CCG-Lehrgang SE 3.02 September 1991

Echtzeit-Bildfolgenverarbeitung zur räumlichen Erfassung von Bewegungen und zur Steuerung von Fahrzeugen (Land-, Luft- und Raumfahrzeuge)

Dynamic vision for locomotion control - An evolutionary path to intelligence

Ernst D. Dickmanns*

Abstract

A paradigm for dynamic machine vision is presented which takes time and 3D space in an integrated manner as the underlying framework for internal representation of the visually observed outside world. This world is considered to consist of material and mental processes evolving over time. The concept of state and control variables developed in the natural sciences and engineering over the last three centuries is exploited to find a new, more natural access to dynamic real-time vision and intelligence. A. Schopenhauer's conjecture of 'The world as evolving process and internal representation' (1819) is combined with modern recursive estimation techniques [Kalman 1960] and some ingredients from geometry and AI in order to arrive at a very efficient scheme for autonomous robotic agents dealing with evolving processes in the real world in real time. Application to autonomous mobile robots is discussed.

* Prof. Dr.-Ing., Universität der Bundeswehr München, Germany

Content

Introduction

Computer vision has evolved from digital image processing over the last three decades. Therefore, it is usually embedded in a quasistatic framework of snapshot interpretation. On the contrary, biological vision systems seem to have developed for motion detection and control in an ever changing physical environment. Are the best suited methods for both tasks the same or are there fundamental differences?

In the Artificial Intelligence (AI) community the vision problem has initially been tackled as a quasistatic problem. Much effort has been devoted to the inversion of the perspective mapping process taking several (consecutive) frames into account; for a survey see [Nagel 83]. This does not take advantage of the temporal continuity conditions in the physical world to which all material processes are usually subjected.

In physics, especially in mechanics, powerful methods have been developed over the last three centuries in order to describe the observed behavior of material processes. In engineering, over the last three decades these methods have been supplemented by features well adapted for recursive digital data processing. Recursive in this context means that least squares data interpretation is achieved step by step as new data arrive. The discipline of systems dynamics evolved out of these activities encompassing aspects of several fields: from sensor technology, signal processing, control theory and design, actuator technology through dynamic behavior of systems.

In this article, the systems dynamics approach is applied to the field of visual dynamic scene understanding, motion control and intelligence. Off the beaten track of main stream research into computer vision, this approach has been developed over the last decade. Combining well proven engineering methods with knowledge from geometry (perspective mapping) and some new aspects of AI, a surprisingly powerful and efficient scheme for the general task of dynamic machine vision using distributed processing resulted. The basic connecting link is a very old idea which the German philosopher Arthur Schopenhauer conjectured more than 170 years ago ['Die Welt als Wille und Vorstellung', 1819, freely translated: The world as evolving process and internal representation].

Building on I. Kant's basic result from two centuries ago, which also formed the foundation for Schopenhauer's conjecture, namely that space and time are not attributes of objects but are carried into the world through our perception and analysis system, it was decided to represent space and time directly in the interpretation scheme. In addition, the constraint was deliberately imposed on the approach that it should work in real time, i.e. that the computational progress over time is directly linked to the progress of the physical process observed and controlled, and not limited by the present state of computer hardware performance. Of course, this confined the problems to be treated considerably in the early 80-ies. It had the members of the team look at problems in a different way, however, and both image processing and scene interpretation algorithms developed differently as compared to the results of other groups who worked under the paradigm that the increasing processing power of future miroprocessor generations will solve all the performance problems with respect to real time.

After a decade of steadily increasing complexity of the problems solved and with experience in five different problem areas, it seems timely to present the approach and the basic ideas behind it in a comprehensive way; the seven dissertations in which most of the material has been originally published are in German language and, therefore, not readily accessible to the general public. The survey article [Dickmanns and Graefe 88] triggered much interest which was one of the driving factors for writing this document.

The present article is intended as a general introduction to the '4D approach' for all those interested in machine vision applications in real world dynamical scenes. Emphasis is put on exploiting knowledge about the physical world and temporal processes; image sequences are nothing but discrete and systematically impoverished intermediate carriers of information about the spatio-temporal world. It is the main goal of the article to shift the paradigm for dynamic machine vision from more academic computer science to practical applications in physics and engineering and to the corresponding methods. Practitioners should find it particular attractive to experience the direct connections from this modern, very promising field of development to well proven methods in conventional applied sciences.

Resorting to these tools, hopefully, will not have AI-researchers turn away immediately. It is the blend of methods which will lead to efficient machine intelligence systems.

Lessons learned from the natural sciences, mathematics and engineering

The intention of this approach is not primarily to generate some artificial counterpart of what is called intelligence, but to enable machines with complex sensory systems and the capability of self-controlled locomotion to get around in the real world in a meaningful way; by doing this, some kind of intelligence will emerge more as a side effect in a natural way.

In physics and the engineering sciences mankind has learned over the last centuries how to analyse and represent natural and artificial objects and processes in the environment efficiently. The condensed results of this longterm endeavor of interest to the field of dynamic vision are reviewed briefly in the following sections.

Three-dimensional (3D) space and time

Early geometricians, already millennia ago, discovered that the space we happen to live in can be exhaustively analysed using three independent coordinates. After the more modern French scientist Descartes the orthogonal ('Cartesian') coordinate systems in wide use today are named.

The relationship between space and time has been more obscure for a long time. It was Newton who in the 17-th century invented the differential calculus and applied it to motion analysis. This step in the natural sciences together with the introduction of the inverse square field of gravity brought about a revolution in motion understanding. After this step the geometrically known orbits of planets (Kepler's ellipses) could be linked to a few dynamical motion parameters. The time derivative of the moment of momentum (the second time derivative of position variables in cases of constant mass) was postulated to be proportional to forces, which in a gravity field were in turn linked to position.

The general description of this famous motion law, which despite modern theory of relativity is well justified in conventional mechanics still today, may be written in vector notation as $(^{o} = d(\,)/dt)$

$$\underline{x}^{o} = \underline{f}(\underline{x}, \underline{u}, \underline{p}, t) \ , \tag{1}$$

where $\underline{x}$ is the state vector with n components, $\underline{u}$ the control vector of dimension r to be freely selected at each point in time, and $\underline{p}$ the parameter vector of dimension q characterizing the special problem. In each degree of freedom, since acceleration as the second time derivative is proportional to forces or moments, two state components (position and velocity) have to be taken into account. Therefore a particle moving freely in 3D space has to be described by 12 state variables, 6 for translation and 6 for rotation, 3 each for position and velocity. For motion in a plane, 6 state variables are sufficient.

It is the integral relationship from acceleration to velocity and from velocity to position which constitutes essential (implicit) knowledge about the temporal behavior of massive objects in the real world. We humans do not have to learn this knowledge consciuosly, since it is absorbed subconsciously during the first years of our lives while we learn to crawl and walk and to react to other moving objects or subjects properly. Some individuals develop a special skill in this respect; they are good sportsmen even though they may not be able to explicitly formulate how they behave. A wealth of knowledge about the real world is acquired and coded in our neural nets this way even though it is not yet known how.

3D shape and perspective mapping

A similar situation may prevail with respect to our 3D shape understanding through vision. Geometric mapping has been applied for many millennia in all cultures around the globe. Sensible theories about the vision process are less than one millenium old; a nice survey on early vision theories is given in [Lindberg 76]. The difficult problem in vision is that even though the input into data processing is a 2D matrix (spherically arranged in the eye or planar in a camera) the conscious interpretation should be spatial according to the relative physical positions of objects in the real world. For one single photographic snapshot this problem cannot be solved; much effort in computer vision has been devoted to the problem of how many different images are sufficient for uniquely reconstructing the spatial scene.

The laws of perspective projection, according to which each visible particle emanates or reflects straight line light rays from its spatial position to the receiver, is considered to be a sufficiently good model, discarding all side effects of real lenses and mapping devices.

