Vehicles Capable of Dynamic Vision

Ernst D. Dickmanns
Universitaet der Bundeswehr, Munich
D-85577 Neubiberg, Germany
e-mail: Ernst.Dickmanns@unibw-muenchen.de

Abstract

A survey is given on two decades of developments in the field, encompassing an increase in computing power by four orders of magnitude. The '4-D approach' integrating expectation-based methods from systems dynamics and control engineering with methods from AI has allowed to create vehicles with unprecedented capabilities in the technical realm: Autonomous road vehicle guidance in public traffic on freeways at speeds beyond 130 km/h, on-board-autonomous landing approaches of aircraft, and landmark navigation for AGV's, for road vehicles including turn-offs onto cross-roads, and for helicopters in low-level flight (real-time, hardware-in-the-loop simulations in the latter case).

1 Introduction

Road vehicle guidance based on video-signal processing has been picked up independently in Japan [Tsugawa et al., 1979], in Europe [Meissner, 1982], and in the USA [Klass, 1985]. While in Japan analog signal processing has been used and (quasi-steady) AI-methods predominated in the US, recursive estimation methods well known from systems engineering have been extended to image sequence processing at the author's institute (UBM); the resulting method had been dubbed '4-D approach', in contrast to the 2-D, 2.5-D, and 3-D methods under discussion then, disregarding time as the fourth independent variable in the problem domain. The numerical efficiency and compactness in state representation of recursive estimation which directly allowed control applications for generating behavioral capabilities, finally, led to its wide-spread acceptance in the vision community. Artificial neural nets (ANN) also found wide acceptance in the USA [Pomerleau, 1992] and around the globe even though image resolution used (about 1000 pixel = IKpel), usually, was much less than with recursive estimation (80 Kpel per image, even at a higher image rate).

Both methods allowed road vehicles to run autonomously along highways and other types of roads up to rather high speeds, initially on empty roads only [Dickmanns and Zapp, 1987, Pomerleau, 1989] but finally in normal freeway traffic also [Dickmanns et al., 1994, Pomerleau, 1992]; however, while ANN's stayed confined to either lateral [Pomerleau, 1992; Mecklenburg et al., 1992] or longitudinal control [Fritz, 1996] at a time (the other mode had to be controlled by a human driver), the 4-D approach allowed to detect, track and determine the spatio-temporal state (position and velocity components on a 3-D surface) relative to about a dozen other objects in a range of up to 100 meters in front of and behind the own vehicle pickmanns, 1995a]. The two final demonstrator vehicles in the European project Prometheus: VITA_2 of Daimler-Benz and VaMP of UBM [Ulmer, 1994; Dickmanns et al., 1994], may well be considered as the first two road vehicles of a new species capable of understanding (part of) their environment and of reacting properly to the actual needs on their own (completely autonomous).

Dynamic remote sensing for intelligent motion control in an environment with rapidly changing elements requires the use of valid spatio-temporal models for efficient handling of the large data streams involved. Other objects have to be recognized with their relative motion components, the near ones even with high precision for collision avoidance; this has to be achieved while the own vehicle body carrying the cameras moves in an intended way and is, simultaneously, subject to perturbations hardly predictable.

For this complex scenario, inertial sensing in addition to vision is of great help; negative angular rate feedback to a viewing direction control device allows to stabilize the appearance of stationary objects in the image sequence. Measured accelerations and velocities will, via signal integration, yield predictions for.translational and rotational positions affecting the perspective mapping process. These predictions are good in the short run, but may drift slowly in the long run, especially when inexpensive inertial sensors are used. These drifts, however, can easily be compensated by visual interpretation of static scene elements.

2 Simultaneous representations on differential and multiple integral scales

Combined use of inertial and visual sensing is well known from biological systems, e.g. the vestibular apparatus and its interconnections to eyes in vertebrates. In order to make optimal use of inertial and visual signals, simultaneous differential and integral representations on different scales both in space and in time are being exploited; table 1 shows the four categories introduced: The upper left corner represents the point 'here and now' in space and time where all interactions of a sensor or an actuator with the real world take place. Inertial sensors yield information on local accelerations (arrow 1 from field (1,1) to field (3,3) in the table) and turn rates of this point. Within a rigid structure of an object the turn rates are the same all over the body; therefore, the inertially measured rate signals (arrow 2 from field (1,3) to (3,3)) are drawn on the spatial object level (row 3).

The local surface of a structure may be described by the change of its tangent direction along some arc length; this is called curvature and is an element of local shape. It is a geometrical characterization of this part of the object in differential form; row 2 in table 1 represents these local spatial differentials which may cause specific edge features (straight or curved ones) in the image under certain aspect conditions.

Single objects may be considered to be local spatial integrals (represented in row 3 of table 1), the shapes of which are determined by their spatial curvature distributions on the surface; in connection with the aspect conditions and the photometric properties of the surface they determine the feature distribution in the image. Since, in general, several objects may be viewed simultaneously, also these arrangements of objects of relevance in a task context, called 'geometrical elements of a situation', are perceived and taken into account for behavior decision and reactive control. For this reason, the visual data input labeled by the index 3 at the corresponding arrows into the central interpretation process, field (3,3), has three components: 3a) for measured features not yet associated with an object, the so-

range in time → ↓ in space	point in time	temporally local differential environment	local time integrals basic cycle time	extended local time integrals	→	global time integrals
point in space	'here and now local measurements	temporal change at point 'here' (avoided because of noise amplification)	single step transition matrix derived from notion of (local) 'objects' (row 3)	———		———
spatially local differential environment	differential geometry: edge angles, positions curvatures	1 " 3a	transition of feature parameters 5	feature history		———
local space integrals → objects	object state feature- distribution, shape 4	2 motion 3b constraints: diff. eqs., 'dyn. model'	state transition, changed aspect conditions 'central hub'	short range predictions, > .... object state history	....	sparse predictions, > object state history
maneuver space of objects	local 3c situation	'lead'- information for efficient controllers	single step prediction of situation (usually not done)	6 multiple step prediction of situation; monitoring of maneuvers		———
↓ :						
mission space of objects	actual global situation	———	———	- monitoring		7 mission performance, monitoring

Table 1: Differential and integral representations on different scales for dynamic perception

called detection component; 3b) the object-oriented tracking component with a strong predictive element for efficiency improvement, and 3c) the perception component for the environment which preshapes the maneuver space for the self and all the other objects. Looked at this way, vision simultaneously provides geometrical information both on differential (row 2) and integral scales (rows: 3 for a single objects, 4 for local maneuvering, and 5 for mission performance).

Temporal change is represented in column 2 which yields the corresponding time derivatives to the elements in the column to the left. Because of noise amplification associated with numerical differentiation of high frequency signals $(d/dt(A\ \sin(\omega t) = A\ \omega\ \cos(\omega t))$, this operation is usable only for smooth signals, like for computing speed from odometry; especially, it is avoided deliberately to do optical flow computation at image points. Even on the feature level, the operation of integration with a smoothing effect, as used in recursive estimation, is preferred.

In the matrix field (3,2) of table 1 the key knowledge elements and the corresponding tools for sampled data processing are indicated: Due to mass and limited energy availability, motion processes in the real world are constrained; good models for unperturbed motion of objects belonging to specific classes are available in the natural and engineering sciences which represent the dependence of the temporal rate of change of the state variables on both the state- and the control variables. These are the so-called 'dynamical models'. For constant control inputs over the integration period, these models can be integrated to yield difference equations which link the states of objects in column 3 of table 1 to those in column 1, thereby bridging the gap of column 2; in control engineering, methods and libraries with computer codes are available to handle all problems arising. Once the states at one point in time are known, the corresponding time derivatives are delivered by these models.

Recursive estimation techniques developed since the 60ies exploit this knowledge by making state predictions over one cycle disregarding perturbations; then, the measurement models arc applied yielding predicted measurements. In the 4-D approach, these are communicated to the image processing stage in order to improve image evaluation efficiency (arrow 4 from field (3,3) to (1,3) in table 1 on the object level, and arrow 5 from (3,3) to (2,3) on the feature extraction level). A comparizon with the actually measured features then yields the prediction errors used for state update.

In order to better understand what is going to happen on a larger scale, these predictions may be repeated several to many times in a very fast in advance simulation assuming likely control inputs, for stereotypical maneuvers like lane changes in road vehicle guidance, a finite sequence of 'feed-forward' control inputs is known to have a longer term state transition effect. These are represented in field (4,4) of table 1 and by arrow 6; section 6 below will deal with these problems.

For the compensation of perturbation effects, direct state feedback well known from control engineering is used. With linear systems theory, eigenvalues and damping characteristics for state transition of the closed loop system can be specified (field (3,4) and row 4 in table 1). This is knowledge also linking differential representations to integral ones; low frequency and high frequency components may be handled separately in the time or in the frequency domain (Laplace-transform) as usual in aero-space engineering. This is left open and indicated by the empty row and column in table 1.

