays (shown on the left as converging lines to the lower right, symbolizing the 3-D to 2-D data reduction). They record the incoming light intensity from a certain field of view at a fixed sampling rate. By this imaging process, the information flow is discretized in several ways: There is a limited spatial resolution in the image plane and a

$$
\begin{bmatrix} [d/dt]\lambda \\ \beta \\ y_V \\ \psi_V \\ -- \\ c_{0hm} \\ c_{1hm} \\ c_{1h} \\ -- \\ c_{0vm} \\ c_{1vm} \end{bmatrix}
=
\left[\begin{array}{cccc|ccc|cc}
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
b_F & a_F & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & V & 0 & V & 0 & 0 & 0 & 0 & 0 \\
c_F & 0 & 0 & 0 & -V & 0 & 0 & 0 & 0 \\
-- & -- & -- & -- & -- & -- & -- & -- & -- \\
0 & 0 & 0 & 0 & 0 & V & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & -3V/L & 3V/L & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
-- & -- & -- & -- & -- & -- & -- & -- & -- \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & V \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0
\end{array}\right]
\cdot
\begin{bmatrix} \lambda \\ \beta \\ y_V \\ \psi_V \\ -- \\ c_{0hm} \\ c_{1hm} \\ c_{1h} \\ -- \\ c_{0vm} \\ c_{1vm} \end{bmatrix}
+
\begin{bmatrix} k_\lambda \\ 0 \\ 0 \\ 0 \\ -- \\ 0 \\ 0 \\ 0 \\ -- \\ 0 \\ 0 \end{bmatrix}
\cdot u_\lambda +
\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ -- \\ 0 \\ 0 \\ n_{c1h} \\ -- \\ 0 \\ n_{c1v} \end{bmatrix}
\tag{32}
$$

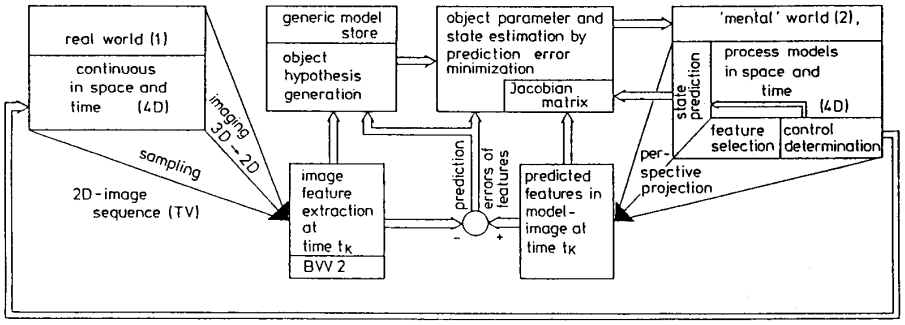


Fig. 9. Survey block diagram of the cybernetic 4-D approach to vision.

temporal discretization of 16.66 or 20 ms, usually including some averaging over time.

Instead of trying to invert this image sequence for 3-D-scene understanding, a different approach by analysis through synthesis has been selected. In an initialization phase, starting from a collection of features extracted by the low-level image processing (BVV_2, lower center left in Fig. 9), object hypotheses including the aspect conditions and the motion behavior (transition matrices) in space have to be generated (upper center left). They are installed in an internal "mental" world representation intended to duplicate the outside real world. This is sometimes called "world_2" as opposed to the real "world_1."

Once an aggregation of objects has been instantiated in world_2, exploiting the dynamical models, the object states can be predicted for that point in time when the next measurement is going to be taken. By applying the *forward* perspective projection to those features that will be well visible, using the same mapping conditions as the TV sensor, a model image, which should duplicate the measured image if the situation has been understood properly, can be generated. The situation is thus "imagined" (right and lower center right in Fig. 9). The big advantage of this approach is that due to the internal 4-D model, not only the actual situation at the present time but also the sensitivity matrix of the feature positions with respect to state changes can be determined (the so-called Jacobian matrix in the upper block in the center right, lower right corner). This rich information is used to bypass the perspective inversion via recursive least squares filtering through feedback of the prediction errors of the features. Unfortunately, space does not allow us to go into more details here.

This approach has several very important practical advantages:

1) No previous images need to be stored and retrieved for computing optical flow or velocity components in the image as an intermediate step in the interpretation process.
2) The transition from signals (pel data in the image) to symbols (spatio-temporal motion state of objects) is done in a very direct way, based on higher level knowledge and with the 4-D world model integrating spatial and temporal aspects.
3) Intelligent nonuniform image analysis becomes possible, allowing concentration of limited computer resources to areas of interest known to carry relevant information.
4) The position and orientation of well visible features can be predicted, and the feature extraction algorithm can be provided with information to more efficiently find the desired ones; outliers can easily be removed, thereby stabilizing the interpretation process.
5) Viewing direction control can be done directly in an object-oriented manner.