The shape of real bodies has to be inferred from intensity distributions over its visible surfaces and their behavior over time during relative motion. Oftentimes, physical edges and region boundaries on the surface lead to intensity edges in the image plane which, when observed under steadily changing aspect conditions, may allow the proper spatial interpretation (shape from X).

For the representation of 3D shapes the engineering sciences have perfected a 2D representation scheme showing parallel projection views from three (or all six) mutually orthogonal directions. If the object has a plane of symmetry, two (four) of these viewing directions should preferably lie within this plane. One or two reference axes are usually chosen in such a way that the object is oriented in a functionally proper way under normal Earth gravity conditions (e.g. a car with all four wheels touching the ground plane). Nonunique interpretation possibilities (e.g. in concavities) may be disambiguated by special 2D cuts through these regions. A skilled and trained person can imagine the proper perspective view of this object from any aspect condition. For practical purposes, only approximately correct 3D views (to within a few percent accuracy) are often sufficient for object recognition; this can be achieved using relatively simple heuristics for fast and efficient computation of the perspective image given the 2D normal views. 2D shapes with smoothly curved contours and corners can be efficiently represented in a translation, rotation- and scale- invariant form by Normalized Curvature Functions (NCF) [Dickmanns 85] which in turn are easily measurable by tangency operations in the image plane.

Dynamical models of physical processes

The term 'dynamical model' in mechanics, systems dynamics and control theory means a generic differential equation description (like in eq. (1)) for some motion process. We confine the discussion here to motion of massive bodies, be it rigid or elastic. In the case of rigid bodies, classical mechanics has shown that the overall motion can be decoupled into translation of the center of gravity (cg) and rotation around the cg. In the case of elastic bodies some deformation may be superimposed which in the case of free motion usually is an oscillation around a reference shape.

For massive rigid bodies, the forces and moments acting on a specific body are usually very limited in magnitude leading to a characteristic motion behavior over time like a ball flying through the air in the gravity field; gravity and its secondary effects like friction in sliding or rolling motion as well as fluid dynamic drag predominate many motion processes in the real world. Once these basic influences are properly understood (internally represented by a model), a prediction of physical motion in 3D space becomes easy. Combining this with the perspective mapping knowledge of the previous section allows to predict motion appearing in the image plane. Note that for the motion in the image plane no similarly simple direct models can be given due to the nonlinear perspective mapping involved.

The use of dynamical models enforces the internal representation to be in space and time simultaneously (4D). Since the image sequence is discretized over time (50 or 60 Hz corresponding to a video cycle time T' of 20 or 16 2/3 ms), this basic cycle time T' or an integer multiple T thereof is used to transform the differential equation (1) into a difference equation leading to a state transition matrix **A** and a control input matrix **B**

$$\underline{x}\,[(k+1)T] = \mathbf{A}\,(\underline{x}, \underline{p}, kT) \cdot \underline{x}\,(kT) + \mathbf{B}\,(\underline{x}, \underline{p}, \underline{u}, kT) \cdot \underline{u}\,(kT)\;, \tag{2}$$

which yield a very compact knowledge representation for the temporal evolution of physical processes in the real world. Note that in the second additive term on the right hand side the effect of control action is contained; this makes this type of representation especially attractive since it allows to include the intelligent motion control part into the prediction scheme. For more long term prediction, probably for investigating the effect of some future control time history of the own vehicle (maybe even several alternatives thereof) this eq. has to be evaluated as many times as requested into the future, thereby allowing a simple means for temporal reasoning. Entire action sequences may be investigated (simulated) this way before decision taking.

State and control variables, process parameters

In an efficient description of real world processes there are three types of variables involved:

1. Those which can be changed at any time at will: e.g. steering wheel turn rate of a car, voltage applied to an electromotor, force applied to an aircraft control stick, throttle position of an engine. These variables are called **control variables $\underline{\mathbf{u}}(\mathbf{t})$**. Note that this definition is somewhat arbitrary: If the force applied to an aircraft control stick is such that the desired control stick position is reached before the aircraft starts moving in its eigenmodes, the control stick position could have been chosen as the control variable (as has been done with the engine throttle). The essential point is that the control motion has to have a dynamic behavior at least one order of magnitude faster than the controlled process.
2. Those variables which can not be changed directly but which only evolve over time: these are the socalled **state variables $\underline{\mathbf{x}}(\mathbf{t})$**. Their evolution over time is as characteristic for an object in the temporal domain as shape is in the spatial domain. Exploiting this knowledge about moving objects in addition to shape constancy results in much more efficient recognition and tracking schemes for moving objects. Note that the spatial velocity components of objects are state variables in this sense; again, this is a strong argument for favoring an internal representation in 3D space and time via dynamical models.
3. Variables which are fixed over periods of time and which may be selected at some discrete point in time, including the system design phase: socalled **system parameters $\underline{\mathbf{p}}$**. Typical examples are shift gear position in a car, landing flap position in an aircraft, switch positions etc. and the constants in the system matrices A and B. This set of system variables can be considered constant over time for short

term motion behavior even though there may occur a slow change due to wear and tear or environmental effects like temperature or humidity.

Knowledge about a dynamical system is firstly coded in the set of parameters $\underline{p}$ and the structure of the matrices A and B as well as their numerical entries. Equally important in the temporal domain is, however secondly, knowledge of how the system is going to behave with respect to its state variables in response to some control input over time. Especially, the question of how a desired set of state components can be achieved efficiently by appropriate control input time histories is practically relevant; the entire field of 'optimal control theory and application' is devoted to this problem. Mathematicians have developed the calculus of variation for this purpose [Euler 1744] and the 'Maximum principle' [Pontryagin et al. 62], which especially in aerospace engineering but also in many other fields has important and widespread applications since the time that digital computers allow to solve the corresponding difficult numerical problems [Bryson, Ho 75].

To intelligent agents the control variables are of special importance since they constitute the only means through which any influence can be exerted on an evolving process in the real world. Discretely selectable parameters like a switch or flap position may be viewed as 'control parameters' and handled correspondingly. Controls in this sense are the extremely important parts of a system where 'a free will' working on information collected by sensors can exert an influence on the process under control. The provocative term 'free will' will be discussed later.

Feedforward and feedback control loops (cybernetics)

When an experienced person drives a car and wants to switch lane on a highway she or he implements an approximately sinusoidal steering wheel maneuver over time without thinking about it. The amplitude and the time rate are adjusted in such a way that the car finishes this maneuver approximately in the center of the new lane. This can be done in one smooth overall maneuver. A beginner, on the contrary, since unfamiliar with the behavior of the car, will tend to use small incremental control inputs and observe the reaction of the car which in turn will lead him to select the next control input step until the car will finally also end up in the new lane, however, much later and without a smooth control time history. The experienced person since knowing the temporal response of the car to a 'feedforward control' time history made use of this knowledge leading to better performance; the beginner observing the actual discrepancy between desired and actual state used the difference in some way to feed the control input according to some rule (e.g. a constant factor times the negative difference).

By applying a 'feedback control law' the behavior over time of the controlled vehicle is fixed, but modified relative to the 'open loop'-behavior without any control input. The actuator need not be a person but may be some suitable technical subsystem like an electro-motor or an hydraulic actuator leading to an automated system.

Control engineering and mathematics have developed theoretical and numerical methods which allow designing closed-loop systems with complex eigenbehavior. Literature abounds in this field; just one among many others is [Kailath 80].

Dynamic sytems design

With the powerful digital microprocessors available today, combinations of event-triggered feedforward control and robust feedback control laws for different subtasks allow the development of very flexible and high performance automatic systems.

Even though the theories developed are mostly based on the assumption of a linear system description, a very large percentage of the generally nonlinear 'plants' (the technical systems to which automation is applied) can be handled this way since linearisations around the actual reference point usually are sufficiently good approximations to the system, especially since feedback controllers keep the system actively in this domain by their functioning. By adding a systems identification component, the temporal change of system parameters can be detected and the control scheme may be adjusted accordingly without human intervention.