The various feed-forward and feedback control laws which may be used in superimposed modes constitute behavioral capabilities of the autonomous vehicle. If a sufficiently rich set of these modes is available, and if the system is able to recognize situations when to activate these behavioral capabilities with which parameters for achieving mission goals, the capability for autonomous performance of entire missions is given. This is represented by field (n,n) (lower right) and will be discussed in sections 6 to 8. Essentially, mission performance requires proper sequencing of behavioral capabilities in the task context; with corresponding symbolic representations on the higher, more abstract system levels, an elegant symbiosis of control engineering and AI-methods can thus be realized.

3 Task domains

Though the approach is very general and has been adapted to other task domains also, only road and air vehicle guidance will be discussed here.

3.1 Road vehicles

The most well structured environments for autonomous vehicles are freeways with limited access (high speed vehicles only) and strict regulations for construction parameters like lane widths, maximum curvatures and slopes, on- and off-ramps, no same level crossings. For this reason, even though speed driven may be high, usually, freeway driving has been selected as the first task domain for autonomous vehicle guidance by our group in 1985.

Six perceptual and behavioral capabilities are sufficient for navigation and mission performance on freeways: 1. Lane recognition and lane following with adequate speed, 2. obstacle recognition and proper reaction, e.g. transition into convoy driving or stopping; 3. recognition of neighboring lanes, their availability for lane change, and lane changing performance; 4. reading and obeying traffic signs, 5. reading and interpreting navigation information including proper lane selection, and 6. handling entries and exits.

On well kept freeways it is usually not necessary to check surface structure or to watch for humans or animals entering from the side. None-the-less, safe reaction to unexpected events must be required for a mature autonomous system.

On normal state roads the variability of road parameters and of traffic participants is much larger; especially, same level crossings and oncoming traffic increase relative speed between objects, thereby increasing hazard potential even though traveling speed may be limited to a much lower level. Bicyclists and pedestrians as well as many kinds of animals are normal traffic participants. In addition, lane width may be less in the average, and surface state may well be poor on lower order roads, e.g. potholes, especially in the transition zone to the shoulders.

In urban traffic, things may be even worse with respect to crowdedness and crossing of subjects. These latter mentioned environments are considered to be not yet amenable to autonomous driving because of scene complexity and computing performance required.

However, driving on minor roadways with little traffic, even without macadam or concrete sealing, has been attacked for research purposes in the past, and may soon be performed safely with the increasing computing power becoming available now. If it is known that the ground is going to support the vehicle, even cross country driving can be done including obstacle avoidance. However, if compared to human capabilities in these situations, there is still a long way to go until autonomous systems can compete.

3.2 Air vehicles

As compared to ground vehicles with 3, full 6 degrees of freedom are available for trajectory shaping, here. In addition, due to air turbulence and winds, the perturbation environment may be much harder than on roads. For this reason, inertial sensing is considered mandatory in this task domain, in addition, visual navigation guidelines like lanes on roads are not available once the aircraft is airborne at higher altitudes. Microwave electronic guidelines have been established instead.

Vision allows the pilot or an autonomous aircraft to navigate relative to certain landmarks; the most typical task is the landing approach to a prepared runway for fixed wing aircraft, or to the small landing site usually marked by the large letter H for a helicopter. These tasks have been selected for first demonstrations of the capabilities of seeing aircraft. Contrary to other electronic landing aids like ILS or MLS, machine vision also allows to detect obstacles on the runway and to react in a proper manner.

For flights close to the Earth surface, terrain formations may be recognized as well as buildings and power lines, thus, obstacle avoidance in nap-of-the-Earth flights is a natural extension of this technique for unmanned air vehicles, both with fixed wings and for helicopters. For the latter, the capability of recognizing structures or objects on the ground and of hovering in a fixed position relative to these objects despite perturbations, will improve rescue capabilities and delivery performance.

Motion control for fixed wing aircraft and for helicopters is quite different from each other; by the use of proper dynamical models and control laws it has been shown that the 4-D approach allows to turn each craft into an autonomous agent capable of fully automatic mission performance. This will be discussed in section 8.

4 The sensory systems

The extremely high data rates of image sequences are both an advantage (with respect to versatility in acquiring new information on both environment and on other objects/subjects) and a disadvantage (with respect to computing power needed and delay time incurred until the information has been extracted from the data). For this reason it makes sense to also rely on conventional sensors in addition, since they deliver information on specific output variables with minimal time delay.

4.1 Conventional sensors

For ground vehicles, odometers, speedometers as well as sensors for positions and angles of subparts like actuators and pointing devices are commonplace. For aircraft, pressure measurement devices yield information on speed and altitude flown; here, inertial sensors like accelerometers, angular rate- and vertical as well as directional gyros arc standard. Evaluating this information in conjunction with vision alleviates image sequence processing considerably. Based on the experience gained in air vehicle applications, the inexpensive inertial sensors like accelerometers and angular rate sensors have been adopted for road vehicles too, because of the beneficial and complementary effects relative to vision. Part of this has already been discussed in section 2 and will be detailed below.

4.2 Vision sensors

Because of the large viewing ranges required, a single camera as vision sensor is by no means sufficient for practical purposes. In the past, bifocal camera arrangements (see fig.l) with a wide angle (about 45°) and a tele camera (about 15° aperture) mounted fix relative to each other on a two-axis platform for viewing direction control have been used [Dickmanns, 1995a]; in future systems, trinocular camera arrangements with a wide simultaneous field of view (> 100°) from two divergently mounted wide angle cameras and a 3-chip color CCD-camera will be used [Dickmanns, 1995b]. For high-speed driving on German Autobahnen, even a fourth camera with a relatively strong tele-lens will be added allowing lane recognition at several hundred meters distance.

All these data are evaluated 25 times per second, the standard European video rate.

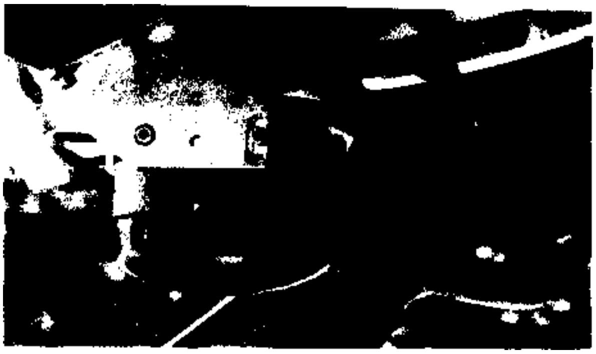


Figure 1: Binocular camera arrangement of VaMP

4.3 Global Positioning System (GPS-) sensor

For landmark navigation in connection with maps a GPS-receiver has been integrated into one system in order to have sufficiently good initial conditions for landmark detection. Even though only the least accurate C/A code is being used, in connection with inertial sensing and map interpretation good accuracies can be achieved after some time of operation [Furst et al, 1997]; GPS signals are available only once every second.

5 Spatio-temporal perception: The 4-D approach

Since the late 70ies, observer techniques as developed in systems dynamics [Luenberger, 1964] have been used at UBM in the field of motion control by computer vision [Meissner, 1982; Meissner and Dickmanns, 1983). In the early 80ies, H.J. Wuensche did a thorough comparison between observer- and Kalman filter realizations in recursive estimation applied to vision for the original task of balancing an inverted pendulum on an electro-cart by computer vision [Wuensche, 1983]. Since then, refined versions of the Extended Kalman Filter (EKF) with numerical stabilization (UDU^T-factorization, square root formulation) and sequential updates after each new measurement have been applied as standard methods to all dynamic vision problems at UBM.

Based on experience gained from 'satellite docking' [Wuensche, 1986], road vehicle guidance, and on-board autonomous aircraft landing approaches by machine vision, it was realized in the mid 80ies that the joint use of dynamical models and temporal predictions for several aspects of the overall problem in parallel was the key to achieving a quantum jump in the performance level of autonomous systems based on machine vision. Beside state estimation for the physical objects observed and control computation based on these estimated states it was the feedback of knowledge thus gained to the image feature extraction and to the feature aggregation level which allowed for an increase in efficiency of image sequence evaluation of one to two orders of magnitude. (See fig. 2 for a graphical overview.)

Following state prediction, the shape and the measurement models were exploited for determining:

- viewing direction control by pointing the two-axis platform carrying the cameras;
- locations in the image where information for most easy, non-ambiguous and accurate state estimation could be found (feature selection),
- the orientation of edge features which allowed to reduce the number of search masks and directions for robust yet efficient and precise edge localization,
- the length of the search path as function of the actual measurement uncertainty,
- strategies for efficient feature aggregation guided by the idea of the 'Gestalt' of objects ,and
- the Jacobian matrices of first order derivatives of feature positions relative to state components in the dynamical models which contain rich information for interpretation of the motion process in a least squares error sense, given the motion constraints, the features measured, and the statistical properties known.

