Journal of Applied Intelligence 2, 251–270 (1992)

A General Dynamic Vision Architecture for UGV and UAV

E.D. DICKMANNS
Universität der Bundeswehr München, Fakultät für Luft-und Raumfahrttechnik (LRT)
Institut für Systemdynamik und Flugmechanik W. Heisenberg Weg 39, 8014 Neubiberg, Germany

Received January 31, 1992, Revised February 24, 1992

Abstract. The expectation-based 4D approach to dynamic machine vision exploiting integral spatio-temporal models of objects in the real world is discussed in the application domains of unmanned ground and air vehicles. The method has demonstrated superior performance over the last half decade in autonomous road vehicle guidance with three different vans and busses, with an AGV on the factory floor and with completely autonomous relative state estimation for a twin turboprop aircraft in the landing approach to a runway without any external support; in all application areas only a small set of conventional microcomputers was sufficient for realizing the system. This shows the computational efficiency of the method combining both conventional engineering type algorithms and artificial intelligence components in a well balanced way.

The modularity of the approach is demonstrated in a simulation set-up serving both the ground- and the air vehicle applications. Experimental results in both areas are discussed.

Key words: Machine vision, vision architecture, vehicle guidance, state estimation, modeling

1. Introduction

The problem of cognition has long bothered humans with respect to the dualistic aspects of the 'real' external and the 'mentally imagined' internal world. For many centuries, philosophers had tried to adjust the mental interpretation to the 'true' external world. It was I. Kant who, about two centuries ago after all the frustrating efforts over millennia, inverted the problem and herewith laid the foundation for a consistent scientific interpretation model. He recognized that reasoning has to start with the mind of an individual, embodied in a biological system having sensory contact to the real world outside. So, his main task was to clarify, what types of statements about the world could be made on safe ground. His main works 'Critiques of . . .' are dedicated to this problem.

He asked, what do we carry into the world with our sensing and analysis system, independent of objects and subjects which we observe? The conclusion was that 3D space and time are fundamental ('a priori') properties of our cognition system; they are not attributes of objects.

In evolution-oriented modern terms we would say that successful survival of our species is a solid foundation for the assumption that our sensing system is reasonably well adapted to the real world and that, therefore, we are justified to rely on the sensory signals from the outside world, in general. It is well known, however, that there are some flaws and that one has to be very critical with respect to the internal interpretations which have evolved over time from this. The multitude of cultural forms to be observed and the resulting different interpretation schemes for the same events (perceived signals) give an indication of the mental variety developed.

The idealist philosophers after Kant even turned the world upside down with their interpretation that the outside world is created by the mind. In this mental environment, it was A. Schopenhauer about 175 years ago who 'wanted to put the world back to its feet again' and who clearly stated the misinterpretations since Kant.

In his main work 'Die Welt als Wille und Vorstellung' (freely translated: 'The world as evolving process and internal representation') he emphasized and further elaborated the hypothesis of the dualistic relationship between mind and matter in one whole (unified) world.

Now that we are about to create machines with both material and mental components we should take advantage of basic insights gained formerly, which have been deepened in some scientific disciplines in the meantime. The natural and the engineering sciences have developed very good models for many processes in the real world exploiting the mathematical tool of differential equations (deq). Modern technology simply would not exist without numerical simulation on this basis. A set of deq describing a process in the real (4D) spatio-temporal world is called a dynamical model of the process.

Based on these dynamical models, recursive estimation techniques have been developed over the last three decades in order to optimally recover the process state from noisy measurements of some output variables. In this very successful technique originally due to Kalman [1], missing measurements, e.g., complete state variable time histories, may be substituted by knowledge via a model, observability given. The 4D approach to dynamic machine vision developed at UniBwM over the last decade extended the recursive estimation technique to image sequence processing. The required perspective inversion from the 2D image to 3D space is achieved at no extra cost once the Jacobian matrix of the imaging process has been determined.

The dynamical models are well suited for recovering the actual state of generically known motion processes; these models become useful for decision making only when solution integrals are known (at least some ones, possibly only locally valid); approximately valid solutions are often sufficient for practical applications. These solutions link process states over longer distances of time in the sense of state transitions. This step may form the missing link between the deq-based engineering type methods and the artificial intelligence methods which usually do not refer to time except for a time tag. This will be discussed in the next section.

The following section will then give a brief review on the 4D approach to dynamic vision. The essential points of view for grouping and modularizing activities in the overall approach will be treated next. The fast and efficient transition from high- level models to low-level features and vice versa will be discussed in the subsequent section exploiting the 'Gestalt' idea of psychology which has been incorporated into the approach. Then the resulting system architecture will be discussed. Following a section on the modular simulation facility developed at UniBwM for real-time dynamic machine vision, applied to a wide range of problem areas, the real-world experiments with road vehicles and aircraft will be reviewed.

2. The Sliding Point 'Here and Now'

A sensor system in the real world always is at a certain point in space at the one and only 'present' time; this point in time is part of a continuous 'time ray'. A physical object cannot be at two different locations at the same time. In order to move from one location to another, energy is required for ac- and deceleration, and time will go by because the energy available to effect locomotion is bounded. These facts constitute constraints on the motion process which may help considerably when tracking locomotion of objects, especially when ac/decelerations are very limited in magnitude as is the case in most of the occurrences in our natural and even technical environment (exceptions being bullets shot by guns or the driving hit of a golf ball for example).

Therefore, we are in a much better position to understand the next image of a real-time sequence if an internal representation is available which allows us to predict how the process under observation is going to evolve over time, taking certain control or perturbation inputs into account. If this prediction model is approximately correct one can concentrate the limited data processing capabilities on the data originating in the local environment of the predicted spot, thereby making the sensing process much more efficient; in addition, also the data processing algorithms may be adjusted to the predicted situation, thereby further increasing efficiency. This positive feedback favors the evolution of powerful prediction capabilities since, in spite of additional computing resources required for prediction, the overall requirements may be decreased for the same performance level; on the other hand, completely new performance levels and new qualities of deeper understanding of environmental processes may be achievable with this approach.

It might be argued, that human culture and its achievements are an outgrowth of nature having discovered this positive feedback during evolution of the human species.

In fig. 1 a qualitative display of internal representation density over the sliding time axis which moves from right to left is given. At the point 'here and now' (shown stationary at the cross-section of the two orthogonal axes) sensors provide data on the actual state of the real world. These data are interpreted taking high-level spatio- temporal world models into account. These dynamical models are derived from those developed for system design and analysis in engineering. In addition, it is taken into account that measurement data, usually, are superpositions of actual process states (the desired quantities to be recovered) and of measurement noise which is to be deleted. In order to be able to make this distinction, the models representing temporal behavior have to contain both the 'eigen-'characteristics (that means how states change over time when left on their own) and the response characteristics with respect to control- or pertubation inputs.

Once this is represented, predictions of the state evolution over time may be obtained at relatively low cost. Since usually neither the control nor the perturbation inputs of the future are known, prediction usually stops at one cycle (for the normal prediction-error-feedback state estimation process) or after only a few cycles in order not to incur too much uncertainty. For well known feedforward control time history inputs in order to perform some maneuver element (for example lane change in road vehicle guidance with a sine-like steering angle time history input with appropriate parameters for period T and amplitude A), reliable predictions over longer temporal ranges (seconds) are possible. Taking standard perturbation statistics into account, even longer ranges of prediction over entire maneuver sequences may be meaningful, at least for some aspects of the process (like prediction of

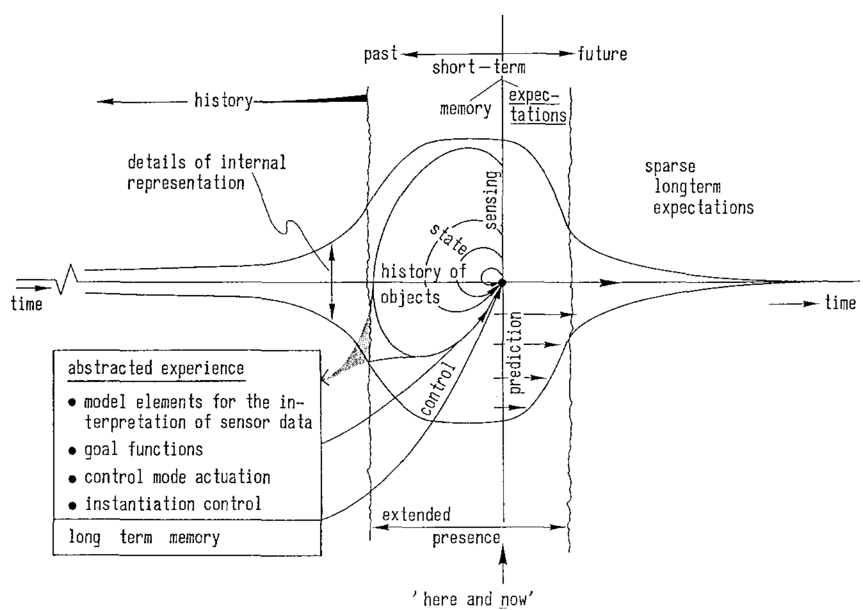

Fig. 1. Temporal structuring for details of internal representation (qualitatively)

the time needed to go from point A to B). In the average, however, the number of predicted events will vanish on the future time scale to the right.