Processing a variable number of features extracted from frame to frame is alleviated by using the sequential filtering version. For improving the numerical performance, the UD-factorized version of the square root filter is used [1]. Details may be found in [25], [2], and [22]. By exploiting the sparseness of the transition matrix in the dynamical model, a speedup can be achieved.

Special care has to be taken in the initialization phase when good object hypotheses are in demand. From feature aggregations that may have been collected in a systematic search covering extended regions of the image, the presence of objects has to be hypothesized. For roads, the coexistence of left- and right-hand side boundaries in a narrow range of meaningful distances (say 2 to 15 m, depending on the type of road) and with low curvatures are the guidelines for a systematic search. Fig. 10 shows some results with a search region of six horizontal stripes; in this case, at the left-hand side, the lane boundary marked by a gap in the concrete surface filled with tar has been accepted as a lane boundary. Realistic lane widths are known to the vehicle (say 2.5 to 4.5 m). A low curvature line model is fit to each boundary in a least squares sense. Note that the vehicle need not be positioned on the road as in a normal driving mode initially; it may be parked on one side or have some larger angle relative to the road tangent. For this reason, only in the initialization phase, the general inversion problem for road and relative ego state has to be solved. Assuming constant road (lane) width and planar surface, knowing the camera elevation above the ground, this can be done in a straightforward manner; close to the vehicle, a linear approximation to the road boundary may be sufficient for starting the recursive estimation process, which can then be extended to further distances. The recursive estimation process by itself has a certain range of convergence to the proper solution so that an approximate initialization is

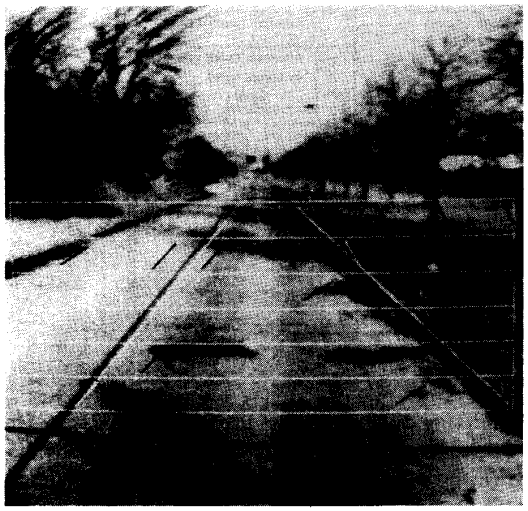


Fig. 10. "Nearby" image region (lower part) is analyzed for initial recognition of the road. Each of six stripes is searched with correlation masks of different orientations for points with maximal correlation values. Only those with consistent position and orientation are retained as lane boundary candidates.

sufficient. This process is not done on one static image but by using active vision with proper viewing direction control; however, there is no need for sticking to a small, fixed cycle time as during locomotion.

Once the road has been recognized, the task becomes much more simple by taking the temporal continuity conditions captured in the dynamical model into account.

IX. Feature Extraction and Processing Structure

The following relatively simple processing structure has evolved out of the 4-D approach in road vehicle guidance applications (see Fig. 11):

The vertical structure consists of four processing layers: the horizontal subdivision depends on the number of objects of interest in the image (two in Fig. 11, the road (right), and one obstacle (left)). The lowest vertical level is formed by 2-D image data A/D conversion (256*240 * 8 b) and distribution to each parallel processor (PP) for feature extraction via a videobus. Each PP in the image sequence processing system BVV2 presently in use can grab up to 4 KB of data by software control from a rectangular subimage of any position, shape, and regular sparseness [16] (see Fig. 11, squares, both left and right, and cross-shaped rectangles, left). PP's for feature extraction form the second processing level of the system. No image preprocessing or conventional "low-level" image processing (into a modified image) is being done. Instead, linearly extended features (edge elements = edgels) are extracted with respect to orientation and position by finding maxima in ternary correlations with special elementary templates or masks along certain search directions. Both the selection of mask orientations and of search directions (areas) may be controlled from the higher interpretation level according to the actual needs, and are continuously fine tuned to the object hypothesis under consideration. Thus, there is a frequent communication exchange both bottom up and top down among the lower three levels, which all reside in the multiprocessor system BVV2.