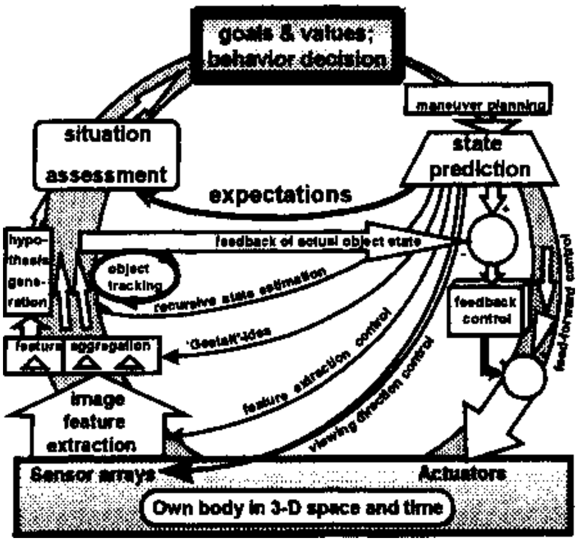


Figure 2: Multiple feedback loops on different space scales for efficient scene interpretation and behavior control: control of image acquisition and -processing (lower left corner), 3-D 'imagination-space in upper half; motion control (lower right corner).

This integral use of

1. dynamical models for motion of and around the center of gravity taking actual control outputs and time delays into account,
2. spatial (3-D) shape models for specifying visually measurable features,
3. the perspective mapping models, and
4. prediction error feedback for estimation of the object state in 3-D space and time

simultaneously and in closed loop form was termed the '4-D approach'. It is far more than a recursive estimation algorithm based on some arbitrary model assumption in some arbitrary subspace or in the image plane.

It *is* estimated from a scan of recent publications in the field that even today most of the papers referring to 'Kalman filters' do not take advantage of this integrated use of spatio-temporal models based on physical processes.

Initially, in our applications just the ego-vehicle has been assumed to be moving on a smooth surface or trajectory, with the cameras fixed to the vehicle body. In the meantime, solutions to rather general scenarios are available with several cameras spatially arranged on a platform which may be pointed by voluntary control relative to the vehicle body. These camera arrangements allow a wide simultaneous field of view, a central area for trinocular (skew) stereo interpretation, and a small area with high image resolution for 'tele'-vision. The vehicle may move in full 6 degrees of freedom; while moving, several other objects may move independently in front of a stationary background. One of these objects may be 'fixated' (tracked) by the pointing device using inertia! and visual feedback signals for keeping the object (almost) centered in the high resolution image. A newly appearing object in the wide field of view may trigger a fast viewing direction change such that this object can be analysed in more detail by one of the tele-cameras; this corresponds to 'saccadic' vision as known from vertebrates and allows very much reduced data rates for a complex sense of vision. It essentially trades the need for time-sliced attention control and sampled-data based scene reconstruction against a data rate reduction of 1 to 2 orders of magnitude as compared to full resolution in the entire simultaneous field of view.

The 4-D approach lends itself for this type of vision since both object-orientation and the temporal ('dynamical') models are available in the system already. This complex system design for dynamic vision has been termed EMS-vision (from Expectation-based, Multi-focal and Saccadic); it is actually being implemented with an experimental set of four miniature TV-cameras on a two-axis pointing platform named 'Multi-focal active/reactive Vehicle Eye' MarVEye Pickmanns, 1995b].

In the rest of the paper, major developmental steps in the 4-D approach over the last decade and results achieved will be reviewed. As an introduction, in the next section we summarize the basic assumptions underlying the 4-D approach.

5.1 Basic assumptions underlying the 4-D approach

It is the explicit goal of this approach to take, as much as possible, advantage of physical and mathematical models of processes happening in the real world. Models developed in the natural sciences and in engineering over the last centuries, in simulation technology and in systems engineering (decision and control) over the last decades form the base for computer-internal representations of real-world processes:

1. The (mesoscopic) world observed happens in 3-D space and time as the independent variables; non-relativistic (Newtonian) models are sufficient for describing these processes.
2. All interactions with the real world happen *'here and now'* , at the location of the body carrying special input/ouput devices; especially the locations of the sensors (for signal or data input) and of the actuators (for control output) as well as those body regions with strongest interaction with the world (as for example the wheels of ground vehicles) are of highest importance.
3. Efficient interpretation of sensor signals requires background knowledge about the *processes* observed and controled, that is both its spatial and temporal characteristics. Invariants for process understanding may be abstract model components not graspable at one point in time. Similarly,
4. efficient computation of (favorable or optimal) control outputs can only be done taking complete (or partial) process models into account, control theory provides the methods for fast and stable reactions.
5. Wise behavioral decisions require knowledge about the longer-term outcome of special feed-forward or feedback control modes in certain situations and environments; these results are obtained from integration of the dynamical models. This may have been done beforehand and stored appropriately, or may be done on the spot if analytical solutions are available or numerical ones can be derived in a small fraction of real-time as becomes possible now with the increasing processing power available. Behaviors are realized by triggering the modes available from point 4 above.
6. Situations are made up of arrangements of objects, other active subjects, and of the own goals pursued; therefore,
7. it is essential to recognize single objects and subjects, their relative state, and for the latter also, if possible, their intentions in order to be able to make meaningful predictions about the future development of a situation (which is needed for successful behavioral decisions).

8. As the term re-cognition tells, in the usual case it is assumed that objects seen are (at least) generically known already, only their appearance here (in the geometrical range of operation of the senses) and now is new; this allows a fast jump to an object hypothesis when first visual impressions arrive through sets of features. Exploiting background knowledge, the model based perception process has to be initiated. Free parameters in the generic object models may be determined efficiently by attention control and the use of special algorithms and behaviors.
9. In order to be able to do step 8 efficiently, knowledge about 'the world' has to be provided in the context of 'task domains' in which likely co-occurrences are represented. In addition, knowledge about discriminating features is essential for correct hypothesis generation (indexing into the object data base).
10. Most efficient object (class) descriptions by invariants is usually done in 3-D space (for shape) and time (for motion constraints or stereotypical motion sequences); modern microprocessors are sufficiently powerful to compute the visual appearance of an object under given aspect conditions in an image (in a single one, or even in several ones with different mapping parameters in parallel) at runtime. They are even powerful enough to numerically compute the Jacobian matrices for sensor/object pairs of features evaluated with respect to object state or parameter values; this allows a very flexible general framework for recursive state and parameter estimation. The inversion of perspective projection is thus reduced to a least squares model fit once the recursive process has been started. The underlying assumption here is that local linearizations of the overall process are sufficiently good representations of the nonlinear real process; for high evaluation rates like video frequency (25 or 30 Hz) this is usually the case.
11. In a running interpretation process of a dynamic scene, newly appearing objects will occur in restricted areas of the image such that bottom-up search processes may be confined to these areas. Passing cars, for example, always enter the field of view from the side just above the ground; a small class of features allows to detect them reliably.
12. Subjects, i.e. objects with the capability of self induced generation of control actuation, are characterized by typical (sometimes stereotypical, i.e. predictive) motion behavior in certain situations. This may also be used for recognizing them (similar to shape in the spatial domain).
13. The same object/subject may be represented internally at different scales with various degrees of detail; this allows flexible and efficient use in changing contexts (e.g. as a function of distance or degree of attention).

5.2 Structural survey on the 4-D approach

Figure 3 shows the main three activities running in parallel in an advanced version of the 4-D approach:

1. Detection of objects from typical collections of features not yet assigned to some object already tracked (center left, upward arrow); when these feature collections are stable over several frames, an object hypothesis has to be formed and the new object is added to the list of those regularly tracked (arrow to the right).
2. Tracking of objects and state estimation is shown in the loop to the lower right in figure 3; first, with the control output chosen, a single step prediction is done in 3-D space and time, the 'imagined real world'. This step consists of two components, a) the 'where'- signal path concentrating on progress of motion in both translational and rotational degrees of freedom, and b) the 'what'- signal path dealing with object shape. (In order not to overburden the figure these components are not shown.)
3. Learning from observation is done with the same data as for tracking; however, this is not a single step loop but rather a low frequency estimation component concentrating on 'constant' parameters, or it even is an offline component with batch processing of stored data. This is an actual construction site in code development at present which will open up the architecture towards becoming more autonomous in new task domains as experience of the system grows. Both dynamical models (for the 'where'-part) and shape models (for the 'what'-part shall be learnable.

Another component under development not detailed in figure 3 is situation assessment and behavior decision; this will be discussed in section 6.