If good internal models are available for generating rich actual internal representations from the actual data measured, it will be impossible to store all these data as a 'personal history of adventures'; it is not necessary, though. Since the time histories of all state variables may be regenerated from stored initial conditions and control as well as perturbation time history inputs, once a proper model for the dynamic behavior is available, only the latter (reduced) data sets need be stored. For these again, instead of pointwise storing each individual time history, parameterized generic models as functions over time would allow very efficient storage since a dense data input vector may be replaced by a few parameters needed to feed the proper function call. This shows that proper temporal models may be very efficient in reducing memory requirements if things are properly organized. Past process state time histories and events may then be reconstructed actively from combining only a few stored historical data with stored model knowledge. This principle is the basic advantage of the 4D approach combining space and time in an integrated manner.

This type of data compression into valid models is symbolized in fig. 1 by the formation of a reduced tail on the past time axis (left). Quasi-static knowledge resulting from this is used later on for triggering proper control activities depending on the situation encountered. Standard perturbations are counter-acted by feedback control laws which are implemented by a direct loop from the sensory data to the corresponding actuators via internal state variables of recognized objects (see center of fig. 1); this allows stable behavior under perturbed conditions without the explicit knowledge levels having to interact with the high frequency data stream. Only unforseen situations and unpredicted new features discovered lead to an activation of the more knowledge based hypothesis generation part controlling the active set of internal dynamical models (one for each object, lower left in fig. 1).

Seen from this point of view, the entire 'mental' internal world of representations has as its purpose to provide the system with capabilities of data interpretation well suited for control outputs which enable the system to achieve its goals; previous experience may be exploited for this purpose contributing to the rating of a system as being intelligent or not.

3. The 4D-Approach as the Core of Expectation-Driven Vision

The dynamical models link time to spatial motion, in general. 3D shape models exhibit the spatial distribution of visual features which allow objects to be recognized and tracked. In order to exploit both dynamical and shape models at the same time, the prediction error feedback scheme for recursive state estimation developed by Kalman and successors in the 1960's has been extended to image sequence processing by our group [2]. There are many publications on this approach so that only a short summary will be given here (see e.g., the survey articles [3,4]).

Fig. 2 shows the resulting overall blockdiagram of the real-time core of the vision system based on these principles. To the left, the real world is shown by a block; control inputs to the vehicle carrying the camera may lead to changes in the visual appearance of the world either by changing the viewing direction or through egomotion. The continuous changes of objects and their relative position in the world over time are sensed by CCD-sensor arrays (shown 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 several ways: There is a limited spatial resolution in the image plane and a temporal discretization of 16 2/3 or 20 ms, usually including some averaging over time. Bifocal vision with a wide angle lens for a large viewing area nearby, and a tele-lens for good resolution further away, has become standard for road vehicle applications; active

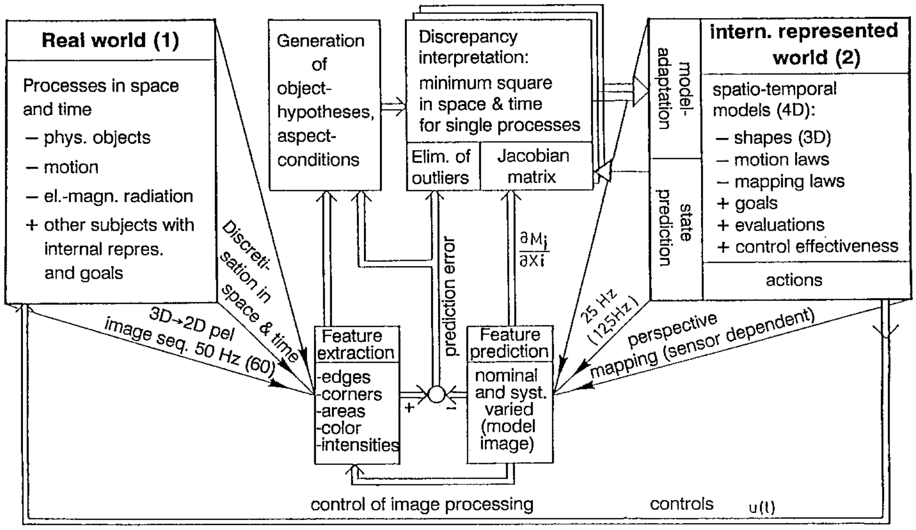

Fig. 2. Survey block diagram of the cybernetic 4D approach to vision

viewing direction control with fast pan and tilt platforms has been studied and applied since 1984.

Instead of trying to directly invert the perspective projection in the image sequence for 3D-scene understanding, a different approach by analysis through synthesis has been selected. Based on previous human experience, generic models of objects in the 4D-world have been installed 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 *p*icture *el*ement (pel) processing (lower center left in fig. 2), object hypotheses including aspect conditions and motion models in space (transition matrices) have to be generated (upper center left). They are installed in an internal 'mental' world representation intended to duplicate the outside real world, at least in the aspects relevant to the task at hand. This is sometimes called 'world_2' as opposed to the real 'world_1'.

Once an aggregation of objects has been instantiated in the world_2, exploiting the dynamical models, the object states can be predicted for that point in time when the next measurement is going to be taken. By applying the *forward* perspective projection to those features which will be well visible, using the same mapping conditions as 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. 2). 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 with respect to state changes can be determined, the so called Jacobian matrix (upper block in center right, lower right corner). This rich information is used for bypassing the direct perspective inversion via recursive least squares filtering through feedback of the prediction errors of the features. This means that the perspective inversion can be achieved at no extra cost once the Jacobian has been computed; based on the dynamical model, the **spatial** state is estimated in a least squares error sense including its **spatial** velocity components. Note that this signal to symbol transition from pels via edge features to a high-level spatio-temporal object state is achieved in just two steps; however, both data driven bottom-up and model-driven top-down components are traversed in each of the frequent (12.5 or 25 Hz) cycles in this approach, as will be explicated in the next section.

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; one might consider the method to exploit a 'virtual' optical flow, if one likes this notion, where the reference is stored not in a previous image but in a symbolic, predicted spatio-temporal state of objects in 4D;
- the transition from signals to symbols (spatio-temporal motion state of objects) is done in a very direct way, well based on scientific knowledge, that is 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 extraction algorithms can be provided with information for more efficiently finding the desired features; outliers can easily be removed, thereby stabilising the interpretation process; the 'Gestalt' idea of 'objects in motion' allows the elimination of combinatorial explosion in feature aggregation which otherwise would hamper object recognition in natural environments (especially in environments with many shadows);
- viewing direction control can be done directly in an object-oriented manner; known egomotion can be compensated for in order to achieve better fixation performance.

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 [5]. Details may be found in [2, 6–8]. By exploiting the sparseness of the transition matrix in the dynamical models a speedup can be achieved.

Special care has to be taken in the initialization phase when good object hypotheses are in demand. From feature collections which may have been compiled in a systematic search covering extended regions of the image, the existence of objects has to be hypothesized.

4. Structuring Points of View

The main problem in image sequence processing is data reduction without loss of information. Humans (and possibly most biological systems) have good capabilities in this respect: an image of homogeneous gray values is immediately characterized as uniformly gray. In a 1K by 1K pel image with 1 Byte (B) per pel for intensity coding, the amount of 1 MB of data is directly reduced to a few Bytes of information by specifying the average gray value and the range of validity 'everywhere'. In an average image, this is not true all over, but regionwise for certain areas, usually in correlation with object surfaces. Therefore, one might say that the information in an image rests in *nonuniform* intensity changes (or, in general, in color intensity changes). That it is not just the intensity gradients by themselves, may be seen immediately from regular repetitions of intensity or shape patterns; they have received a special term and are called 'texture'. It is always the relationship from local to global which determines efficient characterization.