The interpretation level three is geared to single physical objects and their appearance in space and time. Here, the dynamical models are installed, and there are specialists for recognizing, tracking, and interpreting certain object classes.

Presently, a road specialist and one obstacle specialist are in operation. The PP's are Intel 8086 and 286 16-b processors, and the 4-D object processors on the interpretation level are 32–b 80386's. The latter ones work at an update rate of 25 Hz; PP's usually work at 50 Hz (full video rate). In general, processing cycle time increases with the hierarchy level; vehicle control operates at 12.5 Hz (80 ms). For situation assessment, in many cases, still longer processing times are acceptable since reflexlike behavior, adapted to the general situation recognized, is assumed to be able to deal with small deviations from the normal. Taking human performance as a reference, 300–500 ms or even 1 s seem to be sufficient on this supervisory level. Up until very recently, the uppermost level has been implemented on a ruggedized version of a PC (an 8-MHz Intel 80286) in the vehicle. In order to improve the system capability of dealing with several objects at a time, of handling more complex situations by selecting proper behavioral modes realized by different feedback laws, and of performing knowledge processing, the upper levels are being ported onto a 32-b multiprocessor system.

Two black and white CCD TV sensors are mounted fixed to each other on a pan and tilt platform with high dynamics. One has a short focal length for a wide viewing angle in order to see a large portion of the road nearby; the other has a telelens for good resolution further down the road in order to be able to detect obstacles reliably at a rather large distance (center of Fig. 11). The platform is controlled by a microprocessor integrated into the image processing system so that a fixation type vision mode can be implemented easily; the object on which to fixate is decided by the situation level. Inertial sensor signals are being integrated into the viewing direction control so that stabilization on uneven terrain becomes possible.

Road recognition, as discussed here, has been performed with the right processor cluster in Fig. 11 with four PP Intel 80286's working on eight window areas of 48*48 pel. Obstacle recognition (left cluster) has been reported in [12] and [34]. Data rate reduction upwards through the levels is significant: From the 1.6 MB/s per camera videodata rate, each PP has to catch only 120 KB/s. It extracts from each window a small set of candidates for road boundary elements; from this, an input data rate into the interpretation level of less than 8 KB/s per object in the scene results. Its output is the nine-element state vector every 80 ms representing a data stream of less than 1 KB/s.

X. Experimental Results

A. Planar Case

A wealth of experience has been accumulated with the differential geometry road representation for the planar case

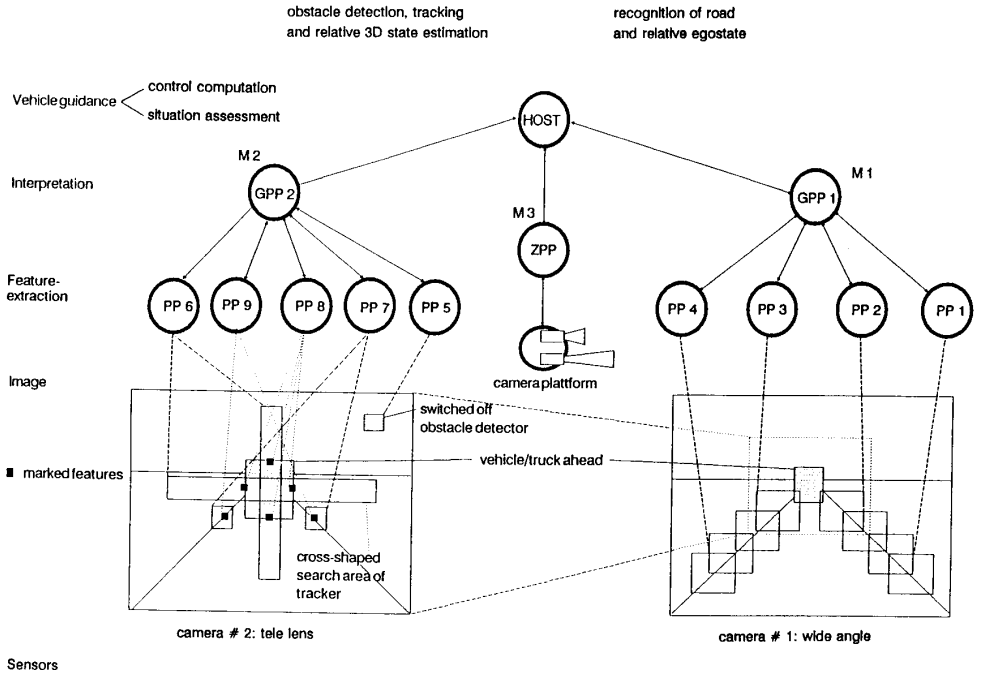


Fig. 11. Processing structure for vehicle guidance through dynamic real-time vision using two cameras.