5.3 Generic 4-D object classes

The efficiency of the 4-D approach to dynamic vision is achieved by associating background knowledge about classes of objects and their behavioral capabilities with the data input. This knowledge *is* available in generic form, that is, structural information typical for object classes is fixed while specific parameters in the models have to be adapted to the special case at hand. Motion descriptions for the center of gravity (the translational object trajectory in space) and for rotational movements, both of which together form the so-called 'where'-problem, are separated from shape descriptions, called the 'what'-problem. Typically, summing and averaging of feature positions is needed to solve the where-problem while differencing feature positions contributes to solving the what-problem.

Motion description

Possibilities for object trajectories are so abundant that they cannot be represented with reasonable effort. However, good models are usually available describing their evolution

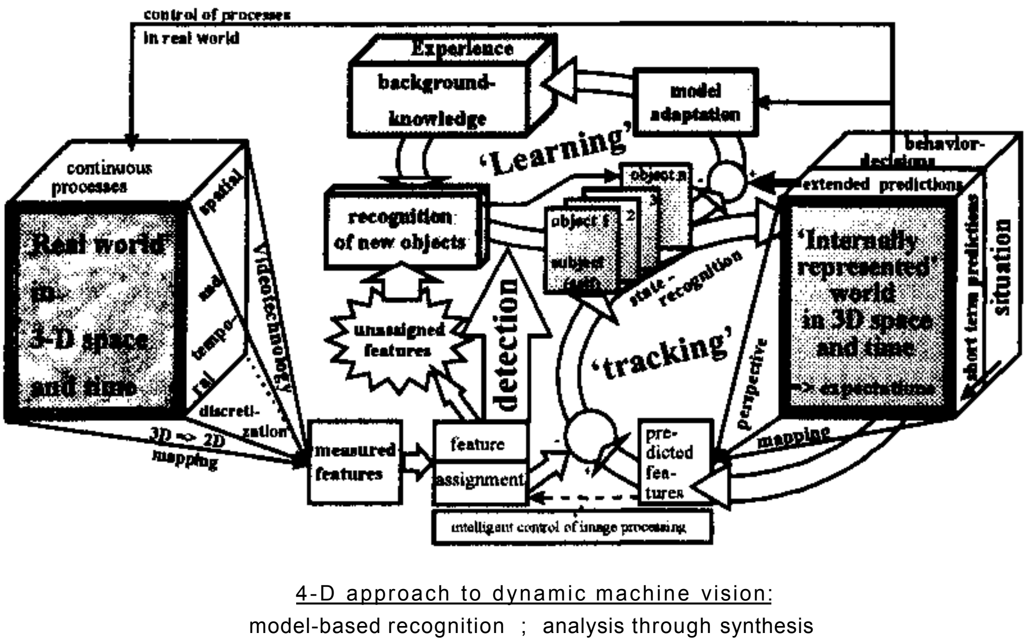


Figure 3: Survey on the 4-D approach to dynamic machine vision with three major areas of activity: Object detection (central arrow upwards), tracking and state estimation (recursive loop in lower right), and learning (loop in center top), the latter two being driven by prediction error feedback.

over time as a function of the actual state, the control- and the perturbation inputs. These so-called 'dynamical models', usually, are sets of nonlinear differential equations ($\dot{\underline{x}} = \underline{f}(\underline{x}$, u. v' t)) with x as the n-component state vector, u as r-componcnt control vector and v' as perturbation input.

Through linearization around a nominal trajectory $\underline{x}_N(t)$, locally linearized descriptions are obtained which can be integrated analytically to yield the (approximate) local transition matrix description for small cycle times T

$$x[(k+1)T] = A\, x[kT] + B\, u[kT] + v[kT]. \qquad (1)$$

The elements of the matrices A and B are obtained from $F(t) = \partial \underline{f}/\partial \underline{x}|_N$ and $G(t) = \partial \underline{f}/\partial \underline{u}|_N$ by standard methods from systems theory.

Usually, the states cannot be measured directly but through the output variables $\underline{y}$ given by

$$\underline{y}[kT] = h(\, \underline{x}[kT], \underline{p}, kT) + \underline{w}[kT], \qquad (2)$$

where h may be a nonlinear mapping (see below), $\underline{p}$ are mapping parameters and $\underline{w}$ represents measurement noise.

On the basis of eq.(1) a distinction between 'objects' proper and 'subjects' can be made: If there is no dependence on controls $\underline{u}$ in the model, or if this $\underline{u}(t)$ is input by another agent we speak of an 'object', controlled by a subject in the latter case. If $\underline{u}[kT]$ may be activated by some internal activity within the object, be it by pre-programmed outputs or by results obtained from processing of measurement data, we speak of a 'subject'.

Shape and feature description

With respect to shape, objects and subjects are treated in the same fashion. Only rigid objects and objects consisting of several rigid parts linked by joints have been treated; for elastic and plastic modeling see [DeCarlo and Metaxas, 1996]. Since objects may be seen at different ranges the appearance in the image may vary considerably in size. At large ranges the 3-D shape of the object, usually, is of no importance to the observer, and the cross-section seen contains most of the information for tracking. However, this cross-section depends on the angular aspect conditions; therefore, both coarse-to-fine and aspect-dependent modeling of shape is necessary for efficient dynamic vision. This will be discussed briefly for the task of perceiving road vehicles as they appear in normal road traffic.

Coarse-to-fine shape models in 2-D: Seen from behind or from the front at a large distance, any road vehicle may be adequately described by its encasing rectangle; this is convenient since this shape just has two parameters, width b and height h. Absolute values of these parameters are of no importance at larger distances; the proper scale may be inferred from other known objects seen, like road or lane width at that distance. Trucks (or buses) and cars can easily be distinguished. Our experience tells that even the upper limit and thus the height of the object may be omitted without loss of functionality (reflections in this spatially curved region of the car body together with varying environmental conditions may make reliable tracking of the upper body boundary very difficult); thus, a simple U-shape of unit height (corresponding to about 1 m turned out to be practical) seems to be sufficient until 1 to 2 dozen pixels can be found on a line crossing the object in the image. Depending on the focal length used, this corresponds to different absolute distances.

Fig. 4a shows this shape model. If the object in the image is large enough so that details may be distinguished reliably by feature extraction, a polygonal shape approximation as shown in fig. 4b or even with internal details (fig. 4c) may be chosen; in the latter case, area-based features like the licence plate, the tires or the signal light groups (usually in yellow or reddish color) may allow more robust recognition and tracking.

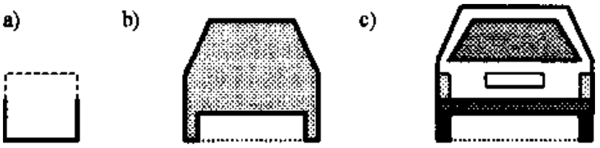


Figure 4: Coarse to fine shape model of a car in rear view:
a) encasing rectangle (U-shape); b) polygonal silhouette,
c) silhouette with internal structure.

If the view is from an oblique direction, the depth dimension (length of the vehicle) comes into play. Even with viewing conditions slightly off the axis of symmetry of the vehicle observed, the width of the car in the image will start increasing rapidly because of the larger length of the body and due to the sine-effect in mapping. Usually, it is impossible to determine the lateral aspect angle, body width and -length simultaneously from visual measurements; therefore, switching to the body diagonal as a shape representation has proven to be much more robust and reliable in real-world scenes [Schmid, 1994].

Just for tracking and relative state estimation, taking one of the vertical edges of the lower body and the lower bound of the object body has proven to be sufficient in most cases [Thomanek, 1996]; this, of course, is domain specific knowledge which has to be introduced when specifying the features for measurement in the shape model.

In general, modeling of well measurable features for object recognition has to be dependent on the aspect conditions. Experience tells that area based features should play an important role in robust object tracking. Initially, this has been realized by observing the average grey value on the vehicle-side of edge features detected; with more computing power available, color profiles in certain cross-sections yield improved performance.

Full 3-D models with different degrees of detail Similar to the 2-D rear silhouette, different models may also be used for 3-D shape. The one corresponding to fig. 4a is the encasing box with perpendicular surfaces; if these surfaces can be easily distinguished in the image, and their separation line may be measured precisely, good estimates of the overall body dimensions may be obtained from small image sizes already. Since space does not allow more details here, the interested reader is referred to [Schick and Dickmanns, 1991, Schmid 1995].

5.4 Image feature extraction

Due to space restrictions, this topic will not be detailed here; the interested reader is referred to [Dickmanns and Graefe, 1988] and an upcoming paper [Dickmanns et al., 1997]. Figure 5 shows a survey on the method used.