With the processing power available in the past, there was no hope to achieve full image processing at video rate with an extraction of texture features; therefore, we confined ourselves to intensity edge features as the carrier of information in the image plain. Larger regions with (constant) gradients may be handled efficiently by resorting to pyramid image representations.

Knowledge about the real world has been accumulated by human culture around the physical units termed objects or subjects. In order to understand more complex processes involving several objects or subjects, the term situation is used for grouping classes, which may be handled in a similar manner. In correspondence to this well proven scheme, the structuring points of view of features, physical objects (subjects) and situations have been adopted in the integrated 4D approach.

4.1. Features

Since the absolute intensity level of light in outdoor scenery may vary very frequently due to sunshine, clouds and time of day as well as year, robust methods for image processing should concentrate on invariants in the image plane at least approximately independent of these changes. The most prominent ones are intensity edge features measured against a common standard and evaluated relative to each other. Ternary correlation [3, 6, 9] has been developed to a standard tool. The presence in abundance of neural receptive fields in vertebrate vision systems performing similar functions hint to the adequacy of this operator.

Correlation masks of different length (from 3 to 16 pel) and orientation (up to 32 for a half circle) are being used, depending on the situation, for searching local correlation extrema along simply defined search paths (horizontal, diagonal to left and right, and vertical). Corners may be evaluated by proper coupling of two edge masks.

This technique coupled with intelligent control of the evaluation parameters exploiting the 'Gestalt' idea for objects, to be discussed in the next section, allows efficient real-time image sequence processing with today's microprocessors.

4.2. Objects

The units to which humans attach their knowledge about the real world are objects, usually characterized by being coherent entities; they may be composed of parts and may be parts of larger units. It has taken a relatively long time until in computer science an equivalent construct has been introduced; now that it is available as a programming paradigm, it is rapidly taking over for handling complex tasks.

Objects may be stationary or movable with respect to angular orientation or translation; with respect to shape they may be rigid, flexible or deformable. Stationary objects are completely described by their 3D shape, orientation and location. Movable objects need a full rotational and translational state vector including the velocity components at each point in time in order to be fully defined; when at rest, they may be treated as stationary as long as no forces or moments move them.

When moving, objects usually have a specific characteristic over time in addition to shape. Therefore, in order to be efficient with respect to motion analysis, both these properties constituting essential knowledge about objects should be taken into account; this is captured in the structure and the values of the parameters of dynamical models. Through these models, time is introduced directly into the interpretation system as a basic property characterizing processes in the real world.

4.3. Subjects

This notion has been introduced in [10] in order to be able to properly characterize systems capable of starting actions in the real world 'at will'. Exploiting high-gain power magnification in properly constructed units, a transition from the information level (micro- or milliwatt) to real-world control actuation (kilowatt-level) is possible. An electro-hydraulic power amplifier, for example, bridges six orders of magnitude in power in just one unit.

This enables autonomous systems to become active dependent on sensory signals and on the results of their processing. Therefore, in order to completely describe such a 'subject', its internal 'mental' state has to represented and to be known in addition to the states needed for objects.

Specific subjects usually react according to some generic prototypical schemes; if the basic structures of these schemes and their triggering conditions depending on the external situation are known, the reaction of a subject may become predictable, at least in a fuzzy (probabilistic) way. Including the effects of control actuation into the dynamical model as is usually being done, movements to be expected may be predicted which again may alleviate the real-time data processing task if properly organized. Therefore, the control actuation patterns available to subjects and their triggering conditions constitute essential knowledge for intelligent autonomous systems; they should be exploited for characterizing specific classes.

4.4. Situations

An arrangement of objects and subjects in the framework of a specific task context is called a situation. Note that due to the internal states and to the task dependence, the same arrangement of objects and subjects may constitute a different situation for the different subjects.

All this knowledge combined yields the background against which an intelligent subject will evaluate the sensory signals received in order to come up with good results in dynamic scene understanding. In an actual situation, the 'Gestalt' idea including all dynamical aspects of other objects and subjects is the unifying consideration.

5. The 'Gestalt' Idea for Object Recognition and Tracking

In order to be able to efficiently recognize and track objects in image sequences of a dynamic environment, their spatio-temporal invariances including the constraints resulting from the mapping process should be exploited.

Temporal invariances are captured in the dynamical models which link the changing aspect conditions over time. Spatial invariances are the 3D shapes of rigid objects under observation. Combining both via the laws of perspective projection yields the basis of dynamic scene understanding with the 4D-approach. Thus, dynamic image sequences are interpreted by exploiting a priori knowledge on objects and relevant motion processes in 3D space.

Measurable features on the surface of objects constitute the link between picture element (pel-) processing in the image plane and a symbolic reconstruction of an analog internal representation of the mapped object in the interpretation process within the computer.

Measuring and collecting all features in every image and then trying to match groups of those with all possible object interpretations in every frame would lead to combinatorial explosion. Instead, relying on the temporal continuity conditions in certain task domains and motion processes, considerably reduces the number of meaningful interpretation possibilities. Objects move steadily and usually in a well predictable manner; images of objects behave similarly except for occlusion effects. If an independent motion of the camera is superimposed on dynamic changes in the scene, an internal representation of the camera motion as linked to another bodily motion process (egomotion) is required.

For an autonomous vehicle, therefore, both the egomotion and the motions of other objects have to be represented separately. Both motion components may change the aspect conditions of the objects observed. Inertial measurements may give independent information on the egomotion in a fast and reliable way so that image interpretation can be alleviated taking this information into account. The corresponding effects of changing aspect conditions of other objects through egomotion may thus be eliminated, at least roughly. This reduces the search area for an object in the new frame. Actively predicting the remaining changes in aspect conditions and corresponding changes in feature appearance for a known 3D shape (not 2D in the image plane which is not invariant) allows to considerably reduce the amount of image processing workload.

In natural environments under sun shine condition, shadow boundaries may yield much more pronounced intensity edges than do body boundaries; for moving objects these shadow boundaries move over the body surface. Under these conditions the 'Gestalt' idea of looking for a reasonably coordinated set of features indicative of the object searched or tracked, oftentimes is the only chance for *re*-cognizing the object amongst a multitude of distracting similar features. One has to know what one is looking for in order to find it. This approach will be detailed in the sequel for road and road vehicle recognition in outdoor scenery. It allows to interpret rather complex scenes with comparatively simple image processing schemes, however, with frequent bottom-up and top-down traversal of the system hierarchy encompassing low- to high-level processing and knowledge representations (12.5 resp. 25 Hz).

In [11] this approach has been extended to generic object models with unknown shape parameters. First simulation results indicate the viability of this approach for simultaneous motion and body shape recognition, at least for some restricted classes of technically fabricated objects like road vehicles.

6. System Integration

The method described above lends itself to a well structured system integration as shown in table 1. This is being implemented at present on a

	activity level	processors	operation	result
scene understanding	control level	MPS	compute expectations control viewing direction apply vehicle control	action
scene understanding	task level	MPS	relative goal state evaluation	planning, decisions
scene understanding	object level	MPS	situation assessment parameter adaption	situation
state estimation	feature level	4D-OP	feature aggregation	objects in space/time
state estimation	pel level	PP	feature extraction	features in image plane

Table 1. Modular processing structure for complex tasks

transputer network added to the existing BVV/PC-system (compare [3]). From each of the feature extraction groups called 2D object processors, a transputer link exists into the network for communicating data in parallel between the higher and the lower levels. The 4D object processors are all T800-transputers at present. Once transputers had been integrated into the system, they started to take over also lower level functions from the BVV. Therefore, the system is likely to become an all-transputer system in the near future.

Feature extraction (lowest line in table 1) and its control from the higher levels has remained unchanged, in essence. For the task of 4D object recognition, specialists for certain classes of objects like lanes, crossroads and vehicles have emerged (4D- OP, formerly GPP); at this level (second lowest line), time is introduced into the interpretation process and exploited for avoiding the combinatorial explosion during feature aggregation to objects. A large amount of implicit knowledge about the appearance of real world objects in perspective images is represented at this level.