($C_v \equiv 0$) over the last six years. Initially, it had been tested in real-time simulation with image-sequence-processing hardware in the loop. Compared with previous results [6], this leads to much more stable road following performance due to the curvature feedforward to lateral control [7]. At that time, the processing power available allowed just two windows: one nearby and the other one somewhat down the road; in each window, only one position with the highest correlation value for just one angular template orientation had been determined. This was sufficient in low-noise simulation environments.

In 1986, our experimental vehicle for autonomous mobility and computer vision **VaMoRs**, which is a 5-ton van, became available for tests on real roads. When these had good lane markings or a good contrast to the shoulders (like bright concrete or dark bitumen versus soil or fresh grass) road boundary recognition was quite reliable. Speeds up to 100 km/h, limited only by the engine power, and distances of more than 20 km had been achieved in 1987 in a fully autonomous driving mode on a new stretch of Autobahn that had not yet been turned over to the public [8]. Even driving under light rain conditions was shown to be possible. Changing lanes at an Autobahn entry was demonstrated as well. Speed V was autonomously adjusted to road (path) curvature C so that a preset acceleration limit $a_{y\,\max}$ (say 0.1 Earth gravity $\approx 1m/s^2$) was never exceeded. Therefore, both the lateral and the longitudinal guidance on free roads was done fully autonomously.

On cross-country roads without lane markings and with shadows from trees, poles, or buildings under an oblique angle on the road, this simple approach had problems in selecting the proper edge track representing the road boundary. A more discriminative description of the appearance of roads and its boundary slopes as elementary measurement units was needed. The solution was sought and found by what psychologists call the "Gestalt idea": In the image sequence, combinations of line features that correspond to a generically known object are searched out; the influence of perspective mapping is fully taken into account. In our case of road recognition, the basic Gestalt properties in 3-D space are 1) two or more parallel lines, 2) low curvature ($C \lesssim 0.03m^{-1}$), 3) on a plane almost parallel to the viewing direction. In the image plane, this leads to pencil-type road images, where the axis of the pencil may be curved; these structures extend over a considerable part of the image.

In Figs. 12 and 13, two examples of different degrees of difficulty are shown. Fig. 12 details the determination of road edge candidates by controlled correlation along a horizontal search path under a certain angular orientation of the correlation mask. Only those that combine to a smoothly curved (locally close to collinear) line are accepted as more reliable road edge candidates; they need not be globally maximal with respect to the correlation value along the search path. This also points to the fact that in real-world situations, there may be no need for a compute-intensive "optimal" feature extractor ("optimal" in image processing terms) since there are task- and situation-specific aspects overriding the pattern analysis ones (context, semantics).

In Fig. 13, a situation (evening with low standing sun) is depicted where it is almost hopeless to recognize the lane without the "Gestalt" idea telling the system the combinations

Fig. 12. Road boundary feature extraction exploiting knowledge about the situation (aspect conditions).

Fig. 13. Supporting feature extraction by the "Gestalt" idea of a road lane under perspective projection.

of features that it should be looking for. Only by adding to this the concept of temporal continuity and knowledge about the egomotion conventionally measured, these types of difficult situations can be handled with such little computing power as available in **VaMoRs**.

For more details, refer to [24] and [25]; many different types of roads and lighting as well as weather conditions have been tried with success. In the sequel, we will discuss 3-D (hilly) roads.

B. Simulation Results for 3-D Roads

Figs. 14 and 15 show results from a hardware-in-the-loop simulation with video-projected computer-generated imagery interpreted with the BVV2. This setup has the advantage over field tests that the solution is known to high accuracy.

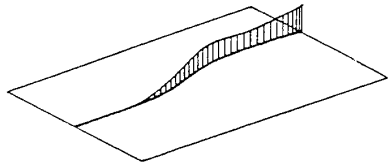
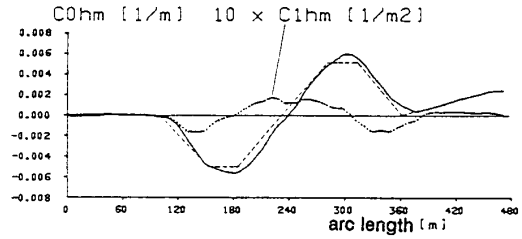
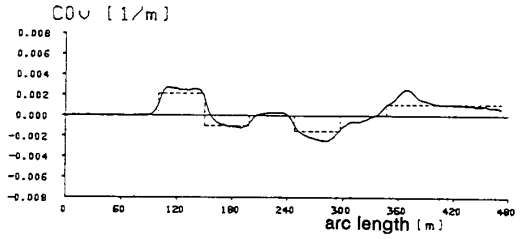


Fig. 14. Perspective view simulated 3-D road.