Modern trends go towards coupling automatic control systems with expert systems in order to improve flexibility and robustness of the overall system under a wide variaty of operating conditions. The system discussed in the sequel for real time machine vision may be subsumed under this category.

Kalman's recursive state estimation technique

For interpreting measurements, modern control systems theory has deviced an elegant scheme, how optimal estimates of the actual state of internally represented objects from the real outside world may be arrived at in an efficient way exploiting dynamical models about spatio-temporal relationships of the processes involved. It allows recovering the full state vector even in cases where only partial measurements of some output variables can be taken. These output variables have to be linked to the state variables by some smooth functional relationship. This scheme is extremely well suited to vision processes where the depth component is systematically lost during imaging and where partial occlusions are more the rule than an exception.

Measurements usually are noise corrupted. Therefore, good state estimation can only be achieved when processing many more data than are minimally required. A brief sketch of the historical development of this technique is given in the following subsections.

Gauss's model based least squares scheme for measurement interpretation: When the structure of the motion trajectory is known in advance like for ellipses in planetary motion around the central star, this knowledge can be used efficiently in order to smooth noisy measurement data. The mathematician K.F.Gauss has introduced the technique of fitting curves of known structure to noisy data by minimizing the sum of the squares of the residues. This has lead to much improved accuracies in orbit determination and general curve fitting.

Note, that this improvement is achieved by using solution curves of motion processes, and that a set of measurement data has to be batch-processed at a time.

From generic solution curves to differential equation models: If the goal is to have good actual motion state estimates while motion is in progress one would like to have a scheme which gives an incremental update at each point in time when new data become available. If the process observed can be influenced by control input, no a priori structure for the solution curve can be given. In these cases, instead of exploiting solution curves the underlying generic differential equations are more appropriate. For the linear case with known noise statistics [Kalman 1960] has given a recursive least squares scheme which allows optimal state estimation from a reduced set of output measurements. Space does not allow to go into details here; the interested reader is referred to [Maybeck 79]. The known system structure of eq. (2) allows to recover state components which are not directly measured by substituting structural knowledge for missing measurements, observability given. The error covariance matrix plays an important role in this process and may be exploited for the removal of outliers, thereby stabilizing the interpretation process.

The big advantage of this recursive state estimation scheme is that always only the last measurements are used for updating the best estimates without the need for storing previous data, which is especially rewarding in image sequence processing where each image comprises enormous amounts of data (10^5 to 10^6 Bytes). The result of all previous data is the present best estimate for the state vector of objects and the covariance matrix corresponding to a storage requirement in the order of magnitude 10^2 per object tracked.

Extended and sequential (numerically favorable) recursion schemes: In the case of nonlinear components in the system description, the socalled extended Kalman filter has been developed based on linearisations around the actual reference point.

In order to keep the covariance matrix symmetric, the upper triangle factorization UDU^T has been introduced [Bierman 75; Maybeck 79]. It is numerically more efficient and stable and is being widely used.

If the state update is computed every time one single measurement component is acquired, the use of two-dimensional arrays in the program may be reduced, leading to faster execution. In addition, this scheme allows an easy adjustment for image sequence processing in the case where - due to occlusion or some other cause - the number of measurement components varies from frame to frame. In our software, this feature has been adopted as a general standard [Wuensche 88, Christians 89, Mysliwetz 90].

Real-time vision, in our approach, is considered to be a measurement process with remote access to the systematically transformed object state (by perspective projection); identification of the object has to be achieved simultaneously with the determination of the motion state.

For image sequence processing, the recursive estimation scheme had to be further extended for the nonlinear perspective mapping of point and line features. In addition, the relationship between the dynamical model for cg-motion and the position and orientation of features on the surface of the body had to be incorporated. The resulting overall scheme will be described next.

Stimuli from philosophical thoughts

Humans with their capability of locomotion and complex information processing may be considered as very complex dynamical systems with a mental component by far not yet understood. Philosophers for millennia have tried to understand human performance in different fields. The natural sciences joined in this endeavor since more than three centuries in a more systematic fashion, but still one is way from having satisfactory answers, though considerable progress has been made recently with the help of information processing technology.

On the basis of Newton's laws of motion and the new understanding of time, Kant in the 18-th century clarified the situation in philosophy by his main works 'Critiques .....' [Kant 1780-ies] to a considerable extent. He separated space and time from attributes of objects granting the former ones a special basic quality. He also introduced a clear distinction between a material object (the 'thing by itself' = "das Ding an sich" (in German)) and a human's notion about this object. The succeeding 'Idealist'-philosophers at the turn from the 18-th to the 19-th century may have turned world interpretation 'upside-down' by giving ideas priority over matter and over the outside world; at least, this was Schopenhauer's impression. In an attempt to put the world from this position 'back onto the feet again', he speculated about the interdependence between the material processes in the world and mind. The basic idea behind the second part of his book title 'The world as will and **internal representation**' [Schopenhauer 1819] may be considered to be a major breakthrough in concepts about cognition.

This basic idea has been adopted as the focal point in our approach to machine vision irrespective of all previous philosophical and psychological controversy. It is not intended to get involved into this discussion as far as humans are concerned; however, this idea has been - probably for the first time - put to work in the context of cognitive machines.

Let us assume there is a material world to which an autonomous agent, say based on a conventional wheeled road vehicle, itself being part of this world, has limited access to (with regard to physical state measurements). This may be achieved through a multi-sensor system encompassing properly calibrated odo- and velocimeters, sensors for control inputs, inertial sensors for translation (accelerometers) and rotation (angular rate and position sensors), a microphone for audio-input and imaging sensors in some spectral bands. All these signals are fed into a computer system with properly suited data processing programs.

The autonomous system is assumed to be endowed with all the relevant knowledge components discussed in the previous section. Provision has been taken that the engine is running, the sensory and motor control systems are operative and that there is enough computing power available for properly processing the sensory data; the computer system has access to the control actuation subsystems (even including voice output, say).

The yet open question is: Is it possible to generate an overall system capable of demonstrating a behavior which is qualitatively similar to that of intelligent humans?

The integrated 4D approach to dynamic vision

The main goal of this approach from its beginning in the early 80-ies has been to take advantage of the full spatio-temporal framework for internal representation and to do as few reasoning as possible in the image plane and in between frames. Instead, temporal continuity in physical space according to some model for the motion of objects is being exploited in conjunction with spatial shape rigidity in this 'analysis-by-synthesis' approach.

Basic scheme

Dynamical models link time to spatial motion, in general. The shape models exhibit the spatial distribution of visual features on the surface which allow objects to be recognized and tracked. 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 successors has been extended to image sequence processing by our group [Kalman 60; Wuensche 88]. There are so many publications on this approach that only a short summary will be given here (see e.g. the survey article [Dickmanns and Graefe 88]).