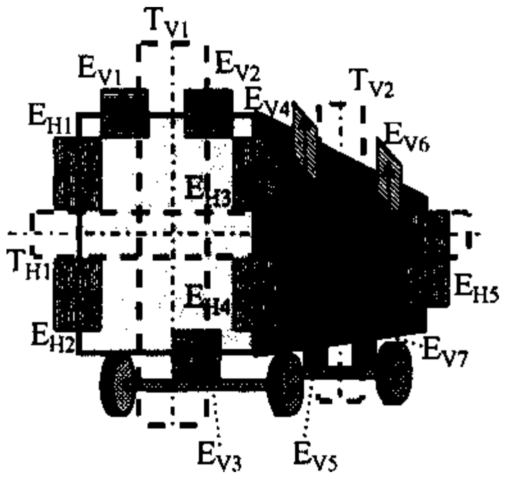


Figure 5: Intelligent control of image feature extraction parameters in the algorithms CRONOS (for edges, marked with a window lable E_{ij}) and Triangle' (labeled T, large rectangles with broken lines for efficient object tracking and state estimation in the 4-D approach

Two types of feature extraction algorithms are used: Oriented edge features extracted by ternary mask correlations in horizontal or vertical search paths (a rather old component), and area-based segmentations of 'stripes' of certain widths, arbitrarily oriented in the image plane (a new one).

The intelligent control of the parameters of these algorithms is essential for efficient tracking. In the 4-D approach, these parameters are set by predictions from the spatio-temporal representations and application of perspective mapping. From fig.5 it may be seen that a small percentage of image data properly analysed allows to track objects reliably and precisely when used in a tight bottom-up and top-down loop traversed frequently (25 Hz); this has to be seen in the context of figure 2.

5.5 State estimation

The basic approach has been described many times (see [Wuensche, 1986; Dickmanns, 1987; Dickmanns, 1992; Behringer, 1996; Thomanek, 1996]) and has remained the same for visual relative state estimation over years by now. However, in order to be able to better deal with the general case of scene recognition under (more strongly) perturbed ego-motion, an menially based component has been added [Werner et al., 1996; Werner, 1997].

This type of state estimation is not new at all if compared to inertial navigation, e.g. for missiles; however, here only very inexpensive accelerometers and angular rate sensors are being used. This is acceptable only because the resulting drift problems are handled by a visual state estimation loop running in parallel, thereby resembling the combined use of (relatively poor) inertial signals from the vestibular apparatus and of visual signals in vertebrate perception. Some of these inertial signals may also be used for stabilizing the viewing direction with respect to the stationary environment by direct negative feedback of angular rates to the pointing device carrying the cameras. This feedback actually runs at very high rates in our systems (500 Hz, see [Schiehlen, 1995]).

Inertially based ego-state estimation (IbSE)

The advantage of this new component is three-fold: 1 Because of the direct encoding of accelerations along, and rotational speed components around body fixed axes, time delays are negligeable. These components can be integrated numerically to yield predictions of positions. 2. The quantities measured correspond to the forces and moments actually exerted on the vehicle including the effects of perturbations; therefore, they are more valuable than predictions from a theoretical model disregarding perturbations which are unknown, in general. 3. If good models for the eigenbehavior are available, the inertial measurements allow to estimate parameters in perturbation models, thereby leading to deeper understanding of environmental effects.

Dynamic vision

With respect to ego-state recognition, vision now has reduced but still essential functionality. It has to stabilize longterm interpretation relative to the stationary environment, and it has to yield information on the environment, like position and orientation relative to the road and road curvature in vehicle guidance, not measurable inertially. With respect to other vehicles or obstacles, the vision task also is slightly alleviated since the high-frequency viewing direction component is known now; this reduces search range required for feature extraction and leads to higher efficiency of the overall system.

These effects can only be achieved using spatio-temporal models and perspective mapping, since these items link inertial measurements to features in the image plane. With different measurement models for all the cameras used, a single object model and its recursive iteration loop may be fed with image data from all cameras relevant. Jacobian matrices now exist for each object/sensor pair.

The nonlinear measurement equation (2) is linearized around the predicted nominal state XN and the nominal parameter set pN yielding (without the noise term)

$$\begin{aligned} \underline{y}[kT] &= \underline{y}_N[kt] + \underline{\delta y}[kt] \qquad (3)\\ &= \underline{h}(\,\underline{x}_N[kT], \underline{p}_N, kT) + C_x\,\underline{\delta x} + C_p\,\underline{\delta p}, \end{aligned}$$

where $C_x = \partial\underline{h}/\partial\underline{x}|_N$ and $C_p = \partial\underline{h}/\partial\underline{p}|_N$ are the Jacobian matrices with respect to the state components and the parameters involved. Since the first terms to the right hand side of the equality sign are equal by definition, eq. (3) may be used to determine $\underline{\delta x}$ and $\underline{\delta p}$ in a least squares sense from $\underline{\delta y}$ as the prediction error messured (observability given); this is the core of recursive estimation.

5.6 Situation assessment

For each object an estimation loop is set up yielding best estimates for the relative state to the ego-vehicle including all spatial velocity components. For stationary landmarks, the velocity is the negative of ego-speed, of course. Since this is known reliably from conventional measurements, the distance to the landmark can be determined even with monocular vision exploiting motion stereo [Hock, 1994; Thomanek, 1996;Muller, 1996].

With all this information available for the surrounding environment and the most essential objects in it, an interpretation process can evaluate the situation in a task context and come up with a conclusion whether to proceed with the behavioral mode running or to switch to a different mode. Fast in-advance simulations exploiting dynamical models and alternative stereotypical control inputs yield possible alternatives for the near-term evolution of the situation. By comparing the options or by resorting to precomputed and stored results, these decisions are made.

6 Generation of behavioral capabilities

Dynamic vision is geared to closed-loop behavior in a task context; the types of behavior of relevance, of course, depend on the special task domain. The general aspect is that behaviors are generated by control output. There are two basically different types of control generation:

1. Triggering the activation of (generically) stored time histories, so-called feed-forward control, by events actually observed, and

2. gearing actual control to the difference between desired and actual state of relevant systems, so-called feedback control.

In both cases, actual control parameters may depend on the situation given. A very general method is to combine the two given above (as a third case in the list), which is especially easy in the 4-D approach where dynamical models are already available for the part of motion understanding.

The general feed-forward control law in generic form is

$$\underline{u}(\tau) = \underline{g}(\underline{p}_M, \tau_M), \qquad \text{with } 0 < \tau = t - t_{Trig} < (\tau_M), \tag{4}$$

where pM may contain averaged state components (like speed).

A typical feed-forward control element is the steer control output for lane change: In a generic formulation, for example, the steer rate λ-dot is set in five phases during the maneuver time τ_M; the first and the final control phase of duration τ_P each, consist of a constant steer rate, say R. In the second and fourth phase of same duration, the amplitude is of opposite sign to the first and last one. In the third phase the steer rate is zero; it may be missing at all (duration zero). The parameters R, τ_M, τ_P have to be selected such that at $(\tau_M + \Delta\tau_D)$ the lateral offset is just one lane width with vehicle heading the same as before; these parameters, of course, depend on the speed driven.

Given this idealized control law, the corresponding state component time histories $\underline{x}_C(\tau)$ for $0 < \tau = t - t_{Trig} < (\tau_M + \Delta\tau_D)$ can be computed according to a good dynamical model; the additional time period $\Delta\tau_D$ at the end is added because in real dynamical maneuvers the transition is not completed at the time when the control input ends. In order to counteract disturbances during the maneuver, the difference $\Delta\underline{x}(\tau) = \underline{x}_C(\tau) - \underline{x}(\tau)$ may be used in a superimposed state feedback controller to force the real trajectory towards the ideal one.

The general state feedback control law is

$$\underline{u}(\tau) = -K^T \Delta\underline{x}(\tau), \tag{5}$$

with K being the r by n gain matrix. The gain coefficients may be set by pole placement or by a Riccati design (optimal linear quadratic controller) well known in control engineering [Kailath, 1980]. Both methods include knowledge about behavioral characteristics along the time axis: While pole placement specifies the eigenvalues of the closed loop system, the Riccati design minimizes weighted integrals of state errors and control inputs.

The simultaneous use of dynamical models for both perception and control and for the evaluation process leading to behavior decision makes this approach so efficient. Figure 6 shows the closed-loop interactions in the overall system.