(a)



(b)

Fig. 15. (a) Horizontal curvature values over arc length: input (dashed) for C_{0h} and estimates (solid); (b) vertical curvature values over arc length: input (dashed) for C_{0V} and estimates (solid).

Fig. 14 is a perspective display of the tested road segment with both horizontal and vertical curvature. Fig. 15 shows the corresponding curvatures recovered by the estimation process described (solid) as compared with the ones used for image generation (dashed). Fig. 15(a) displays the good correspondence between the horizontal curvature components containing four clothoid elements and two circular 200-m-radius arcs. Even the C_{1hm} curve is relatively smooth.

Under cooperative conditions in the simulation loop, vertical radii of 1000-m curvature could be recognized reliably with a look-ahead range of 20 m. The relatively strong deviation at 360 m in Fig. 15(b) is due to a pole close to the road (with very high contrast), which has been mistaken as part of the road boundary. The system recovered from this misinterpretation all on its own when the pole was approached; the local fit with high vertical curvature became increasingly contradictory to new measurement data of road boundary candidates in the far look-ahead range. The parameter C_{0V} then converged back to the value known to be correct (400 to 450-m arc length in Fig. 15(b)). Since this approach is often questioned as to whether it yields good results under stronger perturbations and noise conditions, a few remarks to this point seem in order. It is readily seen that the interpretation is most reliable when it is close to the vehicle for several reasons:

1) The resolution in the image is very high; therefore, there are many pel per unit area in the real world from which to extract information; this allows achievement of

relatively high estimation accuracies for the lane (road) width and the lateral position of the camera on the road.

2) The elevation above the road surface is well known, and the vehicle is assumed to remain in contact with the road surface due to Earth gravity; because of surface roughness or acceleration/deceleration, there may be a pitching motion, the influence of which on feature position in the image is, again, smallest nearby. Therefore, predictions through the dynamical model are trusted most in those regions of the image corresponding to an object region spatially close to the camera; measured features at positions outside the estimated 3σ range from the predicted value are discarded (σ is the variance determinable from the covariance matrix, which in turn is a byproduct of recursive estimation).
3) Features actually close to the vehicle have been observed over some period of time while the vehicle moved through its look-ahead range; presently, this is 40 to 70 m. For a speed of 30 m/s (108 km/h), this corresponds to about 2 s or 50 frames traveling time (at 40 ms interpretation cycle time). If there are some problems with data interpretation in the far range, the vehicle will have slowed down, yielding more time (number of frames) for analysis when the trouble area is approached.
4) The "Gestalt" idea of a low curvature road under perspective projection and the egomotion (under normal driving conditions, no skidding) in combination with the dynamical model for the vehicle including control input yield strong expectations allowing selection of those feature combinations that best fit the generic road (lane) model even if their correlation value from oriented edge templates is only locally but not globally maximal in the confined search space. In situations like the one shown in Fig. 13, this is more the rule than the exception.

In the general case of varying road width, an essential Gestalt parameter is left open and has to be determined in addition to the other ones from the same measurements; in this case, the discriminative power of the method is much reduced. It is easy to imagine that any road boundary image from a hilly road can also be generated by a flat road of varying width (at least in theory and for one snapshot). Taking temporal invariance of road shape into account and making reasonable assumptions about road width variations, this problem may also be resolvable, at least for the region nearby, when it has been under observation for some time (i.e., due to further extended look-ahead ranges). Due to limitations in image resolution at far look-ahead distance and in computing power available, this problem had not been tackled in the past.

With viewing direction stabilization and larger focal lengths becoming available in addition to increased processing capabilities, this area of development is the natural next step being tackled presently. If processing power permits, several "Gestalt" hypotheses may be tested in parallel over time and compared with each other relative to some yet-to-be-defined performance measure.

Note that the only low-level image operations are correlations with local (7- to 16-pel long) edge templates of various orientations (covering the full circle at discrete values, e.g., every 11°). Therefore, there is no problem of prespecifying other feature operators. Selecting those to be applied is done by the higher system levels depending on the context. In order to exploit continuity conditions of the real world roads, sequences of feature candidates to be measured in the image are defined from near to far (bottom up in the image plane), taking adjacency and neighboring orientation conditions into account.