A dynamical object data base collecting all information on actual objects constitutes the interface to the higher levels. There, the situation is assessed taking the own task into account (middle line, table 1); decisions are being made whether to continue with the control mode running or to switch to a different maneuver element (second line from top). Note, that these high levels do not directly act on the controls but that there is a special control level implemented which closes the loop (shown in top line, table 1; a more appropriate way of displaying the physical organization and the signal flow is given in figures 3 and 4 to be discussed below). In order to incur as little time delay as possible into the actuation process, feedback control works directly upon data in the dynamical data base. This way, the higher levels are somewhat decoupled from direct actuation, alleviating them from hard cycle time constraints. The control output is, however, fed back to all internal state representation and prediction instances in order to compute corresponding expectations.

Fig. 3 shows an overall block diagram which better indicates the signal flow in the system. The inner core represents the lower two levels and the uppermost one of table 1. The shaded areas implement the recursive estimation and single step prediction of object center of gravity states (engineering type methods). The 'geometric reasoning' block (lower right center) adds the shape aspects. This and the peripheral functions (outer inverted U-shape) might be called the artificial intelligence parts of the system. Initialization (right bar of inverted U in fig. 3) is achieved by matching feature groups with projections of an hypothesized object. Three ingredients are always necessary for instantiation of an object model: the dynamical model for the temporal evolution of motion, the shape model for feature distribution around the center of the body, and the aspect conditions for computing the perspective projection. The generic object models will be kept in an object data base under development (upper right corner).

The upper bar of the inverted U in fig. 3 represents the real-time monitoring and decision making part of the system. If prediction errors (arrow upwards between the inner and outer right blocks) are consistently large, object hypotheses will have to be adjusted; this may be done by parameter changes or by switching to a completely different model. The method presented in [11] may allow to perform the parameter adjustment for known generic objects in a recursive way similar to and with the same methods as for state estimation.

Since the scene is time varying and new fea-

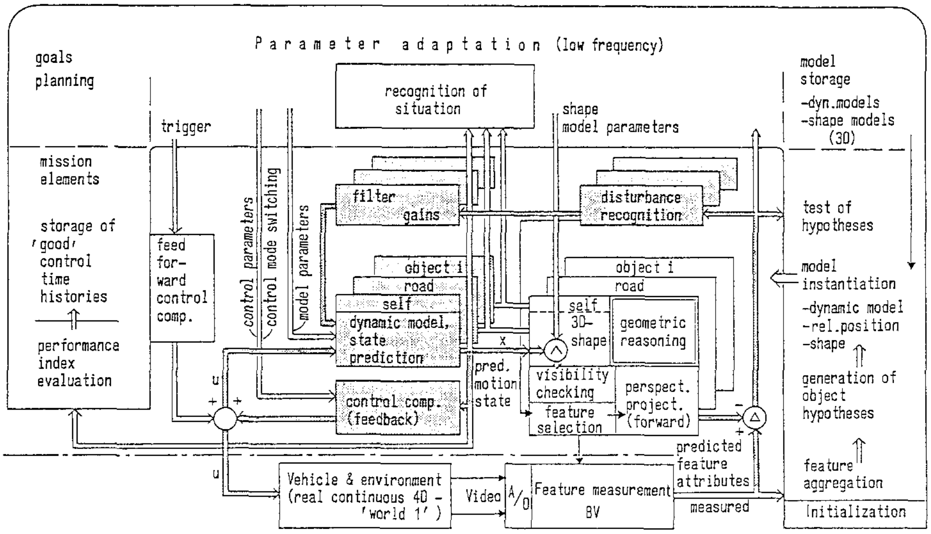

Fig. 3. Block diagram of dynamic vision; control oriented AI (4D prediction error minimization)

tures belonging to yet unknown objects may occur, a steady monitoring has to be done in order to detect those and to generate proper object hypotheses. In ambivalent cases, several object hypotheses may have to be started in parallel; they will be pruned over time when enough information has been collected for dropping the less likely ones.

The main task of the situation recognition level is to trigger proper use of the generic control procedures in the inner block for transforming measurements into actuator outputs. This may be done by either feedback or feedforward modes. The regulatory control tasks like lane keeping and convoy driving are realized by state feedback (lowest shaded block); maneuvering control tasks like lane change and turning off are realized by parameterized feedforward control time histories (block attached to the right of the left vertical bar in fig. 3). Proper feedback may be superimposed during later parts of feedforward control maneuvers for compensating disturbances which might have occurred. All control outputs are sent to both the real-world vehicle and the internal models (lower center left).

The left vertical bar is reserved for future autonomous learning capabilities based on state time histories experienced (lower input to left bar), caused by control time history inputs (attached block to the right). Both have to be evaluated in conjunction with respect to some performance index (goal function for mission elements) in order to be able to select favorable control time histories. More information on the intelligent control aspects may be found in [12], from which fig. 4 shows yet another way of displaying the system architecture; it emphasizes temporal structuring with low cycle times at the bottom (full or 1/2 to 1/4 video rate for feature extraction and 1/2 to 1/4 video rate for state estimation and control computation) and large ones (possibly asynchronous) at top. The intermediate level is purely event-driven.

7. Implementational Aspects

The overall system architecture sketched above has evolved over a decade of research in the field of autonomous visual guidance of vehicles based on the engineering approach of recursive estimation. The artificial intelligence aspects for visual object recognition had to be incorporated into this approach.

There are two distinctly different phases of operation: 1. The initialization phase, when nothing is known about the scene to be interpreted, and 2. the real-time operation phase in which spatio-temporal models are available, alleviating the frame-to-frame interpretation considerably.

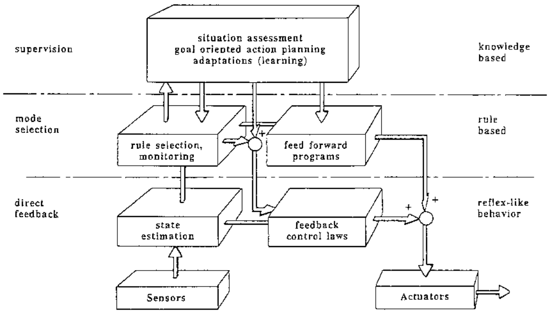

Fig. 4. Hierarchical scheme for adaptable fast control determination with temporal structuring

Phase one is much more difficult than phase two, and solutions are available only for very special cases and task domains. On the other hand, during initialization, usually, there is no strict time limit for the orientation to be performed; it may last for several seconds or even minutes, whereas in real-time operation the cycle time is less than one tenth of a second, normally, if the performance level of humans is taken as standard for comparisons.

Since initialization may be very involved, it is not discussed here due to space limitations; the interested reader is referred to [6] for road vehicle initialization. For the example of real-time road recognition with shadows on the road, the realization of the 'Gestalt' idea and the handling of uncertainties in object recognition will be discussed in this section.

Fig. 5 shows a campus road in summer with shadows from trees crossing the road almost normally; due to the leaves, there are almost homogeneous shade areas on the road. From previous interpretations, the system knows the lane width, the curvature parameters of the road and the own position and orientation relative to the lane center line. With the known elevation of the camera above the ground and the known camera mapping parameters the appearance of both lane border lines in the next frame can be predicted taking the last control input to the vehicle into account. From lateral position on the lane and from the curvature parameters, areas of interest for the collection of meaningful information on both lane boundaries are predicted; the digitized video image is stored by the feature extraction (parallel) processors in these regions only. The predicted slope of the boundaries is used for selecting proper mask orientations for controlled ternary correlation in each window individually; due to road curvature which is part of the

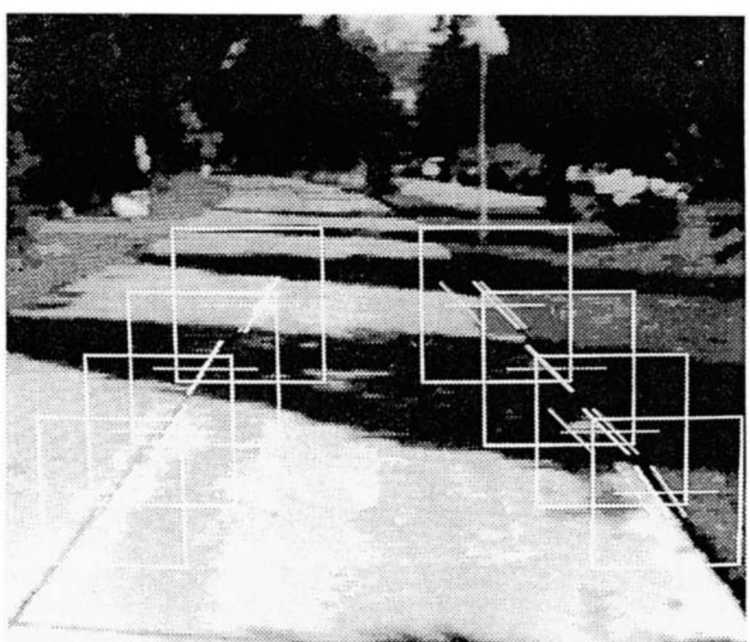

Fig. 5. Application of the 'Gestalt' idea to road recognition

'Gestalt' idea, the tangent direction changes smoothly with look-ahead distance, that is with decreasing line number in the image. The correlation, then, is performed in a restricted region around the predicted spot along a search path, in this application usually along the horizontal axis; the length of the search path may be adjusted according to the uncertainty of the prediction, which of course is a function of the cycle time. This may lead to the fact that shorter cycle times with fewer correlation evaluations to be performed, may be advantageous relative to longer ones with high search loads and the increasingly difficult problem of feature correspondence from frame to frame.