Figure 1 shows the resulting coarse overall blockdiagram of the vision system based on these principles. To the left, the real world is shown by a block; control inputs to the own 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 ob-

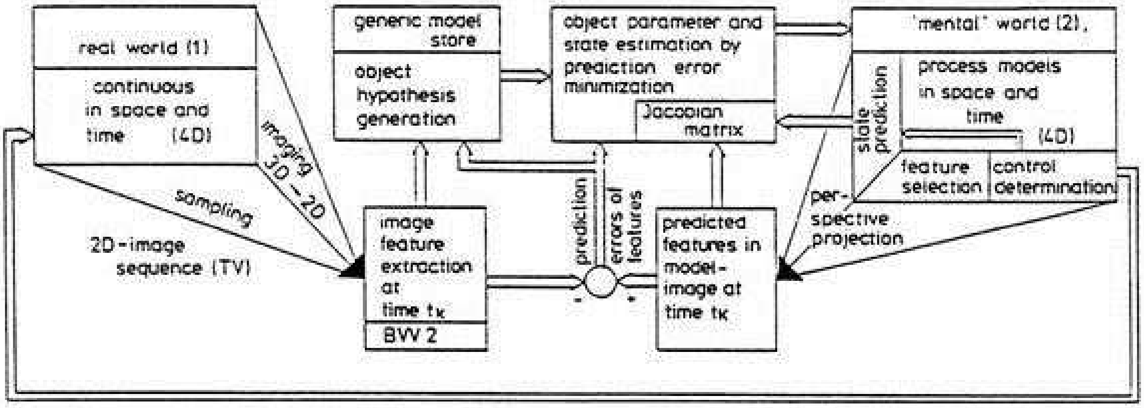

Figure 1: Basic scheme for 4D-image sequence understanding by prediction error minimization

jects and their relative position in the world over time are sensed by CCD-sensor arrays (shown as converging lines to the lower right, symbolizing the 3D to 2D 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 two ways: There is a limited spatial resolution in the image plane and a temporal discretization of 16 2/3 or 20 ms (due to the different video standards), usually including some averaging over time.

Instead of trying to invert this image sequence for 3D-scene understanding, a different approach by analysis through synthesis has been selected, taking advantage of the available recursive estimation scheme after Kalman. From previous human experience, generic models of objects in the 3D-world are known in the interpretation process. This comprises both 3D shape, recognizable by certain feature aggregations given the aspect conditions, and motion behavior over time. In an initialisation phase, starting from a collection of features extracted by low level picture element (pel) processing (BVV_2, lower center left in fig. 1), object hypotheses including the aspect conditions and the motion behavior (transition matrices) in space have to be generated (upper center left in fig.1). They are installed in an internal 'mental' world representation intended to duplicate the outside real world. After the philosopher K.Popper this is sometimes called 'world_2', as opposed to the real 'world_1'.

The initialisation is the most difficult part and has been solved for well defined simple problems only. A more general capability is being developed presently. It consists of both data driven bottom up and model driven top down components cooperating over time as discussed in the next section.

Once an aggregation of objects has been instantiated in the world_2, exploiting the dynamical models for those objects allows the prediction of object states for that point in time when the next measurements are going to be taken. By applying the <u>forward</u> perspective projection to those features which will be well visible, using the same mapping conditions as in the TV-sensor, a model image can be generated which should duplicate the measured image if the situation has been understood properly. The situation is thus 'imagined' (right and lower center right in fig. 1). The big advantage of this approach is that due to the internal 4D-model not only the actual situation at the present time but also the sensitivity matrix of the feature positions and orientations with respect to all state component changes can be determined, the socalled Jacobian matrix (upper block in center right, lower right corner). This need not necessarily be done by analytical means but may be achieved with little programming effort by numerical differentiation exploiting the mapping subroutines already implemented for the nominal case.

This rich information is used for bypassing the perspective inversion via recursive least squares filtering through feedback of the prediction errors of the features. Unfortunately, space does not allow to go into more details here (see [Dickmanns and Graefe 88]).

This approach has several very important practical advantages:
- no previous images need be stored and retrieved for computing optical flow or velocity components in the image plane as an intermediate step in the interpretation process,
- the transition from signals (pel data in the image) to symbols (spatio-temporal motion state of objects) is done in a very direct way, well based on higher level knowledge, the 4D world model integrating spatial and temporal aspects;
- intelligent nonuniform image analysis becomes possible, allowing to concentrate limited computing resources to areas of interest known to carry meaningful information;
- the position and orientation of well visible features can be predicted and the feature extraction algorithms can be provided with information for more efficiently finding the desired ones; outliers can easily be removed thereby stabilising the interpretation process.
- viewing direction control can be done directly in an object-oriented manner.

Processing a variable number of features measured from frame to frame is alleviated by using the sequential filtering version. For improving numerical performance, the UD-factorized version of the square-root-filter is used [Bierman 75]. Details may be found in [Wuensche 88; Mysliwetz 90; Bierman 77; Maybeck 79]. By exploiting the sparseness of the transition matrix in the dynamical model a speedup may be achieved.

Two interpretation phases have to be distinguished: First the initialisation phase when no previous knowledge about the scene is available, and second the continuous tracking phase, when objects have been recognized and their future behavior is being observed.

From features to physical objects in space and time

In the first phase, usually not time critical, like initialisation while at rest, regions in the image are systematically searched for feature groupings indicative of some known object (lower center of fig. 2). From the collection of features found, object hypotheses have to be generated as to which objects are being viewed under which aspect conditions.

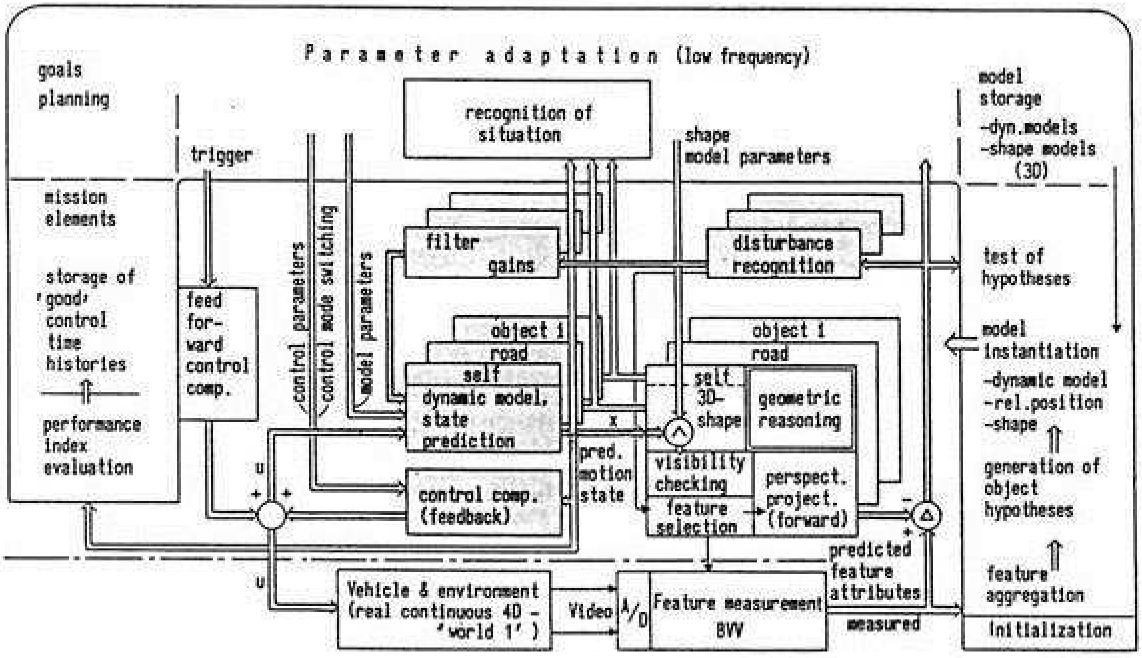

Figure 2: Gross flow chart of the 4D approach to real-time vision

Depending on the task context the higher levels to which the results of feature extraction are reported have to come up with hypotheses for generic objects fitting these data by proper parameter adjustment. Several such hypotheses will usually be generated. They allow to make specific predictions as to where which other features should be found if the hypothesis is correct. Checking these predictions, the best hypothesis will hopefully be arrived at by eliminating the less likely ones.

With this information, suitable dynamical models together with body-shapes and aspect conditions have to be instantiated in the recursive estimation loop (shaded blocks in center of figure 2, started by the right column of the inverted U-shaped outer frame). The dynamical models are then used to predict the cg-motion and body rotations around the cg. This information is combined with geometrical shape in order to determine the spatial position and orientation of well visible features. Their positions in the image plane are predicted and the feature extractors in the image processing system BVV are directed to these regions and orientations ('geometric reasoning'-block in lower center right of fig. 2).