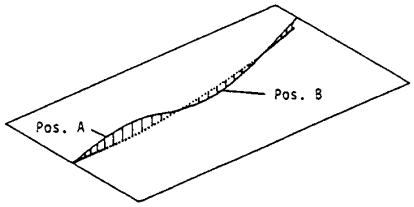


Fig. 16. Perspective view on rural road as determined from the estimated curvature parameters.

C. Real-World Experiments

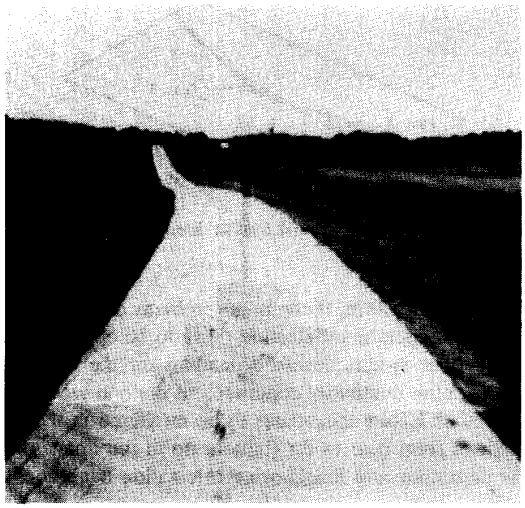
Fig. 16 shows a narrow sealed rural road with a cusp in a light curve to the left followed by an extended positively curved section that has been interpreted while driving on it with **VaMoRs** (see also Fig. 17).

For vertical curvature estimation, road width is assumed to be constant. Ill-defined or irregular road boundaries, as well as vehicle oscillations in pitch, affect the estimation quality correspondingly. These effects are considered to be the main causes for the fluctuations in the estimates of the vertical curvature in Fig. 18.

In order to improve these results in the framework of the 4-D approach geared to dynamical models of physical objects for the representation of knowledge about the world, it is felt that the pitching motion of the vehicle has to be taken into account. There are several ways of doing this:

1) The viewing direction of the camera may be stabilized by inertial angular rate feedback. This well-known method has the advantage of reducing motion blur. There are, however, drift problems if there is no position feedback. Therefore, the feedback of well discriminable visual features yields nice complementary signals for object fixation.
2) The pitch motion of the egovehicle is internally represented by another dynamical model of second order around the pitch axis. Tracking horizontal features far away (like the horizon) vertically allows to estimate pitch rate and angular position of the vehicle recursively by prediction error feedback. Again, knowledge about the pitching dynamics of the massive inertial body is exploited for measurement interpretation. Picking features far away on the longitudinal axis of the body decouples this motion component from other ones.
3) Purely visual fixation (image registration from frame to frame) may be implemented.

The first two are being investigated presently by members of our group. The third one is being studied elsewhere, e.g. [4], [27].



(a)

(b)

Fig. 17. Two snapshots from road of Fig. 16 with camera elevation of 1.8 m (Note brighter areas of evaluated image windows): (a) Sealed rural as seen from position A; (b) same road as seen from position B.

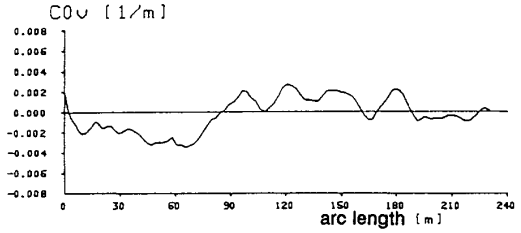


Fig. 18. Estimated vertical curvature of rural road in Fig. 17 while driving on it with **VaMoRs**.

XI. Conclusions

The 4-D approach to real-time 3-D visual scene understanding allows the spatial interpretation of both horizontally and vertically curved roads while driving. Exploiting recursive estimation techniques that have been well developed in the engineering sciences, this can be achieved at a high evaluation rate of 25 Hz with a rather small set of conventional microprocessors. If road width is completely unconstrained, ill-conditioned situations may occur. In the normal case of parallel road boundaries, even low curvatures may be recovered reliably with modest look-ahead ranges.

Adding temporal continuity conditions to the spatial invariance properties of object shapes allows to reduce image processing requirements by orders of magnitude. Taking physical objects as units for representing knowledge about the world, there results a spatio-temporal internal representation of situations in which the object state is continuously servoed according to the visual input taking perspective projection and motion constraints to the changing aspect conditions into account.