Since correlation of larger areas is more immune to noise corruption, long masks of 16 or 17 pel edge length have proven to be beneficial in natural environments with high spatial noise frequencies. Note that in fig. 5, due to this selectivity in intensity gradient direction and window position, the amount of features and data to be handled could be reduced considerably.

For each lane boundary feature to be determined, a couple of candidates will be delivered by the extraction processes; these perform the search for the predicted and the two neighboring mask orientations. It is not even tried to decide on this level which of the candidates are good ones and which are not; the information for this decision is missing on this purely data driven level. (Therefore, from this point of view, the question of an 'optimal' feature detection does not make sense. For example, in natural scenes with hard sunshine, the shadow boundaries very often yield higher intensity gradients than the transitions from road to shoulder.) In fig. 5 some candidates for edge features are marked by line segments.

Up to 24 feature candidates from 8 windows may be presented to the interpretation process. From these, at most one per window will be selected yielding a least squares error sum fit for the curvature model with slightly adjusted parameters (adaptation of the 'Gestalt' idea). If the measured feature position in some windows is outside the 3 sigma region of the predicted value, the contribution from this window will be discarded for the actual cycle; sigma is the standard deviation evaluated as a side product during recursive estimation. The measured feature position will also be discarded if the correlation value is below a preset threshold in magnitude. For this reason, a many-fold redundancy in feature extraction has been introduced with the specification of four window areas on each side of the lane. In the case of fig. 5, the measured features of only 5 windows have been considered sufficiently good for inclusion.

In fig. 6 from a winter day with low standing sun, there are many perturbations from branches without leaves yielding many spurious features; in cases like this one, the features of more than half of the windows may be rejected because the many shadow boundaries prevent the real road (lane) boundary from being extracted. With the BVV_2 image sequence processing system [3, 6], speeds up to 30 km/h have been achieved under these conditions. In these situations, longer look-ahead ranges with a tele-lens and a wider range of feature extractors (multiple scales) would probably yield better results. However, without the 'Gestalt' idea projected into image evaluation and the uncertainty treatment by the recursive estimation procedure, situations like these could hardly be handled with that low computing power in real time at 25 Hz.

These types of images also are very hard for initialization because the shadow from the trunk of the tree yields an extended linear intensity gradient and would well qualify for a candidate of one road boundary; only the missing adjacent boundary according to the 'Gestalt' idea of a perspectively mapped road, with a pair of converging lines including a large area nearby, allows to eliminate these shadow boundaries from the candidate list.

In the rest of the paper, the spatio-temporal modeling and experimental results in two application examples will be discussed.

8. Application Examples

The area of autonomous mobile robots or unmanned vehicle systems has attracted the attention of many researchers over the last years. The majority of activities seems to be in the ground vehicle area, but also underwater, air and space vehicle applications are being studied; numerous

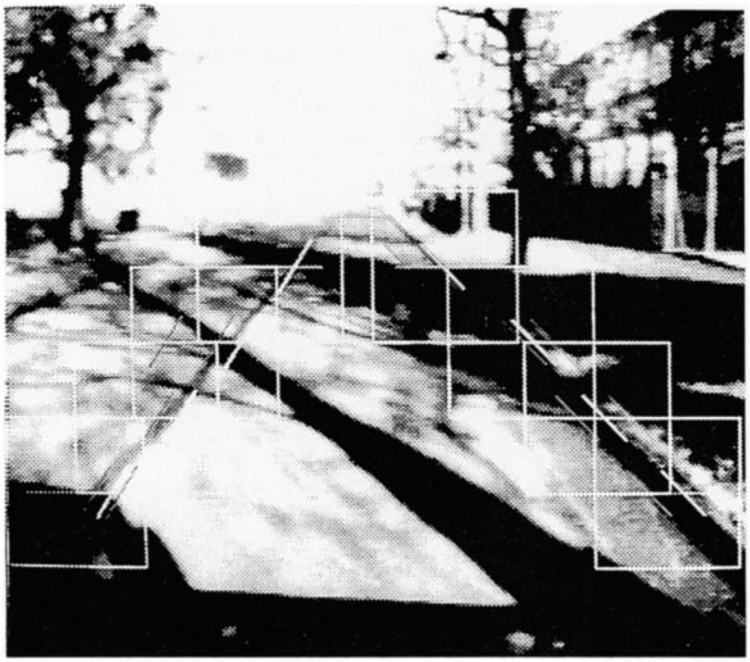

Fig. 6. Difficult winter scene with shadows on the road

conferences and technical journals are devoted to these fields. With [13] a collection of articles giving an overview over major ground vehicle activities has appeared just recently; references to many of the existing projects may be found there. In the USA, the DARPA project on autonomous land vehicles drew much attention; in Europe, EUREKA-projects like 'Prometheus' and ESPRIT projects have spurred much interest in road and ground vehicle guidance. All of these have fostered activities in this application area worldwide.

In the sequel, some of our application results with the 4D approach to dynamic vision will be discussed.

8.1. Real-Time Hardware-in-the-Loop Simulation

For software development it is essential to have facilities available which allow exact reproduction of test conditions; with vehicles in real world environments this is almost never given. Therefore, it has been decided in 1976 when dynamic machine vision as research subject has been picked, to build a unique simulation facility for this purpose, similar to those known for human pilot training in aviation. Fig. 7 shows the actually existing version of this simulator which, in the meantime, has been used in various configurations over more than a decade for developing visual dynamic scene understanding and control algorithms both for road vehicle guidance and for the aircraft landing approach. The rotational and translational motion of the system to be investigated (road vehicle or aircraft dynamics) is simulated numerically in all relevant degrees of freedom on the simulation processor exploiting a nonlinear set of differential equations (upper left block in fig. 7). Input to this simulation process is exactly the same control output as will be used for the real vehicle. This control input is numerically integrated over one step in time (40 ms typically), in order to obtain the next set of state variables; perturbations may be added during this process for improved similarity to real world processes (like gusts in aircraft flight). For this state, the camera position and orientation is computed; these 'eyepoint conditions' determine the perspective view on the synthetic landscape in the graphics processor of the Computer Generated Image system (CGI, upper right block in fig.

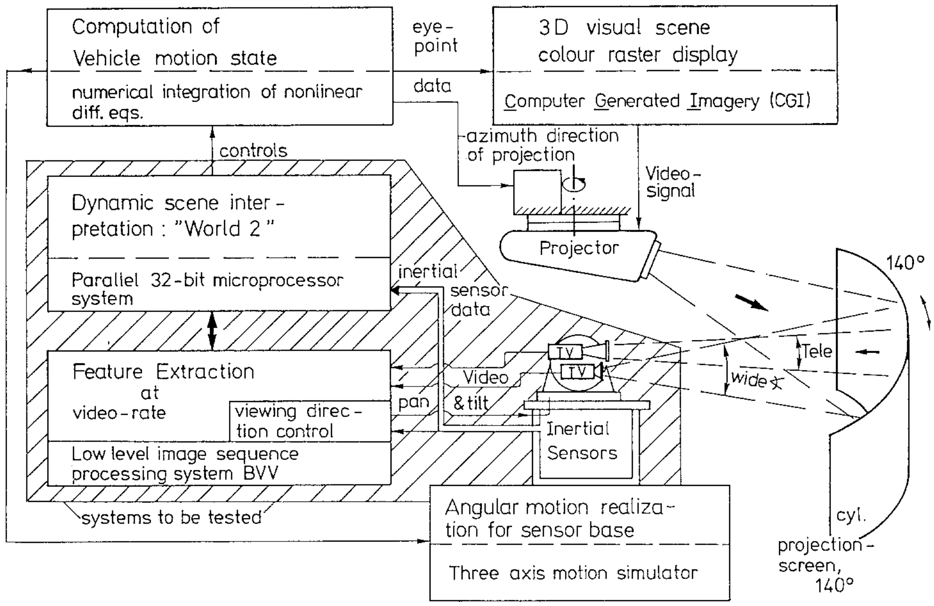

Fig. 7. Hardware in the loop simulation for machine vision at UniBw München

7). This image is projected onto a cylindrical screen with 2.5m radius (bottom right) by a video projector.