The differences between measured and predicted feature data are used in conjunction with the filter gain matrix in order to update the predicted state variables after removal of disturbances recognized (upper right center in fig. 2). The temporal sequence of errors is also used for checking the validity of the hypotheses underlying the actual recursive computation. If consistently poor predictions are obtained, the corresponding hypothesis has to be adjusted; this may concern shape components, parameters in the dynamical model or the complete model. This part up to now has been implemented in a rather rudimentary form. For more complex dynamical scenes than the ones treated up to now, an object oriented data base (in the computer science sense) for a variety of physical objects (in the common sense) has to be implemented; this work has just been started (upper right corner in fig. 2).

A dynamical model has to be instantiated for each physical object capable of being moved. In road vehicle guidance this is not only the ego-vehicle and other vehicles but also the road, the appearance of which varies while driving upon it, at least in the general case with horizontal and/or vertical curvature. This is indicated in fig. 2 by the perspectively shown multiple boxes in the recursive center part.

The state of several objects in conjunction with environmental parameters and the active goal function of the ego-vehicle constitute a situation, to be discussed below. After recognizing the situation (center of upper bar in fig. 2) control modes or actual control time histories may be selected and implemented in an efficient way.

Reflex-like egomotion behavior

Since in the internal representation scheme chosen both the spatio-temporal state variables and the controls at the disposal of the system are explicitly represented, it is straightforward to apply the concept of state variable feedback in order to obtain optimal behavior for well defined tasks. Modern control theory provides the proven background for this approach. For each class of tasks, like lane following, convoy driving etc. in visual road vehicle guidance, a special feedback control law tuned to the actual dynamic parameters of the vehicle yields a characteristic behavioral mode.

Since the computation required is but a matrix-vector-multiplication, this simple operation can be done additionally at the lower level where the recursive state estimation is performed, thereby alleviating the higher levels from any involvement in high frequency control computation; in addition, this eliminates the incremental time lag which would have been introduced by the communication between the hierarchical levels required. With this workload sharing the higher levels may run at considerably lower cycle times (limited only by the requested lumped reaction time delay to some event requiring control mode switching). For systems with dynamical capabilities in the range of humans, several hundred milliseconds reaction time delay may be acceptable, while the recursive state estimation with reflex-like feedback control may run at 40 to 120 ms cycle time (two to six video cycles) typically.

In case a new event in the outside world requires special action, like the detection of an obstacle in the lane at a certain look-ahead distance, the upper decision level may trigger some predefined feedforward control time history (left in fig.2) with a set of parameters known to be able to deal with this new situation (for example either braking or lane changing).

The concepts up to this point have been implemented and proven to be very efficient computationally and robust enough for real world applications. The following sections deal with extensions under way and planned for the near future. The integrated 4D internal representation including time derivatives of state variables and the effect of control actuation over time yields a much richer background for action planning and prediction of the future evolution of the situation. Thus, based on fast forward simulation, temporal reasoning becomes relatively simple and complex situations may be handled in a straight forward manner.

Objects, subjects and situations

Before dealing in more detail with the notion of situations a brief review of the concept of subjects as introduced in [Dickmanns 89] will be given: Mobile entities in the observed outside world may be classified according to the fact whether or not they have the capability of activating some locomotion or perception system control at their disposal. There exists a large variety of systems with many shades of sophistication. Those which perform internal sensor data processing in such a way that control actuation is not directly coupled to measured data will be called 'subjects'. They are separated from the rest called objects (proper) because they require additional (internal or 'mental') state variables in order to completely describe their state. (Deliberately, no attempt is made to remove the grey zone implicit in this definition.)

For most real autonomous systems it will be impossible to determine their internal state completely. For most practical applications it will be sufficient to grossly know that part of the internal state of an autonomous partner which is relevant for the task at hand. This may be its actual 'view' of the situation, its actual goal function (or system of goal functions together with a likely control strategy) and its way of arriving at decisions in the situation as perceived.

Since usually all control decisions are based on more or less inexact estimates and since too many parameters of other systems are incompletely known, it seems wise to refrain from computing too detailed expectations of other subjects' behavior but only prepare reactions to the most likely ones; careful observation of the development of motion trajectories of the physical body of other subjects will give indications of its likely intentions. The most likely behaviors to be expected may be derived from decision and control strategies which oneself would adopt in the other subject's situation.

This way of defining a situation is in agreement with the one proposed in [Nagel 88]. Here however, the state of the objects and subjects is assumed to be known as good as possible through the recursive estimation scheme, and one is looking for a suitable control decision, the effect of which on the future evolution of the situation can be predicted by utilizing the dynamical models for all objects and subjects involved (assuming likely control inputs).

Mental states and intelligence

For an independent outside observer the internal representation of objects and their states in another subject constitute an increase in state variables of the entire system since the other subject may base control decisions on its actual 'view of the world'; these 'mental' states will then have their effect on the physical world when the resulting control action starts changing the real physical state of objects in the world. Therefore, these mental states are decisive factors in understanding situations; in the German language the word 'Wirklichkeit', usually translated as a synonym for 'reality', allows a different interpretation including these action-consequence effects: Ideas too may be part of 'reality' in the sense of 'Wirklichkeit' since they may effect changes in the evolution of processes in the real world. (The word 'wirken', from which Wirklichkeit is derived, means 'to effect changes or reactions'.)

Fixing the way how internal representations are arrived at, when sets of input data are given is therefore a decisive factor in the design and shaping of cognitive systems. [Maybe the hard core of human cultures is essentially an equivalent to this process on a very sophisticated level.] The richer an internal representation can be made by linking incoming data to predefined interpretation structures or to previously stored experience with different types of objects and subjects, the better will the system be able to deal with a variety of situations in the sense of achieving its goals despite perturbing factors. If rich interpretational schemes are available, a cognitive system may recognize situations or courses of actions from short subsequences, and it may be able to react early in an efficient, goal oriented way.

This capability seems to be at the core of the ancient definition of intelligence (if our Latin-teacher has been correct): The word 'intelligence'was claimed to have originated from the Latin verb 'inter-legere' meaning to be able to read in between of lines: those facts or hints which are not explicitly written down but which can be concluded from the context. Translated to the more modern usage of the word this would mean that a system could be called intelligent if it is able to recognize an action or a process sequence, especially a future one, from partial observations only; given an early correct interpretation, such a system would be able to also act early and adequately and to have advantages over lower performance competitive systems. This interpretation seems to be in agreement with the general usage of the word intelligence in everyday life. Note that this interpretation is a quite natural outgrowth of the basic approach taking spatio-temporal representations and the definition of controls in this context into account.

Especially with the sense of vision it is possible to apprehend situations 'at a glance' if typical arrangements of objects and subjects and short but typical action fragments can be observed. This, however, is only possible if the temporal domain is adequately represented by proper models.

System architecture based on the integrated 4D approach

In our vision system the main sensors are two passive monocular imaging arrays (CCD-cameras, black and white) mounted on a two-axis-platform fixed to each other with a given relative orientation. Their viewing direction can be controlled by the interpretation system according to its needs in the actual context; the controller is integrated into the image processing system BVV.

Based on the concepts discussed above the system developed has also a temporal structuring besides the usual structuring with respect to subtask hierarchies; both aspects will be discussed in the following subsections.

Temporal structuring

Video signal processing of course is linked to the 50 Hz video frame rate; this yields the basic cycle time of 20 ms for image feature extraction of which all slower cycles are integer multiples. The only faster cycle up to now is the viewing direction control for active vision and stabilization; it may use inertial angular rate signals at a small fraction of the video cycle time (typically 5 ms).