Based on object state estimation (lower left corner) events arc detected (center left) and the overall situation is assessed (upper left). Initially, the upper level has to decide

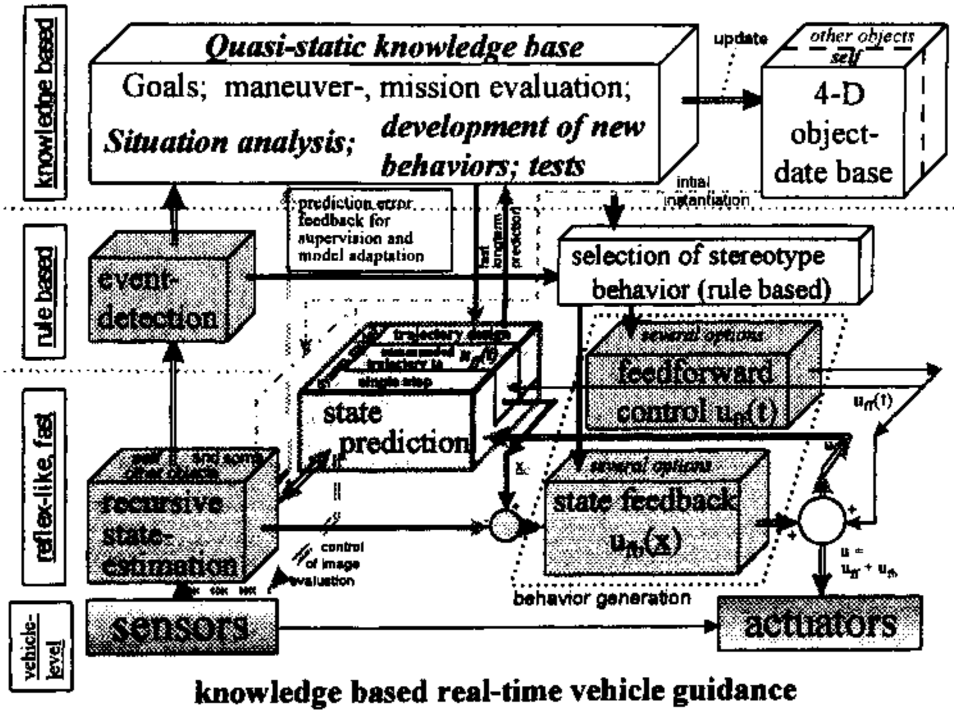

knowledge based
rule based
reflex-like, fast
vehicle-level
Quasi-static knowledge base
Goals; maneuver-, mission evaluation;
Situation analysis; development of new behaviors; tests
update
other objects
self
4-D object-date base
initial instantiation
prediction error feedback for supervision and model adaptation
fast
longterm
prediction
event-detection
selection of stereotype behavior (rule based)
trajectory design
single step
state prediction
several options
feedforward control $u_{ff}(t)$
$u_{ff}(t)$
x_c
several options
state feedback $u_{fb}(\underline{x})$
+
$u = u_{ff} + u_{fb}$
recursive state-estimation
control of image evaluation
behavior generation
sensors
actuators
knowledge based real-time vehicle guidance

Figure 6: Knowledge based real-time control system with three hierarchical levels and time-horizons.

which of the behavioral capabilities available are to be used: Feed-forward, feedback, or a superposition of both; lateron, the feedback loops activated are running continuously (lower part in fig. 6 with horizontal texture) without intervention from the upper levels, except for mode changes. Certain events also may trigger feed-forward control outputs directly (center right).

Since the actual trajectory evolving from this control input may be different from the nominal one expected due to unforseeable perturbations, commanded state time histories $\underline{x}_C(\tau)$ are generated in the block 'state prediction' (center of fig. 6, upper right central part) and used as reference values for the feedback loop (arrow from top at lower center). In this way, combining feed-forward direct control and actual error feedback, the system will realize the commanded behavior as close as possible and deal with perturbations without the need for replanning on the higher levels.

All, that is needed for mission performance of any specific system then is a sufficiently rich set of feed-forward and feedback behavioral capabilities. These have to be activated in the right sequence such that the goals are achieved in the end. For this purpose, the effect of each behavioral capability has to be represented on the upper decision level by global descriptions of their effects:

1. For feed-forward behaviors with corrective feedback superimposed (case 3 given above) it is sufficient to just represent initial and final conditions including time needed; note that this is a quasi-static description as used in Al-methods. This level does not have to worry about real-time dynamics, being taken care off by the lower levels. It just has to know in which situations these behavioral capabilities may be activated with which parameter set.
2. For feedback behaviors it is sufficient to know when this mode may be used; these reflex-like fast reactions may run over unlimited periods of time if not interrupted by some special event. A typical example is lane following in road vehicle guidance; the integral of speed then is the distance traveled, irrespective of the curvatures of the road. These values are given in information systems for planning, like maps or tables, and can be used for checking mission progress on the upper level.

Performing more complex missions on this basis has just begun. The newly available computing power will lead to quick progress on this mission level, now that the general concept has been defined.

7 Multiple loops in dynamic scene understanding

The principles discussed above have lead to parallel realizations of multiple loops in the interpretation process both in space and in time; figure 2 has displayed the spatial aspects. In the upper half of the figure, the essential scales for feedback loops are the object level, the local situation level, and the global mission performance level on which behavior decisions for achieving mission goals are being done (see table 1 also).

These decisions may be based on both local and extended predictions of the actual situation and on knowledge about behavioral capabilities of the own vehicle and of other subjects in the scene. The multiple loops used in our system in the time domain are displayed in figure 7; they range from the millisecond scale for inertial viewing direction control to several hours for ground and flight vehicles on the mission scale encompassing sequences of maneuvers and feedback behavioral modes.

The outermost two loops labeled 'quasi-static' are closed, up to now, mainly by human operators and software developers. They are being tackled now for automation on the system structure developed; it is felt that a unified approach encompassing systems dynamics, control engineering, computer simulation and animation techniques as well as methods from AI has become feasible.

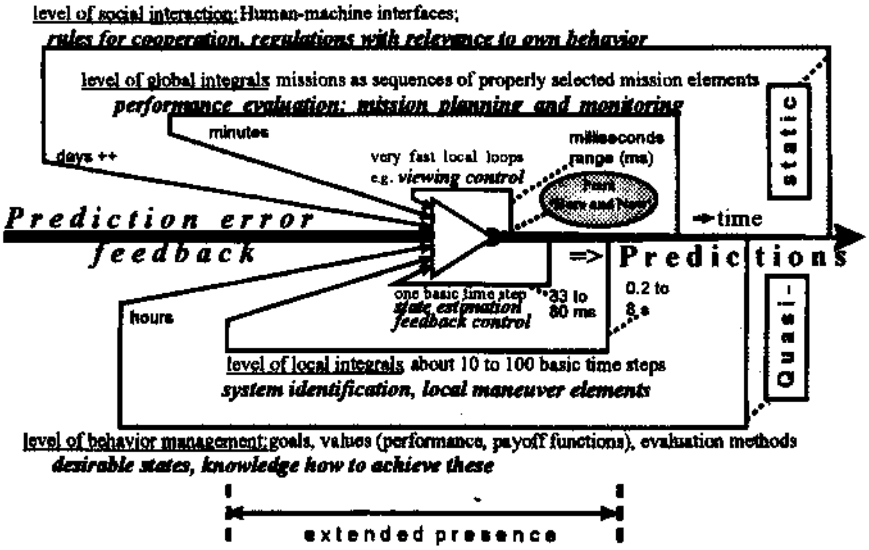

level of social interaction: Human-machine interfaces;
rules for cooperation, regulations with relevance to own behavior
level of global integrals missions as sequences of properly selected mission elements
performance evaluation: mission planning and monitoring
minutes
days ++
very fast local loops
e.g. viewing control
milliseconds
range (ms)
Point 'Here and Now'
Prediction error
feedback
time
=> Predictions
static
Quasi-
one basic time step
state estimation
feedback control
33 to
80 ms
0.2 to
8 s
hours
level of local integrals about 10 to 100 basic time steps
system identification, local maneuver elements
level of behavior management: goals, values (performance, payoff functions), evaluation methods
desirable states, knowledge how to achieve these
extended presence

Figure 7: Multiple feedback loops on different time scales in (visual) cognition systems and corresponding representational levels

8 Experimental results

8.1 Road vehicles

The autonomous road vehicle VaMP (see figure 8) and its twin VITA II of Daimler-Benz have shown remarkable performance in normal freeway traffic in France, Germany and Denmark since 1994. VaMP has two pairs of bifocal camera sets of focal lengths 7.5 and 24 mm; one looks to the front, the other one to the rear. With 320 by 240 pixels per image this is sufficient for observing road and traffic up to about 100m in front of and behind the vehicle. With its 46 transputers for image processing it has been able in 1994 to recognize road curvature, lane width, number of lanes, type of lane markings, its own position and attitude relative to the lane and to the driveway, and the relative state of up to ten other vehicles including their velocity components, five in each hemisphere. At the final demonstration of the EUREKA-project Prometheus near Paris, VaMP has demonstrated its capabilities of free lane driving and convoy driving at speeds up to 130 km/h in normally dense three-lane traffic [Dickmanns et al., 1994], lane changing for passing and even the decision whether lane changes were safely possible have been done autonomously [Kujawski, 1995]. The human safety pilot just had to check the validity of the decision and to give a go-ahead input.