In front of this screen on the cylinder axis, the sensors are mounted on a three axis angular motion simulator which is servo-controlled according to the angular rate and position of the simulated process: Sensors are the TV-cameras on the viewing direction control pan and tilt platform, and a set of angular rate and linear acceleration sensors. These and the parallel processors for data evaluation are exactly those to be used in the real vehicle. Therefore, the entire sensing and guidance hard- and software from signal pick-up till control output on the information level, can be tested in this loop. The control output determined, instead of being power-amplified and applied to the real world, is fed back into the simulation processor, and the loop is closed.

Just by adapting software in the system, the application area may be changed from road vehicle to aircraft guidance (or to guidance of ships as well as space vehicles). The very important advantage beside a nice indoor environment is that the actual state to be recovered, the so called 'ground truth' is not only available (at almost no cost) but that it is so in the best possible form for the purpose at hand, namely as data in the simulation computer. In field experiments, this reference is usually very expensive to collect, especially when strong perturbations are present.

A very high percentage of development work for our real world systems has been performed using this simulation facility with the corresponding simulation models of the real systems. Some results will be discussed below after the models have been explained.

8.2. Road Vehicle Guidance

In the framework of the German information technology program since 1982 and of the European EUREKA-project PROMETHEUS since 1988, this application has been very much advanced over the last half decade in cooperation with our industrial partner Daimler-Benz AG. Three areas are of utmost importance for safe vehicle guidance: 1. Recognition of road geometry in a larger look-ahead range, 2. determination of the own position and orientation relative to the road, and 3. detection of obstacles and estimation of the own relative position and velocity.

Tasks 1 and 2 can only be solved in conjunction since the road is being viewed from the vehicle while driving along it. A hybrid representation of the road using both differential

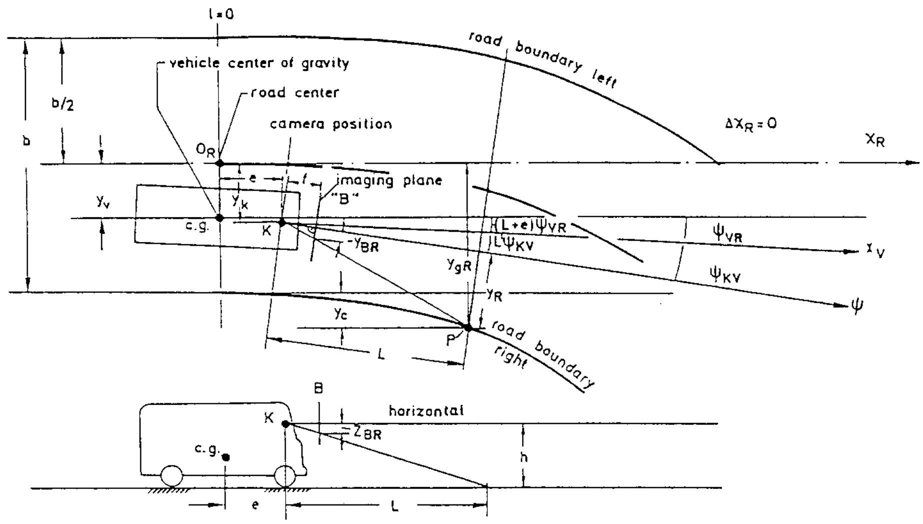

Fig. 8. Vehicle/camera/road-geometry

geometry (for estimation and control purposes) and Cartesian integrals (for perspective projection and the measurement aspects) has proven to be most efficient. For details see [14] and [4]. Fig. 8 shows the basic relations.

Modern high speed roads are built according to so called clothoid models where curvature C(C = 1/R, R = radius of curvature) changes linearly with arc length; this results in constant steer angle rates for turning into a curve at constant speed V. The dynamical models for the road and the own vehicle are (with a = axle distance (3.5m), λ = steer angle, β = lateral slip angle, y_V = lateral position on road (lane), ψ_V = vehicle heading angle, V = vehicle velocity, k_r = lateral tire force coefficient (150 kN/rad), k_λ = steer rate gain factor, u_λ = steer rate control, L = look-ahead range, m = vehicle mass (4000 kg), $p = k_r/(m \cdot V)$, C_{0hm} = local horizontal road curvature of the average model, C_{1hm} = horizontal curvature rate, C_{0vm} = local vertical road curvature, C_{1vm} = vertical curvature rate, n_{c1h} = driving noise for horizontal curvature estimation, n_{c1v} = driving noise for vertical curvature estimation):

lateral vehicle dynamics

$$\begin{bmatrix} \dot\lambda \\ \dot\beta \\ \dot y_V \\ \dot\psi_V \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ V/a-p & -2p & 0 & 0 \\ 0 & V & 0 & V \\ V/a & 0 & 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} \lambda \\ \beta \\ y_V \\ \psi_V \end{bmatrix} + \begin{bmatrix} k_\lambda \\ 0 \\ 0 \\ 0 \end{bmatrix} \cdot u_\lambda + \begin{bmatrix} 0 \\ 0 \\ 0 \\ -V \end{bmatrix} \cdot C_{0hm}$$

road curvature horizontal

$$\begin{bmatrix} \dot C_{0hm} \\ \dot C_{1hm} \\ \dot C_{1h} \end{bmatrix} = \begin{bmatrix} 0 & V & 0 \\ 0 & -3V/L & 3V/L \\ 0 & 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} C_{0hm} \\ C_{1hm} \\ C_{1h} \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ n_{c1h} \end{bmatrix}$$

vertical

$$\begin{bmatrix} \dot C_{0vm} \\ \dot C_{1vm} \end{bmatrix} = \begin{bmatrix} 0 & V \\ 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} C_{0vm} \\ C_{1vm} \end{bmatrix} + \begin{bmatrix} 0 \\ n_{c1v} \end{bmatrix}$$

Since the measurement equations for the general case with vertical curvature are rather involved, the interested reader is referred to [4]. The integral relationship exploited in horizontal curvature measurement is the following: The clothoide model with a linear change of curvature C over arc length l may be written as

$$C(l) = C_0 + C_1 \cdot l.$$

The first integral yields the heading direction χ of the road

$$\chi(l) = \chi_0 + C_0 \cdot l + C_1 \cdot l^2/2.$$

Assuming $\chi - \chi_0$ to be small over the look-ahead range, the cosine of this expression is approximated by 1 and the sine by its argument; then the second integral yields for the lateral offset of the road boundary at the look-ahead distance L due to curvature y_c relative to the tangent direction at the vehicle location ($y_{c0} = 0$)

$$y_c = C_0 \cdot L^2/2 + C_1 \cdot L^3/6.$$

From fig. 8 there follows as measurement equation

$$\pm \frac{b}{2} + y_c - y_V = y_R + \psi_{VR} \cdot (L+e) + \psi_{KV} \cdot L;$$

perspective mapping of y_R as y_{BR} into the image plane at focal distance f yields:

$$y_{BR} = \pm \frac{fb}{L2} + \left[C_0 \cdot L/2 + C_1 \cdot L^2/6 - \psi_{VR} (1 + e/L) - \psi_{KV} - y_V/L \right]$$

8.3. Obstacle Avoidance

The 4D approach, integrating measurement results over time and substituting knowledge about motion processes in 3D-space for determining state variables not directly measured, has the inherent property of motion stereo. This is exploited for monocular range estimation relative to obstacles. Real-time performance has been achieved by assigning this task to an additional group of parallel processors consisting of a set of special feature extraction processors (16 bit word length) and one 4D-object processor (32 bit).