Recursive state estimation is done at the rate necessary for control computation: If the vision based automatic system is expected to have about the same dynamic range as the human operator, its corner frequency should be around 2 Hz. Taking sampled control theory into account, this results in a reasonable sampling frequency of 10 to 25 Hz yielding basic control cycle times from 2 to 5 video cycles (40 to 100 ms). The largest value means at a speed of 30 m/s (108 km/h) a new image every 3 meters, the smallest every 1.2 m. This is considered to be sufficient irrespective of the computing power available.

At this rate the complete physical state of all interesting objects is being recursively estimated. Using state feedback control laws, behavioral competences of the autonomous vehicle can be realized for different tasks and situations by simple matrix vector multiplication. This provides the vehicle with fast reflexlike behavioral modes without having to resort to the higher knowledge levels. Adding the capability of triggering proper control mode sequences as shown in figure 3 depending on simple situation indicators (some feature dependent rules), this may lead to yet relatively complex overall behaviors like lane driving with transitions to convoy driving or stopping and other combinations.

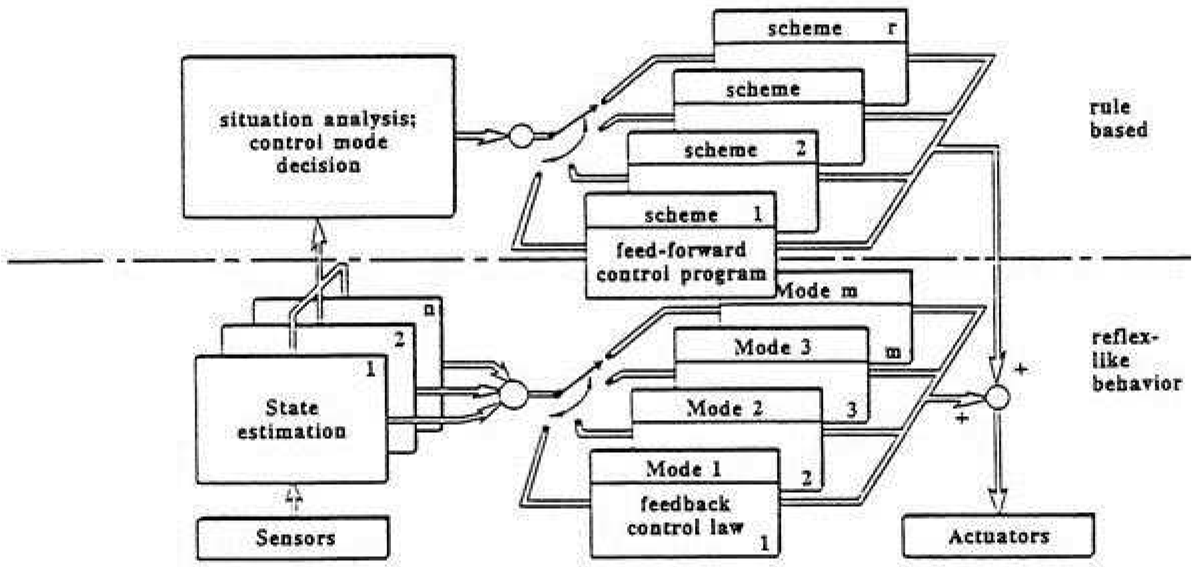

Figure 3: Selectable fast, reflex like feedback control determination with triggered feed forward components; situation dependent control mode decision

When such a pool of basic behavioral modes is available by the fast reacting lower levels the knowledge based higher levels may be allowed slower reaction times, perhaps down to the seconds-range. This figure would still be in agreement with average human performance.

In order to gain additional degrees of freedom for the complex visual perception task it may also be advisible to design overlapping specialised subtasks into the system which work at different time scales but at the same perception problem. One such task which is being studied in our system is the recognition of another object while in motion: There is one subtask which estimates the relative position and spatial speed components rather quickly (40 ms) taking only a very rough (2D) shape representation into account; a second subtask with a different group of processors tries to recognize the full 3D structure of the moving object at a much slower rate. Both may support each other by data or hypothesis exchanges.

On the upper knowledge based levels there is now more time for inferencing using background knowledge in the problem domain. At the same time, relevant environmental parameters may be evaluated and taken into account. In the normal behavioral modes the higher levels just have to monitor the performance of the overall system and to be alert to respond to new situations which may come up. Reaction times of several hundred milliseconds seem acceptable in comparison to human performance. Figure 4 shows the resulting hierarchical scheme.

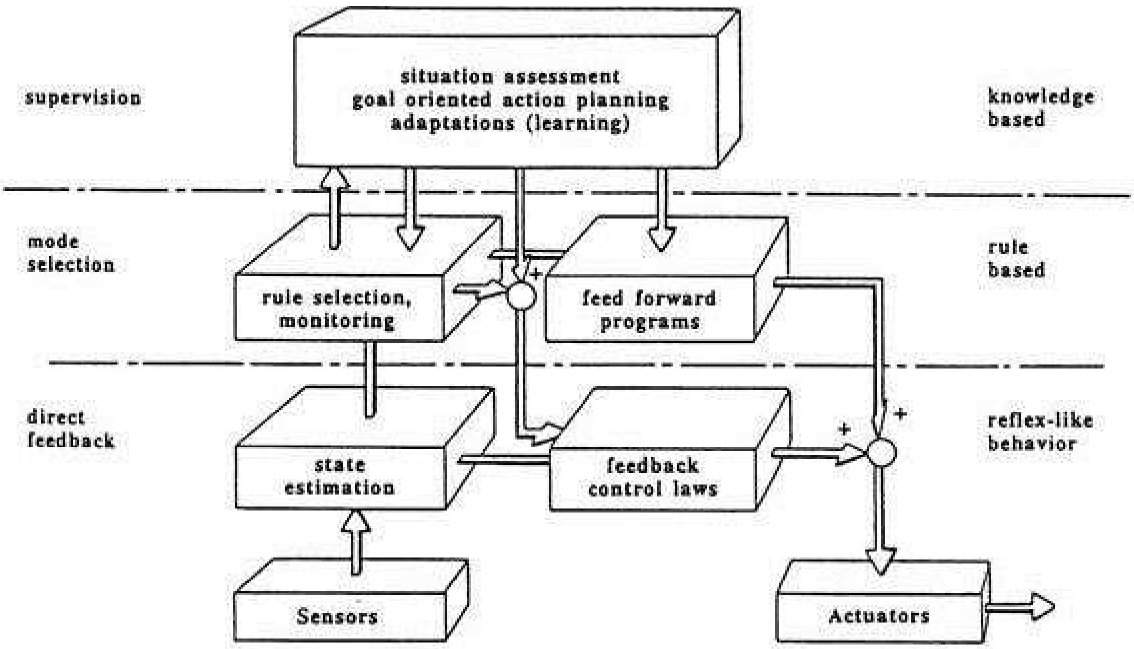

Figure 4: Hierarchical scheme for adapable fast control determinatiobn

Besides the different cycle times there is need for another temporal structuring in a (temporal) range sense. All measurements are taken and all controls are output in an exchange with the real world at the point 'here and now' in space and time, moving continuously and uninteruptably on the time axis. Contrary to the real world, the internal representation - also the temporal one! - can be halted and considered quasistatically. This is usually being done in logical considerations, leading to special problems when dealing with dynamical situations.