Figure 8: The autonomous vehicle VaMP of UBM

In the meantime, transputers had been replaced by PowerPCs MPC 601 with an order of magnitude more computing power. A long range trip over about 1600 km to a project meeting in Odense, Denmark in 1995 has been performed in which about 95% of the distance could be traveled fully automatically, in both longitudinal and lateral degrees of freedom. Maximum speed on a free stretch in the northern German plain was 180 km/h.

Since only black-and-white video signals have been evaluated with edge feature extraction algorithms, construction sites with yellow markings on top of the white ones could not be handled; also, passing vehicles cutting into the own lane very near by posed problems because they could not be picked up early enough due to lack of simultaneous field of view, and because *monocular range estimation* took too long to converge to a stable interpretation. For these reasons, the system is now being improved with a wide field of view from two divergently oriented wide angle cameras with a central region of overlap for stereo interpretation; additionally, a high resolution (3-chip) color camera also covers the central part of the stereo field-of-view. This allows for trinocular stereo and area-based object recognition.

Dual-PentiumPro processors now provide the processing power for tens of thousands of mask evaluations with CRONOS per video cycle and processor.

VaMoRs, the 5-ton van in operation since 1985 which has demonstrated quite a few 'firsts' in autonomous road driving, has seen the sequence of microprocessors from Intel 8086, 80x86, via transputers and PowerPCs back to general purpose Intel Pentium and PentiumPro. In addition to early high-speed driving on freeways [Dickmanns and Zapp, 1987] it has demonstrated its capability of driving on state and on minor unsealed roads at speeds up to 50 km/h (1992); it is able to recognize hilly terrain and to estimate vertical road curvature in addition to the horizontal one [Dickmanns and Mysliwetz, 1992].

Recognizing cross-roads of unknown width and angular orientation has been demonstrated as well as turning off onto these roads, even with tight curves requiring an initial maneuver to the opposite direction of the curve [Muller, 1996; Dickmanns and Muller, 1995]. These capabilities will also be considerably improved by the new camera arrangement with a wide simultaneous field of view and area based color image processing.

Performing entire missions based on digital maps has been started [Hock, 1994] and is alleviated now by a GPS-receiver in combination with inertial state estimation recently introduced [Muller, 1996; Werner, 1997]. The vehicles VaMoRs and VaMP together have accumulated a record of about 10 000 km in fully autonomous driving on many types of roadways.

8.2 Air vehicles

After the feasibility of autonomous control in all six degrees of freedom by dynamic machine vision had been demonstrated for the case of straight-in, unperturbed landing approaches in hardware-in-the-loop simulations [Eberl, 1987), a second effort including inertial sensing and both wind and gust disturbances led to first flight tests in 1991 [Schell, 1992]. Because of the safety regulations, the autonomous vision system was not allowed to control the aircraft, a twin turbo-prop of about 5-ton weight, near the ground; the human pilot did the flying but the vision system determined all 12 state components relative to the runway for distances below 900m from runway threshold.

The next step was to introduce bifocal vision with a mild and a stronger tele lens in connection with the new transputer system in the early 90ies; 1993, in another set of flight experiments with the same aircraft of the University of Brunswick it was proved that visual range could be doubled, essentially, but more computing power would be needed for robust tracking and initialization. The PowerPC satisfied these requirements; it is now possible to detect large obstacles on the runway sufficiently early for safe reactions [Furst et al., 1997].

The most demanding guidance and control task performed up to now in hardware-in-the-loop real-time simulations is helicopter flight near the ground including landmark navigation. The capability of performing a small scale mission starting at one end of the airport of Brunswick, flying along a selected sequence of waypoints on the airport and in the vicinity (road forks), returning to the airport from the other side and slowing down for landing at a helicopter 'H' at the other end has been demonstrated [Werner et al., 1996; Werner, 1997] (see figure 9).

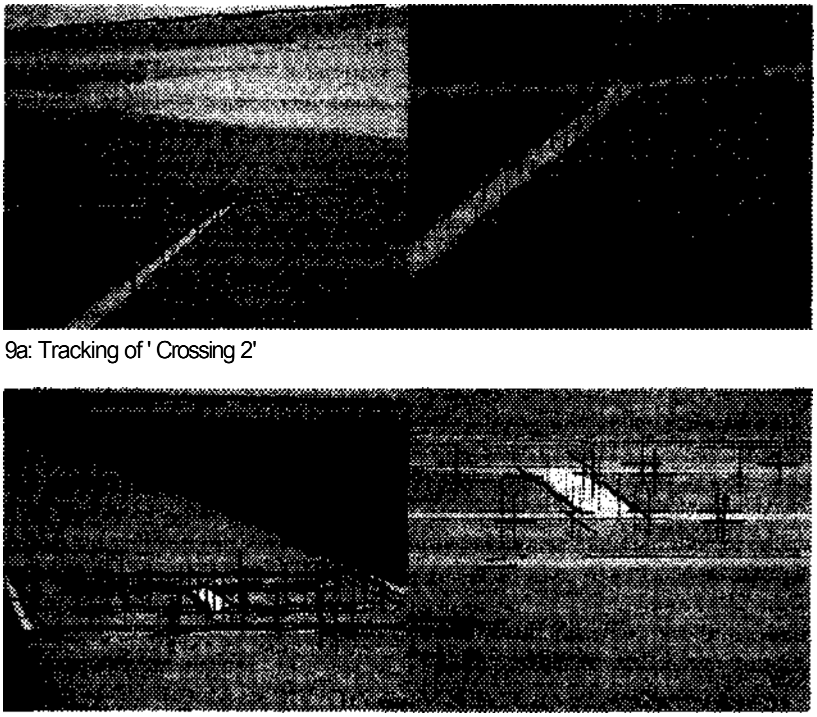
In connection with this demanding task, a complete software package has been developed containing separate inertial and visual state estimation components, integration of GPS signals and data fusion in the context of mission performance. In addition, provisions have been made to integrate coarse-scale image data from a synthetic aperture imaging radar system under development elsewhere. The combined use of all-weather radar images and high-resolution optical or infrared images is considered an optimal solution for future helicopter guidance systems. The capability of interpreting these data streams by an intelligent on-board computer system will unload the pilot from a very difficult task in a situation where he is stressed to the limit already.

9 Technical Beings'?

There is an ongoing discussion as to what technical beings may be like and what the best architecture for realizing these agents might be [Brooks and Flynn, 1989; Steels, 1993]; subsumption architeaure and neural nets have been proposed as roads leading to these type of creatures.

Looking at the results achieved with the 4-D approach to dynamic vision, it does not seem unreasonable to expect that quite a few problems to be encountered in complex scenarios with the other approaches may be avoided taking this route which builds upon long term results in the natural sciences and engineering.

It has the advantage of having a clear notion of space and time, of objects, subjects and processes, and of the spatio-temporal representational structure necessary to handle multiple independent objects and subjects with own intentions, goals and control capabilities.



9a: Tracking of ' Crossing 2'

9b: Tracking of taxiways, frame and Heli H during final approach

Figure 9: Landmark navigation for helicopters has been demonstrated in hardware-in-the-loop, real-time simulations for a small mission near the airport of Brunswick

In [D.Dickmanns, 1997] a corresponding representational framework has been given which allows to handle even complex systems with minimal additional effort on the methodical side; knowledge specific to the task domain may be entered through corresponding data structures. Computing power available in the near future will be sufficient to solve rather complex real-time, real-world problems. A corresponding architecture for road vehicle guidance is discussed in [Maurer and Dickmanns, 1997].

As compared to the other approaches pursued, the pledge is to take advantage of the state of the art in engineering and simulation technology; introducing goal functions for these autonomous systems and providing them with background knowledge of how to achieve these goals, or how to learn to achieve them will be essential. The argument sometimes heard that these systems will be 'closed' as opposed to neural-net-based ones is not intelligible from this point of view.

10 Conclusions

The 4-D approach to dynamic machine vision developed along the lines layed out by cybernetics and conventional engineering long time ago does seem to satisfy all the expectations it shares with 'Artificial Intelligence'- and 'Neural Net'-approaches. Complex perception and control processes like ground vehicle guidance under diverse conditions and in rather complex scenes have been demonstrated as well as maneuver- and mission-control in full six degrees of freedom. The representational tools of computer graphics and -simulation have been complemented for dealing with the inverse problem of computer vision.

Computing power is arriving now for handling realword problems in real-time. Lack of robustness encountered up to now due to black-and-white as-well-as edge-based image understanding can now be complemented by area-based representations including color and texture, both very demanding with respect to processing power.

Taking advantage of well suited methods in competing approaches and combining the best of every field in a unified overall approach will be the most promising way to go. The good old stuff should not be discarded too early.

Literature

[Behringer, 1996] R. Behringer: Visuelle Erkennung und Interpretation des Fahrspurverlaufes durch Rechnersehen fur ein autonomes StraBenfahrzeug. PhD thesis, UniBwM, LRT, 1996.