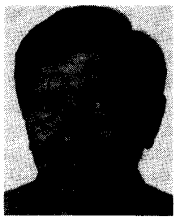
Objects are recognized and tracked by exploiting the Gestalt idea for feature grouping. Critical tests have to be performed in order to avoid "seeing what you want to see." This problem is far from being solved; much more computing power is needed in order to be able to handle more complex situations with several objects.

However, the general approach is considered to be a promising candidate on which to base real-time dynamic machine vision. For more details, refer to [11]–[13].

References

[1] G. J. Bierman, "Measurement updating using the U-D factorization," in *Proc. IEEE Contr. Decision Conf.* (Houston, TX), 1975, pp 337–346.
[2] ——, *Factorization Methods for Discrete Sequential Estimation.* New York: Academic, 1977.
[3] T. J. Broida and R. Chellappa, "Estimation of object motion parameters from noisy images," *IEEE Trans. Pat. Anal. Machine Intell.*, vol. PAMI-8, no. 1, pp. 90–99, 1986.
[4] J. R. Bergen *et al.*, "Dynamic Analysis of Multiple Motion," in *Proc. Israeli Conf. AI Compute. Vision* (Tel Aviv), Dec. 1990.
[5] D. DeMenthon, "A zero-bank algorithm for inverse perspective of a road from a single image," in *Proc. IEEE Int. Conf. Robotics Automat.* (Raleigh, NC), 1987, pp. 1444–1449.
[6] E. D. Dickmanns and A. Zapp, "Guiding land vehicles along roadways by computer vision," in *Proc. Congres Automatique* (Toulouse), 1985, pp. 233–244.
[7] ——, "A curvature-based scheme for improving road vehicle guidance by computer vision," in *Mobile Robots.* Cambridge, MA: SPIE, 1986, pp. 161–168, vol. 727.
[8] ——, "Autonomous high speed road vehicle guidance by computer vision," in *Proc. 10th IFAC World Congress* (Munich), 1987, pp. 232–237.
[9] E. D. Dickmanns, "Dynamic computer vision for mobile robot control," in *Proc. 19th Int. Symp. Expos. Robots* (Sydney, Australia), Nov. 1988.
[10] ——, "Computer vision for flight vehicles," *ZFW*, vol. 12, no. 88, pp. 71–79, 1988.
[11] E. D. Dickmanns and V. Graefe, "Dynamic monocular machine vision," *Machine Vision Applications*, vol. 1, pp. 223–240, 1988; "Applications of dynamic monocular machine vision, *Machine Vision Applications*, vol. 1, pp. 241–261, 1988.
[12] E. D. Dickmanns and T. Christians, "Relative 3D state estimation for autonomous visual guidance of road vehicles," in *Intelligent Autonomous Systems 2* (T. Kanade *et al.*, Eds.). Amsterdam, Dec. 1989.
[13] E. D. Dickmanns, "Dynamic vision for intelligent motion control," in *Proc. IEEE-Workshop Intell. Motion Contr.* (Istanbul), Aug. 1990.
[14] G. Eberl, "Automatischer Landeanflug durch Rechnersehen," Dissertation, Fakultät für Luft- und Raumfahrttechnik der Universität der Bundeswehr Munchen, 1987.