In figure 5 the internal representation density is shown in a qualitative way over the time axis. The sliding point 'here and now' is marked by the vertical line. In a temporal region around this line the internal representation of objects and the environment is kept and updated by recursive estimation exploiting stored knowledge about the processes observed in a fully dynamic spatio-temporal framework. Time histories of interesting state and control variables may be stored over a sliding short term interval

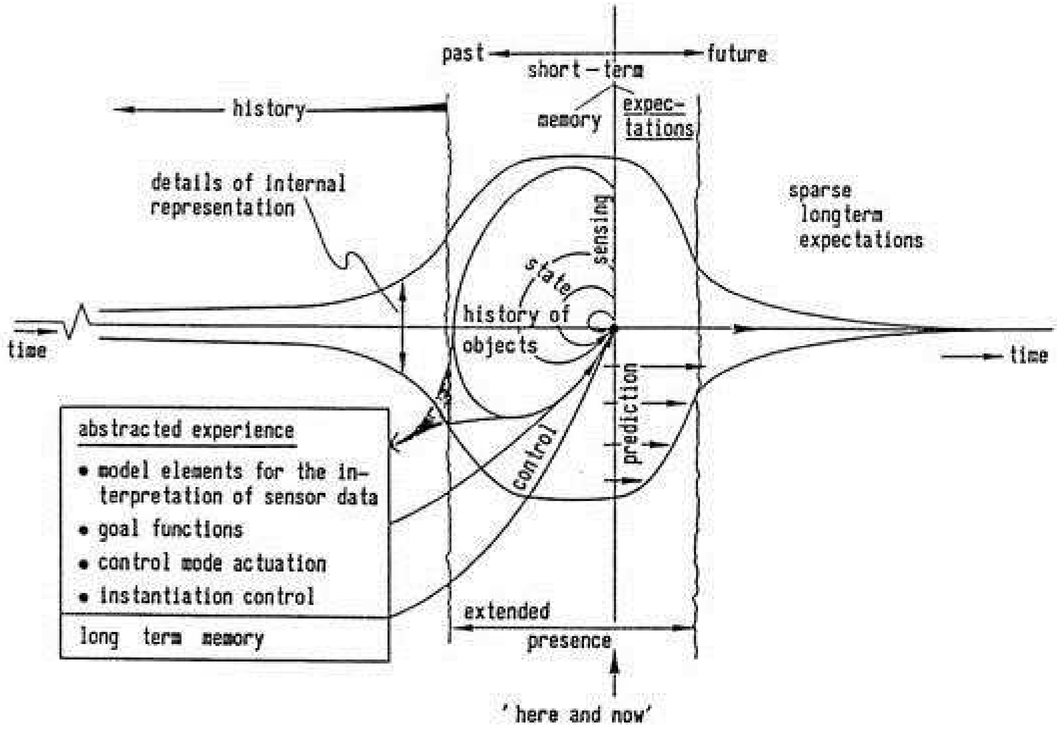

Figure 5: Temporal structuring for details of internal representation (qualitatively)

in order to be able to recognize low frequency process charateristics which may be of advantage for longer term predictions into the future. Prediction density varies with the time range: For one prediction step, all state variables will be predicted in the framework of the recursive estimation scheme for each single dynamic object supporting prediction error minimization. Longer term predictions may be of interest for only some objects, maybe even for only a restricted set of variables (e.g. estimation of collision probability). In order to make reasonable predictions for other subjects it is necessary to recognize its intentions, i.e. its likely control time history application in the framework of some goal it seems to be striving for; because there are so many uncertainties when subjects are involved, predictions usually terminate in the near future.

A somewhat different situation prevails with respect to the past. Here, process time histories when properly measured and stored will allow retrospective analysis correlating control input data with observed state histories; this may be used to derive knowledge about the specific system under scrutiny or for accumulating statistical data about objects and processes.

These activities may run in parallel on an additional processor using software packages developed in the field of control engineering, system analysis and systems identification; the resulting parameters may be used in the decision and control processes thereby allowing adaptations to changing situations and environmental parameters (for example roads on a winter afternoon turning from wet to icy).

In the long run, even more deeply structured temporal activities may be considered: Given the availability of proper software, the system may work on stored data time histories during periods where computing power is not needed for actual locomotion control (in parking condition). Several alternative control time histories and the resulting value of the goal function may be evaluated by simulation with the dynamical model available, for the situation considered. This 're-thinking' of situations with a reference outcome known meanwhile may lead to changes in decision parameters for future action, constituting one component of learning. Another form may be the retrospective comparison of maneuvers performed in the same situation with different control options showing the relative performance achieved; this would be the learning of appropriate behavioral decisions.

Typically during this process, the amount of data to be stored is reduced considerably leading to condensed descriptions of system characteristics (class properties, learning about facts and appropriate behavioral parameters). These characteristics usually are no more state variable time histories but system and control parameters or condensed average state descriptions (e.g. mean values, variances).

In this way, the 'present awareness subsystem' based on the 4D-approach working around the point 'here and now' (central blob in figure 5) can be exploited in several directions by the knowledge based subsystem shown in the rectangular box to the lower left; the latter one represents integral effects derived from experience over time for specific situations and tasks.

Hierarchical structuring

With respect to behavior control fig. 4 gave the resulting hierarchical scheme. Table 1 shows the hierarchical structuring with respect to measurement and scene recognition aspects. No special low level image preprocessing is performed; instead, the algorithms for feature extraction on the basis of controlled correlation [Kuhnert 88; Mysliwetz 90] are designed in such a way as to exhibit good noise reduction properties. Mainly, edge element and corner features have been used up to know. There is no final decision made with respect to 'optimal' features based on bottom up data only; accepted features for object interpretation are selected on the basis of an overall 'Gestalt'-idea derived from perspective mapping of an internal 3D shape representation (second line from bottom in table 1). At the single object level time is introduced via the dynamical models for 4D representation; up to now, no interframe differencing as in optical flow has been applied. The future has to show whether this type of image sequence processing will be necessary at all. (It is well known that nature in its biological systems does make use of it; this has triggered quite a bit of activities in this area also for technical vision systems. Whether and under which circumstances this is advantageous has yet to be determined).

	activity level	processors	operation	result
scene understanding	control level	MPS	compute expectations control viewing direction apply vehicle control	action
	task level	MPS	relative goal state evaluation	planning, decisions
	object level	MPS	situation assessment parameter adaption	situation
state estimation	feature level	GPP	feature aggregation	objects in space/time
	pel level	PP	feature extraction	features in image plane

Table 1: Modular processing structure for complex tasks

The levels discussed up to now have been implemented in the image sequence processing system BVV_2 [Graefe 85; Mysliwetz 90]. The scene understanding (upper) part in table 1 has been implemented on a PC-AT in the past and is being expanded and ported onto a transputer system presently. From several objects and environmental data the situation is recognized and checked against the requirements for task achievement. If no special action is needed the system is continued in its present mode; if some change of the operational mode becomes necessary a replanning is performed and the resulting mode change is triggered.

The control output is fed back to the internal representation via the prediction step updating all the lower levels, thereby adjusting the measurement and interpretation process to the actual state.

This frequent and fast traversion both bottom up and top down in the interpretation scheme assures efficient exploitation of both high level knowledge and most recent measurement data.

The gross flow chart corresponding to table 1 has been discussed already as figure 2 above. It has been arranged in such a way that the procedural recursive state estimation techniques using control engineering methods form the core of the figure while the more knowledge based higher level activities are grouped around this center showing the interaction paths.

A different viewpoint for subdivision showing other facets of the same system has been given at the end of [Dickmanns and Graefe 88]; the completely autonomous simulation capability inherent in this approach, and referred to already above, may even work without any sensory input normally being the driving factor. Stored data may possibly be taken as starting points or as reference trajactories to study variations around; interesting questions with respect to 'mind' and 'dreams' may come up.

Experimental results

The general dynamic machine vision scheme discussed has been developed during parallel application to four different areas after the idea had come up around 1980 in connection with the problem of visually balancing the inverted pendulum on an electrocart [Meissner, Dickmanns 83]. The first application oriented problem was planar docking of a reaction propelled vehicle with three fully independently controllable degrees of freedom [Wuensche 86, 88]. The second area was road vehicle guidance to be discussed in somewhat more detail below. The third one was a birdlike autonomous landing approach for conventional aircraft under visual flight conditions; this may be of interest for unmanned vehicles.

In May 1991 first flight experiments with a twin turboprop aircraft have been successfully performed in Brunswick; the system was capable of estimating its complete 12-

component state vector relative to the landing strip in real time fully autonomously from onboard machine vision.