In the search mode for obstacle detection, the feature extraction processors look for combinations of features above the road surface indicative of a larger obstacle; as soon as some candidates are found and remain stable over time, this group is compared to some generic model; during this step, false alarms are eliminated as far as possible. With these feature data a 4D recognition process is then started. A typical simple feature arrangement is shown in fig. 9.

For road vehicle recognition, pairs of features center each other crosswise in the horizontal and in the vertical direction; symmetry conditions may be imposed for feature acceptance if the aspect conditions are from the back or front. Since the lateral position of the obstacle relative to the lane is of importance for the own reaction, the lane boundaries at the location where the object touches the ground are also measured whenever possible.

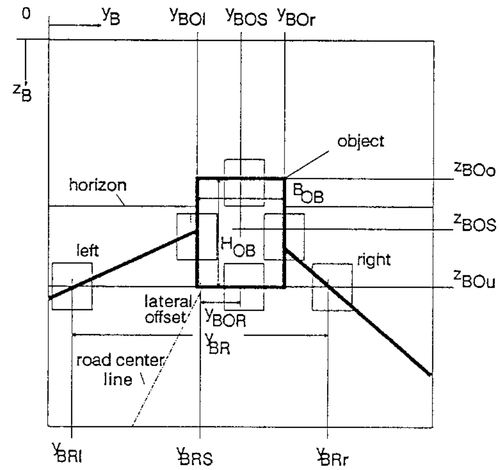

Fig. 9. Feature based obstacle recognition, image geometry

For other vehicles, simple second order dynamical models for decoupled longitudinal (radial) and lateral motion are adopted (for details see [15]). With r = range to the other object, V_O = object speed, y_O = lateral position relative to lane center of lane, V_O = lateral speed of object, n_i = driving noise terms, B = width and H = height of object, the dynamical model is

$$\begin{aligned} \dot{y}_O &= v_{yO} & \dot{r} &= V_O - V \\ \dot{v}_{yO} &= n_{lat}(t) & \dot{V}_O &= n_{long}(t) \\ \dot{B} &= 0 & \dot{H} &= 0 \end{aligned}$$

The measured quantities in the image plain may be seen from fig. 9. Specific estimation results have been discussed in [15, 16].

With a tele-lens of 25 mm focal length, passenger cars can typically be detected at distances of about 100m; at distances of around 50m, range estimation by passive monocular vision typically is around 5 percent accurate. This is considered sufficient for longitudinal control at moderate speeds of around 50 to 80 km/h if the other object moves smoothly. For stationary obstacles and oneself driving at high speed, the performance has to be improved by using larger focal lengths.

9. Experimental Results

9.1. Road Vehicles

The approach described above has matured during half a decade of experimentation with three autonomous vehicles: 1. **VaMoRs,** the experimental vehicle of UniBwM for autonomous mobility and machine vision, a 5-ton van in operation since 1986; 2. since 1988 a 10-ton bus of Daimler-Benz, our industry partner in the framework of the German information technology program for 'Autonomous Mobile Systems', and 3. **VITA,** an autonomous 6-ton van of the Daimler-Benz AG in the EUREKA project PROMETHEUS since 1991. Different sets of computers have been used over the years; in [3] the development of the custom made image sequence processing systems BVV_i has been sketched. Recent developments favor Transputer hardware for easy expandability even before video-busses will become standard in the future.

Always inexpensive PC-type computers have been used for the higher levels: initially, one PC based on the Intel 80286 microprocessor in addition to the BVV_2 with a few 8086 single board computers sufficed for guiding VaMoRs at its maximum speed of 96 km/h on an empty Autobahn in 1987 exploiting the 4D approach. Only through the powerful and intelligent interpretation constraints introduced by the integrated spatio-temporal models has it been possible to achieve these results with that low computing power on board. Since 1989 the Intel 80386 single board computer has been introduced on an intermediate hierarchical level in the BVV_2 [17] resulting in much more robust road recognition under heavily perturbed environmental conditions through shadows from trees; here, the newly introduced 'Gestalt' idea of a perspectively mapped curved road was essential for achieving the performance level demonstrated.

At this stage, also the module for obstacle recognition has been introduced forming a second processor group as discussed above (see [15]).

In 1991 all application software developed up to that point in different computer languages was translated into C and ported onto transputers. In a transition phase now, both BVV_3 and transputers are used jointly; with the availability of transputer processors and corresponding software the BVV will disappear.

The performance level achieved and demonstrated at the PROMETHEUS-display at Torino in September 1991 encompasses the following capabilities of the autonomous vehicles:

- lane keeping on roads with ample shadows from trees under hard sunshine conditions; speed adjustment to road curvature in order not to exceed a preset lateral acceleration limit (for example 0.2 Earth-g)
- driving at night with normal headlights switched on
- detection of obstacles at distances up to about 100m; monocular distance estimation through motion stereo with sufficient accuracy up to about 50m
- stopping in front of an obstacle at a preset distance from speeds up to 50 km/h
- convoy driving behind a vehicle at a speed-dependent distance (2 seconds rule)
- 'stop-and-go' maneuvers behind a preceding vehicle
- lane changes to the left and right triggered by the human operator who has to take care for other vehicles in the neighboring lanes.

With the increased computing power available on board now, work is under progress for coping simultaneously with several other objects and for recognizing object classes from more detailed spatial shape understanding under changing aspect conditions.

The same method has also been applied successfully to the problem of landmark navigation for indoor vehicles on the factor floor (AGV's) [18]. A special kind of landmark navigation in six degrees of freedom will be discussed next.

9.2. Aircraft Landing Approach

One of the most crucial maneuvers in autonomous flight is the final approach phase to the landing strip. Under good visual conditions, human pilots are able to land an aircraft safely without any support from the ground by using just visual cues from the airport environment and the runway. In 1982 we started studying this problem in the simulation loop with the goal to develop methods which would allow autonomous unmanned aircraft with the capability of machine

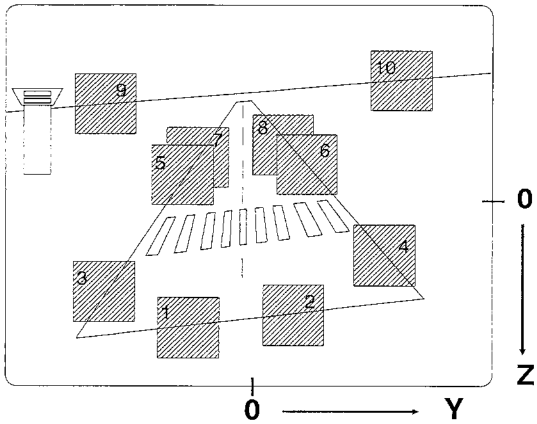

Fig. 10. Simulated landing approach with subareas evaluated for information extraction

vision to do the same. G. Eberl in his dissertation work [19] laid the foundation for the solution available now. From 1987 onward, R. Schell continued the development till the first flight experiments successfully performed in 1991.

The initial nine years of development have been performed in the simulation loop discussed above, exclusively. Results have been published in [20, 21]. Over the years, realism in simulation and the use of real image processing hardware has been steadily increased. Space does not allow to describe the system developed in detail; the interested reader is referred to [22] and forthcoming publications in English.

The achievements may be considered a breakthrough in machine vision application. It has been shown that full spatial motion in all rotatory and translatory degrees of freedom can be controlled by onboard autonomous dynamic machine vision with a relatively small set of today's micro-processors, using the 4D approach. In simulation, the control loop has been closed and landing approaches have been performed from about 1.5 km distance till touchdown, including wind effects and gusts. Fig. 10 shows a simulated approach situation with the hashed squares indicating the image areas evaluated for information extraction. In both the simulation loop and in the real flight experiments the camera was suspended on a two-axis pan-and-tilt platform for visual runway fixation.

In the flight experiments, funded by the German Science Foundation (DFG) and performed with the twin turboprop aircrft Dornier Do-128 of the University of Braunschweig (see fig. 11), inertial angular rates and orientations have been measured by gyros and were fed into the interpretation system, with data fusion performed through the two sixth order dynamical models

Fig. 11. Test aircraft Do-128 of TU-BS

separated for the longitudinal and lateral degrees of freedom.

Since the aircraft was not yet certified for active computer control, only the real-time state estimation part exploiting dynamic vision could be tested. This, however, has been surprisingly successful; after only one week of installation work and interface testing, due to the careful preparations performed in the simulation loop with the complete vision system, first trajectory and state estimation results could be achieved.