[15] D. B. Gennery, "Tracking known three-dimensional objects," in *Proc. Amer. Assoc. Artificial Intell.* (Pittsburgh, PA), 1982, pp. 13–17.
[16] V. Graefe, "Two multiprocessor systems for low-level real-time computer vision," in *Robotics and Artificial Intelligence* (J. M. Brady, *et al.*, Eds.). Berlin, Springer-Verlag, 1984, pp. 301–308.
[17] C. Harris and C. Stennett, "RAPID—A video rate object tracker," in *Proc. BMVC 90* (Oxford, UK), Sept. 1990, pp 73–78.
[18] J. Heel, "Direct estimation of structure and motion from multiple frames," Mass. Inst. Technol., Cambridge, MA, AI Memo 1190, 1990.
[19] R. E. Kalman, "A new approach to linear filtering and prediction problems," *Trans. ASME Series D J. Basic Eng.*, 1960, pp. 35–45.
[20] K. P. Karmann, "Time recursive motion estimation using dynamical models for motion prediction," *ICPR*, vol. I, pp. 268–270, 1990.
[21] K. Kluge and C. Thorpe, "Explicit models for robot road following," in *Proc. IEEE Int. Conf. Robotics Automat.* (Scottsdale, AZ), 1989, pp. 1148–1154.
[22] P. S. Maybeck, *Stochastic Models, Estimation and Control, Vol. 1.* New York: Academic, 1979.
[23] H. G. Meissner and E. D. Dickmanns, "Control of an unstable plant by computer vision," in *Image Sequence Processing and Dynamic Scene Analysis* (T. S. Huang, Ed). Berlin, Springer-Verlag, 1983, pp. 532–548.
[24] B. Mysliwetz and E. D. Dickmanns, "Distributed scene analysis for autonomous road vehicle guidance," in *Proc. SPIE Conf. Mobile Robots* (Cambridge, MA), Nov. 1987, vol. 852.
[25] B. Mysliwetz, "Parallelrechner-basierte Bildfolgeninterpretation zur autonomen Fahrzeugsteuerung," Dissertation, Universität der Bundeswehr, Munchen, LRT, 1990.
[26] S. Ozawa and A. Rosenfeld, "Synthesis of a road image as seen from a vehicle," *Patt. Recogn*, vol. 19, no. 2, pp. 123–145, 1985.
[27] S. Pele and H. Rom, "Motion based segmentation," in *Proc. Int. Conf. Patt. Recogn.* (Atlantic City, NJ), June 1990, pp. 109–113.
[28] S. B. Pollard, J. Porril, and J. E. W. Mayhew, "Experiments in vehicle control using predictive feed-forward stereo," in *Robotics Research. Fifth Int. Symp.* (H. Miura and S. Arimoto, Eds.). Cambridge, MA: MIT Press, 1990, pp. 174–180.
[29] P. Rives, E. Breuil, and B. Espiau, "Recursive estimation of 3-D features using optical flow and camera motion," in *Proc. Conf. Intell. Autonomous Syst.* (Amsterdam), Dec. 1986, pp. 522–532.
[30] K. Sakurai, H. Zen, H. Okta, Y. Ushioda, and S. Ozawa, "Analysis of a road image as seen from a vehicle," *IEEE-ICCV*, pp. 651–656, 1987.
[31] F. -R. Schell and E. D. Dickmanns, "Autonomous automatic landing through computer vision," in *AGARD Conf. Proc. No 455: Adv. Tech. Technol. Air Vehicle Navigation Guidance* (Lissabon), 1989.
[32] H. J. Wuensche, "Bewegungssteuerung durch Rechnersehen, Fachberichte Messen, Steuern, Regeln. Berlin: Springer-Verlag, 1988, vol. 20.
[33] Z. Zhang and O. D. Faugeras, "Tracking and motion estimation in a sequence of stereo frames," in *ECAI 90, Proc. 9th Euro. Conf. Artificial Intell.* (L. C. Aiello, Ed.). London: Pitman, 1990, pp 747–752.
[34] E. D. Dickmanns, B. Mysliwetz, and T. Christians, "An integrated spatio-temporal approach to automatic visual guidance of autonomous vehicles," *IEEE Trans. Syst. Man Cybern.*, vol. 20, no. 6, pp. 1273–1284, 1990.



Ernst D. Dickmanns studied aerospace engineering at the RWTH, Aachen, Germany, from 1956 to 1961 and control engineering at Princeton University, Princeton, NJ, from 1964 to 1965. He received the Dr.-Ing. degree from RWTH in 1969. From 1971 to 1972, he was a postdoctorate research associate with NASA-MSFC, Huntsville, AL, working on shuttle orbiter flight dynamics.

He has been with the German aerospace research organization DFVLR from 1961 until 1975, working in optimal control for atmospheric and satelite vehicles like optimal rocket ascent, positioning of geostationary satellites, and reentry of aerobraking vehicles. Since 1975, he has been a full professor for control engineering at the Department of Aerospace Technology, Universität der Bundeswehr, Munich. His main research area since the late 1970's has been real-time machine vision for intelligent motion control. The 4-D approach developed by his group, which is based on recursive estimation theory, has brought a major advancement to the field of real-time dynamic scene understanding.

Birger D. Mysliwetz was born in Seattle, WA, in 1958. He received the Diplom-Ingenieur degree in electrical engineering in 1984 from the Technical University in Munich, Germany.

From 1984 to 1990, he was with the Department of Aerospace Engineering of the Universität der Bundeswehr, Munich, where he earned his Ph.D. degree. His research interests include parallel architectures and distributed, model-based approaches to real-time image sequence interpretation. Since 1991, he has been with Baasel Lasertech, which is a laser system company in Starnberg, Germany.