[Brooks and Ftynn 1989] R.A. Brooks and A.M. Flynn: Robot beings. *IEEE/RSJ International Workshop on Intelligent Robots and Systems,* Tsukuba, Japan, Sept. 1989, pp 2-10.

PeCarlo and Metaxas, 1996] D. DeCarlo and D. Metaxas: The Integration of Optical Flow and Deformable Models with Applications to Human Face Shape and Motion Estimation. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition,* San Francisco, CA, June 1996, pp 231-238.

[D.Dickmanns, 1997] Dirk Dickmanns: Rahmensystem fur visuelle Wahrnehmung veriinderlicher Szenen durch Computer. PhD thesis, UniBwM, INF, 1997.

[Dickmanns, 1987] ED. Dickmanns: 4-D-Dynamic Scene Analysis with Integral Spatio-Temporal Models. *4th Int. Symposium on Robotics Research,* Santa Cruz, 1987.

[Dickmanns and Zapp, 1987] E.D. Dickmanns and A. Zapp: Autonomous High Speed Road Vehicle Guidance by Computer Vision. *10th IFAC World Congress Munich,* Preprint Vol. 4, 1987, pp 232-237.

[Dickmanns and Graefe 1988] E.D. Dickmanns, V. Graefe: a) Dynamic monocular machine vision. *Machine Vision and Applications,* Springer International, Vol. 1, 1988, pp 223-240. b) Applications of dynamic monocular machine vision, (ibid), 1988, pp 241-261.

[Dickmanns, 1992] E.D. Dickmanns: Machine Perception Exploiting High-Level Spatio-Temporal Models. *AGARD Lecture Series 185* 'Machine Perception', Hampton, VA, Munich, Madrid, Sept./Oct. 1992.

[Dickmanns and Mysliwetz, 1992] E.D. Dickmanns and B. Mysliwetz: Recursive 3-D Road and Relative Ego-State Recognition. *IEEE-Transactions PAMI,* Vol. 14, No. 2, Special Issue on 'Interpretation of 3-D Scenes', Feb 1992, pp 199-213

[Dickmanns et al., 1994] E.D. Dickmanns, R. Behringer, D. Dickmanns, T. Hildebrandt, M. Maurer, F. Thomanek, J. Schiehlen: The Seeing Passenger Car 'VaMoRs-P'. In *Intelligent Vehicles Symposium '94,* Paris, Oct. 1994, pp 68-73.

[Dickmanns, 1995a] E.D. Dickmanns: Performance Improvements for Autonomous Road Vehicles. *Int. Conference on Intelligent Autonomous Systems (IAS-4),* Karlsruhe, 1995.

[Dickmanns 1995b] E.D. Dickmanns: Road vehicle eyes for high precision navigation. In Linkwitz et al. (eds): *High Precision Navigation.* Diimmler Verlag, Bonn, 1995, pp. 329-336.

[Dickmanns and Muller, 1995] E.D. Dickmanns and N. Muller: Scene Recognition and Navigation Capabilities for Lane Changes and Turns in Vision-Based Vehicle Guidance. *Control Engineering Practice,* 2nd IFAC Conf. on Intelligent Autonomous Vehicles-95, Helsinki 1995.

[Dickmanns, et al., 1997] E.D. Dickmanns, S. Furst, A. Schubert, D. Dickmanns: Intelligently controlled feature extraction in dynamic scenes. Technical report UniBwM/LRT/WE-13/FB/97-l, 1997.

[Eberl, 1987] G. Eberl: Automatischer Landeanflug durch Rechnersehen. PhD thesis, UniBwM, LRT, 1987.

[Fritz, 1996] H. Fritz: Model-Based Neural Distance Control for Autonomous Road Vehicles. *Proc. Intelligent Vehicles '96 Symposium,* Tokyo, 1996, pp 29-34.

[Ftirst et al., 1997] S. Ftirst, S. Werner, D. Dickmanns, and E.D. Dickmanns: Landmark navigation and autonomous landing approach with obstacle detection for aircraft. *AeroSense '97, Conference 3088,* Orlando FL, April 20-25, 1997.

[Hock, 1994] C. Hock: Wissensbasierte Fahrzeugfuhrung mit Landmarken fur autonome Roboter. PhD thesis, UniBwM, LRT, 1994.

[Kailath, 1980] T. Kailath: Linear Systems. Prentice-Hall, Inc., Englewood Cliffs, N.J., 1980.

[Klass, 1985] P.J. Klass: DARPA Envisions New Generation of Machine Intelligence. *Aviation Week & Space Technology,* April 1985, pp 47-54.

[Kujawski, 1995] C. Kujawski: Deciding the behaviour of an autonomous mobile road vehicle. *2nd IFAC Conference on Intelligent Autonomous Vehicles,* Helsinki, June 1995.

[Luenberger, 1964] D.G. Luenberger: Observing the state of a linear system. *IEEE Trans on Mil Electronics* 8, 1964, pp 290-293.

[Maurer and Dickmanns, 1997] M. Maurer and E.D. Dickmanns: An advanced control architecture for autonomous vehicles. *AeroSense '97, Conference 3087,* Orlando FL, April 20-25, 1997.

[Mecklenburg et al., 1992] K. Mecklenburg, T. Hrycej, U. Franke and H. Fritz: Neural Control of Autonomous Vehicles. *Proc. IEEE Vehicular Technology Conference '92,* Denver, 1992.

[Meissner, 1982] H.G. Meissner: Steuerung dynamischer Systeme aufgrund bildhafter Informationen. PhD thesis, UniBwM, LRT, 1982.

[Meissner and Dickmanns, 1983] H.G. Meissner and E.D. Dickmanns: Control of an Unstable Plant by Computer Vision. In T.S. Huang (ed): Image Sequence Processing and Dynamic Scene Analysis. Springer-Verlag, Berlin, 1983, pp 532-548.

[Muller, 1996] Muller N.: Autonomes Manovrieren und Navigieren mit einem sehenden StraBenfahrzeug. PhD thesis, UniBwM, LRT, 1996.

[Pomerleau, 1989] D.A. Pomerleau: ALVINN: An Autonomous Land Vehicle in Neural Network. In D.S. Touretzky (ed.) Advances in Neural Information Processing Systems 1. Morgan Kaufmann, 1989.

[Pomerleau 1992] D.A. Pomerleau: Neural Network Perception for Mobile Robot Guidance. PhD thesis, CMU [CMU-CS-92-115], Febr 1992.

[Schmid, 1994] M. Schmid: 3-D-Erkennung von Fahrzeugen in Echtzeit aus monokularen Bildfolgen. PhD thesis, UniBwM, LRT, 1994.

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

[Schell, 1992] F.R. Schell: Bordautonomer automatischer Landeanflug aufgrund bildhafter und inertialer MeBdatenauswertung. PhD thesis, UniBwM, LRT, 1992.

[Schiehlen, 1995] J. Schiehlen: Kameraplattformen fur aktiv sehende Fahrzeuge. PhD thesis, UniBwM, LRT, 1995.

[Steels, 1993] L. Steels: The Biology and Technology of Intelligent Autonomous Agents. NATO-Advanced Study Institute, Ivano, Italy, March 1-12, 1993.

[Thomanek, 1996] F. Thomanek F.: Visuelle Erkennung und Zustandsschatzung von mehreren StraBenfahrzeugen zur autonomen Fahrzeugfuhrung. PhD thesis, UniBwM, LRT, 1996.

[Tsugawa et al, 1979] S. Tsugawa, T. Yatabe, T. Hirose, S. Matsumoto: An Automobile with Artificial Intelligence. *Proceedings 6th IJCAI,* Tokyo, 1979, pp 893-895.

[Ulmer, 1994] B. Ulmer: VITA II - Active collision avoidance in real traffic. *In Intelligent Vehicles Symposium '94,* Paris, Oct. 1994.

[Werner et al., 1996] S. Werner, S. Ftirst, D. Dickmanns, and E.D. Dickmanns: A vision-based multi-sensor machine perception system for autonomous aircraft landing approach. *Enhanced and Synthetic Vision, AeroSense '96,* Orlando, FL, April 1996.

[Werner, 1997] S. Werner: Maschinelle Wahrnehmung fur den bordautonomen automatischen Hubschauberflug. PhD thesis, UniBwM, LRT, 1997.

[Wuensche, 1983] H.-J. Wuensche: Verbesserte Regelung eines dynamischen Systems durch Auswertung redundanter Sichtinformation unter Berucksichtigung der Einflusse verschiedener Zustandsschatzer und Abtastzeiten. Report HSBw/LRT/WE 13a/IB/83-2, 1983.

[Wuensche, 1986] H.-J. Wuensche: Detection and Control of Mobile Robot Motion by Real-Time Computer Vision. In N. Marquino (ed): Advances in Intelligent Robotics Systems. *Proceedings of the SPIE,* Vol. 727, 1986, pp 100-109.