Autonomously guided vehicles for factory floor transportation are the fourth application area; in this context, the capability of landmark navigation has been developed.

Road vehicle guidance

The application area of autonomous road vehicle guidance is by far the most developed one: A 5 ton van 'VaMoRs' of our University as well as a 10 t bus and a 7.5 t van 'VITA' of the Daimler-Benz AG have been equipped with our vision system. In experiments ranging over five years by now, the following capabilities have been demonstrated:

- Lane following at high speed: 100 km/h have been achieved limited only by engine performance of VaMoRs. On well marked empty freeways much higher speeds could be handled by the method; limitations may first come from camera resolution at large look-ahead ranges. Both horizontal and vertical curvatures can be estimated to sufficient accurarcy [Mysliwetz 90] to allow velocity control in order not to exceed preset acceleration limits.

- Lane following on unmarked cross-country roads with shadows from trees and buildings on the road. Speeds up to 60 km/h on empty roads have been demonstrated; even driving under light rain fall with wipers operating in front of the cameras has been shown.

- Night driving on well marked dry roads with normal headlights at low speeds has been performed with the Daimler-Benz bus on their test track.

- Driving on unsealed country roads at speeds below 20 km/h has been achieved by VaMoRs; however, in order to obtain more robust performance, computing power both for image processing and on the higher levels has to be expanded.

- Recognition of well visible obstacles of more than 0,5 m^2 cross-section (black trash can) in a look-ahead range of 30 to 50 m has been demonstrated at speeds up to 40 km/h on unmarked two-lane roads. The situation assessment level decides whether the vehicle is autonomously stopped at a safe distance in front of the obstacle or whether a lane change and passing maneuver is performed. Similar demonstrations have been performed with the Daimler bus stopping in front of another bus. Passenger cars can be detected at ranges up to 100 m with a 25 mm telelens.

- Convoying behind another vehicle has been initially demonstrated in our hardware-in-the-loop simulation facility, lateron with the test vehicles; 'stop-and-go' experiments are a special case of this capability shown in 1990.
- Lane changings to the left and right are the latest achievements.

Conclusions

Vision-based intelligent motion control should take advantage of the recursive state estimation techniques developed in control engineering. The '4D approach' developed at UniBwM over the last decade generalizes the extended Kalman filter to image sequence processing. In its sequential formulation it is well suited for solving major parts of the problem of dynamic scene understanding even under the condition of occlusion. The dynamical models are well suited for knowledge representation in the spatio-temporal domain.

It has been sketched how machine intelligence can possibly be developed based on the feedback scheme for motion control exploiting the high-level spatio-temporal world models which are at the core of recursive state estimation. In human history of science, dynamical models (i.e. differential eqs.) have been a rather late but very consequential achievement in understanding the world we happen to live in. This powerful insight in basic properties of processes in the real world should be exploited for making machine vision more effective.

Literature

Bierman, G.J.: 'Measurement Updating Using the U-D Factorization'. Proc. IEEE Control and Decision Conf., Houston, Tx., 1975, pp 337-346.

Bierman, G.J.: 'Factorization Methods for Discrete Sequenctial Estimation'. Acad. Press, New York, 1977.

Bryson, A.E. jr.; Ho, Y.: 'Applied optimal control'. Hemisphere Publishing Corp., Washington D.C., 1975

Christians, T.; Wünsche H.J.; E.D. Dickmanns: 'Schätzung der räumlichen Relativlage von Objekten durch Rechnersehen'. UniBwM/LRT/WE 13/IB/89-6, 1989.

Dickmanns, E.D.: 'Subject-Object Discrimination in 4D-Dynamic Scene Interpretation for Machine Vision'. Proceedings IEEE-Workshop on Visual Motion, Newport Beach, March 1989, pp 298-304.

Dickmanns, E.D.: Dynamic Vision for Intelligent Motion Control'. IEEE-Int. Workshop on Intelligent Motion Control, Istanbul, Aug. 1990.

Dickmanns, E.D.; Zapp, A.: 'A Curvature-based Scheme for Improving Road Vehicle Guidance by Computer Vision'. In: 'Mobile Robots', SPIE-Proc. Vol. 727, Cambridge, Mass., Oct. 1986, pp 161-168.

Dickmanns, E.D.; Zapp, A.: 'Autonomous High Speed Road Vehicle Guidance by Computer Vision". 10th IFAC World Congress, Munich, July 1987, Preprint Vol. 4, pp 232-237.

Dickmanns, E.D.; Graefe, V.: a) 'Dynamic monocular machine vision' b) 'Application of dynamic monocular machine vision'. J. Machine Vision & Application, Springer-Int., Nov. 1988, pp 223-261.

Dickmanns, E.D.; Christians, T.: 'Relative 3D-state estimation for autonomous visual guidance of road vehicles'. In: Kanade, T. e.a. (ed.): 'Intelligent Autonomous Systems 2', Amsterdam, Dec. 1989, Vol.2 pp. 683-693.

Dickmanns, E.D.; Mysliwetz, B.; Christians, T.: 'Spatio-temporal guidance of autonomous vehicles by computer vision'. IEEE-Transactions on Systems, Man and Cybernetics, 1990, Special Issue on Unmanned Vehicles and Intelligent Systems (to appear 1990).

Kailath, T.: 'Linear systems'. Englewood Cliffs, N.J. Prentice Hall, 1980.

Kalman, R.E.: 'A new approach to linear filtering and prediction problems'. J. Basic Engineering, 1960, pp 35-45.

Kant, I.: Werke in zehn Bänden, Hrsg. W. Weischedel, Wiss. Buchges. Darmstadt, 1983.

Kuhnert, K.D.: 'Zur Echtzeit-Bildfolgenanalyse mit Vorwissen'. Dissertation, Fakultät für Luft- und Raumfahrttechnik der Universität der Bundeswehr München, 1988.

Lindberg, D.C.: 'Theories of Vision from Al-Kindi to Kepler'. University of Chicago Press, Chicago and London, 1976.

Maybeck, P.S.: 'Stochastic models, estimation and control'. Vol. 1, Acad. Press, 1979.

Mysliwetz, B.: 'Parallelrechner-basierte Bildfolgen-Interpretation zur autonomen Fahrzeugsteuerung'. Dissertation, Fakultät für Luft- und Raumfahrttechnik der Universität der Bundeswehr München, 1988.

Nagel, H.H.: 'Overview on image sequence analysis'. In T.S.Huang (ed.): Image Sequence Processing and Dynamic Scene Analysis, NATO-ASI Series F2 Springer-Verlag, Heidelberg, 1983, pp 2-39.

Nagel, H.H.: 'From image sequences towards conceptual descriptions'. Image and Vision Computing, Vol. 6, No. 2, May 1988, pp 59-74.

Pontryagin, L.S.; Boltyansky, V.G.; Gamkrelidse, R.V.; Miscenko, E.F.: 'Mathematical theory of optimal processes'. New York, Wiley, 1962.

Schopenhauer, A.: 'Die Welt als Wille und Vorstellung'. In: Löhneysen, W. (ed) Arthur Schopenhauer, Sämtliche Werke. Suhrkamp, Frankfurt a.M., 1819 (Nachdruck 1986).

Vollmer, G.: Was können wir wissen? Bd. 1: Die Natur der Erkenntnis; Bd. 2: Die Erkenntnis der Natur. S. Hirzel Verlag, Stuttgart,1985.

Wünsche, H.J.: 'Bewegungssteuerung durch Rechnersehen'. Fachberichte Messen, Steuern, Regeln Bd. 20, Springer-Verlag, Berlin, 1988.

Zapp, A.: 'Automatische Straßenfahrzeugführung durch Rechnersehen'. Dissertation, Fakultät für Luft- und Raumfahrttechnik der Universität der Bundeswehr München, 1988.