Fig. 12 shows the visually estimated altitude as compared to a radio-altimeter measurements and those from the Differential Global Positioning System (D-GPS). The landing approaches were abandoned at about 5m altitude in order to make a fly-around for the next trial. It can be seen that visually estimated altitude and radio-altimeter measurements agree very well in the vicinity of the runway (time > 13 sec); aircraft speed was about 55 m/s (~200 km/h). Estimation quality of the longitudinal position was considered sufficiently good whereas lateral position estimation fluctuated with about 2m amplitude relative to the D-GPS-results; this will have to be studied further but seems to be due to delays in data processing and viewing direction control.

10. Conclusions

The 4D approach to machine vision has been developed with the goal in mind to achieve dynamic vision performance similar to the human one, at least in motion control. Introducing time as an independent variable right from the beginning as the basis for integral spatio-temporal object models, allows to develop very efficient data processing schemes. Unlimited image sequences may be processed without the need for storing previous images; the effects of temporal development of the process under consideration are accumulated in the state of physical objects, internally represented in 3D space and time.

It has been shown in several application areas, that microprocessors available today, already allow surprising performance levels when exploiting this method as compared to quasi-steady approaches usually studied in Artificial Intelligence. For high level performance in complex scenes, these engineering-based methods need to be complemented with ones well suited for explicit knowledge representation and decision making. First steps in this direction have been proposed in the architecture discussed.

Even though our systems developed up to now always have a human operator on board, the principles applied are easily adapted to fully autonomous unmanned systems. Designing the system according to human ways of thinking including action sequences over time, makes the interface to humans for interaction or for monitoring relatively simple.

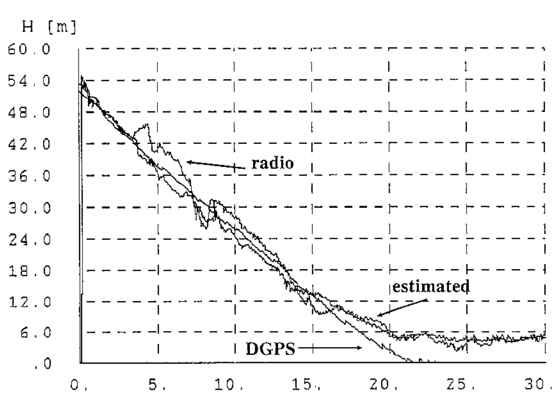

Fig. 12. Estimated altitude time history

Acknowledgments

This research has been funded by the German Federal Ministry of Research and Technology (BMFT), Daimler-Benz AG, Stuttgart and the German Science Foundation DFG.

References

1. R.E. Kalman, "A new approach to linear filtering and prediction problems," *Trans. ASME, Series D, Journal of Basic Engineering*, pp. 35–45, 1960.
2. H.J. Wuensche, "Bewegungssteuerung durch rechnersehen," *Fachberichte Messen*, Steuern, Regeln, No. 20, Springer-Verlag, Berlin, 1988 (Motion control by computer vision).
3. E.D. Dickmanns and V. Graefe, a) "Dynamic monocular machine vision," b) "Application of dynamic mon-

ocular machine vision," *J. Machine Vision Application*, pp. 223–261, 1988.
4. E.D. Dickmanns, and B. Mysliwetz, "Recursive 3D road and relative ego-state recognition," *IEEE-Trans. PAMI, Special Issue on 'Interpretation of 3D Scenes,'* Feb. 1992.
5. G.J. Bierman, "Measurement updating using the U-D factorization," in *Proc. IEEE Control and Decision Conf.*, Houston, TX, 1975, pp. 337–346.
6. B. Mysliwetz, "Parallelrechner-basierte Bildfolgeninterpretation zur autonomen Fahrzeugsteuerung," Dissertation, Fakultät für Luft- und Raumfahrttechnik der Universität der Bundeswehr München, 1990 (Image sequence interpretation for autonomous guidance of vehicles based on parallel processors).
7. G.J. Bierman, *Factorization Methods for Discrete Sequential Estimation*, Acad. Press: New York, 1977.
8. P.S. Maybeck, *Stochastic Models, Estimation and Control*, Vol. 1, Acad. Press: New York, 1979.
9. K. Kuhnert, "Zur Echtzeit-Bildfolgenanalyse mit Vorwissen," Dissertation, Fakultät für Luft-und Raumfahrttechnik der Universität der Bundeswehr München, 1988 (Real-time image sequence analysis using knowledge).
10. E.D. Dickmanns, "Subject-object discrimination in 4D- dynamic scene interpretation for machine vision," *Proceedings IEEE-Workshop on Visual Motion*, Newport Beach, 1989, pp. 298–304.
11. J. Schick and E.D. Dickmanns, "Simultaneous estimation of 3D shape and motion of objects by computer vision," *IEEE-Second Workshop on Visual Motion*, Princeton, 1991.
12. E.D. Dickmanns, "4D dynamic vision for intelligent motion control," *Int. Journal for Engineering Applications of AI (IJEAAI)*, Vol. 4, No. 4, pp. 301–307, 1991.
13. S.S. Iyengar and A. Elfes, "*Autonomous Mobile Robots: Control, Planning, and Architecture*," IEEE Comp. Soc. Press: Los Alamitos, 1991.
14. E.D. Dickmanns and A. Zapp, "A curvature-based scheme for improving road vehicle guidance by computer vision," in *Mobile Robots*, SPIE-Proc. Vol. 727, Cambridge, MA, 1986, pp. 161–168.
15. E.D. Dickmanns, and T. Christians, "Relative 3D-state estimation for autonomous visual guidance of road vehicles," in *Intelligent Autonomous Systems* 2, edited by T. Kanade, Amsterdam, pp. 683–693, 1989.
16. E.D. Dickmanns, B. Mysliwetz, and T. Christians, "Spatio-temporal guidance of autonomous vehicles by computer vision," *IEEE-Trans.on Systems, Man and Cybernetics*, Vol. 20, No. 6, (Special Issue on Unmanned Vehicles and Intelligent Robotic Systems), pp. 1273–1284, 1990.
17. B. Mysliwetz and E.D. Dickmanns, "Distributed scene analysis for autonomous road vehicle guidance," *Proc. SPIE Conf. on Mobile Robots*, Vol. 852, Cambridge, MA, 1987, pp. 72–79.
18. C. Hock, "ATHENE, ein Projekt zur Landmarkennavigation," *6. Fachgespräch Autonome Mobile Systeme*, Karlsruhe, 1990, (Athene, a project for landmark navigation).
19. G. Eberl, "Automatischer Landeanflug durch Rechnersehen," Dissertation, Fakultät für Luft- und Raumfahrttechnik der Universität der Bundeswehr München, 1987, (Automatic landing approach by computer vision).
20. E.D. Dickmanns, "Computer vision for flight vehicles," Zeitschrift für Flugwissenschaft und Weltraumforschung (ZFW), 1988.
21. R. Schell and E.D. Dickmanns, "Autonomous automatic landing through computer vision," in *AGARD Conf. Proc. No. CP-455: 'Advances in Techniques and Technologies for Air Vehicle Navigation and Guidance,'* 1989, pp. 24.1–24.9.
22. R. Schell, "Bordautonomer automatischer Landeanflug aufgrund bildhafter und inertialer Meβdatenauswertung," Dissertation, UniBw München, Fakultät LRT, 1992, (On-board autonomous automatic landing approach based on visual and inertial data evaluation).
23. B. Mysliwetz and E.D. Dickmanns, "A vision system with active gaze control for real-time interpretation of well structured scenes," in *Proc. of 1-st Conference on Intelligent Autonomous Systems (IAS)*, edited by L.O. Hertzberger, Amsterdam, 1986, pp. 477–483.
24. E.D. Dickmanns, "Simulation for the development of a visual autopilot-system for road vehicles," in *Automotive Simulation*, edited by M.R. Heller Springer-Verlag, Berlin, 1989, pp. 11–22.

Ernst D. Dickmanns studied Aerospace Engineering at the Techn. University, Aachen, Germany from 1956 to 1961, and control engineering at Princeton University, N.J., from 1964 to 1965. He has been with the German Aerospace Research organization DFVLR from 1961 to 1975 working in optimal control. Since 1975 he has been a full professor for control engineering at the Universitat der Bundeswehr, Munich, Department of Aerospace Technology. His main research area is real-time machine vision for intelligent motion control.