

Dynamic Vision for Precise Depth Measurement and Robot Control

Volker Graefe, Klaus Peter Wershofen, Johannes Huber

Institut für Meßtechnik
Universität der Bundeswehr München
8014 Neubiberg, Germany

Phone: (+49 89) 6004-3590, -3587
Fax: (+49 89) 6004-3074

Abstract

Dynamic vision is a powerful and effective sensing modality for a broad variety of numerous industrial applications. Two of them will be introduced as examples. One, relates to a novel implementation of *motion stereo* allowing precise distance measurements by a moving robot in real time with an uncalibrated camera. The other, *behavior-based navigation*, allows a robot to accomplish a desired navigation task by activation of an appropriate sequence of system-immanent behavior patterns.

Introduction

Versatility is probably the most striking merit of machine vision as a sensing modality for tomorrow's intelligent systems, be they robots, vehicles, or other machines. A vision system may allow a robot to comprehend a complex situation in real time; the same vision system may also enable it to measure quantitatively and with great accuracy certain geometrical characteristics of its environment.

Two industrial applications of robot vision will be introduced as examples.

The first one is a novel form of *motion stereo* that has great practical advantages. A mobile robot, or a manipulator, carrying a single TV-camera may use this method to determine its distance to external objects accurately and in real time. While the robot approaches an object the accuracy of the measurement improves continuously as the distance decreases, making the method particularly suitable for tasks like collision avoidance, docking, or grasping. The method may be implemented without requiring a camera calibration, provided the lens of the camera is sufficiently free of distortion. Accuracies better than 1 % of the distance have been demonstrated in real-world environments.

The second application is a method for robot control and navigation of mobile robots that relies heavily on a powerful sensing system, such as vision. By this approach, called *behavior-based navigation*, the desired navigation task is accomplished by activating an appropriate sequence of system-immanent behavior patterns. The underlying system concept and the object-oriented vision concept are discussed. A mobile robot operating in a normal office building or similar environment is taken as example. Experiments and results are reported.

Motion Stereo

Principle

Advantages of the approach

Motion stereo has great practical advantages as a means for a mobile robot to determine its precise distance from external objects. A novel approach to motion stereo that is especially suitable for mobile robots has been introduced by the authors in previous papers. It is based on sampling a dynamic scene rapidly (e.g. 25 or 50 images per second) and measuring the minute motion-induced displacements of features relative to each other in the image sequence.

A mobile robot carrying a single TV camera may use this method to measure its distance from objects in front of it accurately and in real time. Mobile robots usually move with a known speed through an environment with many resting objects. If such a robot carries a camera the motion-stereo-based method is particularly adequate since no extra steps have to be taken for providing the required relative motion between the camera and the external objects.

Accuracy is another advantage of the proposed method. Very shortly after detecting an object (possibly an obstacle) and starting the measurement, the robot may already determine its distance from the object with an error that is in typical cases less than $\pm 10\%$ of the true distance. While the robot continues to approach the object the accuracy of the measurement improves rapidly as the distance decreases, making the method particularly suitable for tasks like collision avoidance or docking. Errors below 1% of the true distance are achieved after approaching the object by only a small fraction of the initial distance.

Such accuracies were first verified using a mobile robot in an indoor-environment. Later it was demonstrated that similar results may also be obtained in outdoor-environments, despite of the much greater magnitude of uncontrollable camera motion.

In these outdoor-experiments the camera was mounted inside a truck, and the distance to an object being approached at a speed of about 30 km/h was continuously measured. The error was less than 50 cm at distances of 50 to 60 m.

These accuracies are achieved with a completely uncalibrated camera. Considering the high cost and low stability of camera calibrations, this is a significant advantage of the proposed motion stereo method, especially when comparing it to the well-known multi-camera parallax stereo approach. There, the entire camera system must be precisely calibrated to acquire accurate results.

Theoretical background

The basic concepts of the motion stereo method as well as the theory behind it, and some experimental results, were introduced in previous papers [Graefe 90], [Huber, Graefe 90] and [Huber, Graefe 91].

For summarizing the theoretical background let us assume that two distinct points are visible in the image of a scene, corresponding to two fixed points on the surface of an object in the scene, and that the two object points are equally distant from the image plane of the camera. Let b be the distance between the two points in the image, and Z the unknown distance between the object points and the camera center (measured parallel to the optical axis). In the papers mentioned it was shown that, if the relative motion parallel to the optical axis is known, the distance between the external object and the camera may be computed from b and db/dt or db/dZ :

$$Z = \frac{-b \cdot dz/dt}{db/dt} = \frac{-b}{db/dz} . \quad (1)$$

Note that for computing Z according to this equation no knowledge is needed regarding the internal camera parameters and the camera's velocity components parallel to the image plane.

The time t_0 , remaining until the object will pass through the plane $Z = 0$, may be computed similarly:

$$t_0 = \frac{Z}{dz/dt} = \frac{b}{db/dt} ; \quad (2)$$

t_0 has been called "time-to-collision" by [Lee 76]. It may be a useful quantity if only the speed of the vehicle carrying the camera is known, but not the speed of the object. In this case it is impossible to use equation (1) for computing the distance to the object, but knowing t_0 , at least the distance to the point of "collision" may then be determined.

Measurement Errors

One difficulty with the proposed method is that the motion of features in an image relative to each other, i. e. db/dt , is usually not easy to measure, due to its small magnitude. To solve this problem the scene is sampled at a high rate (25 or 50 images per second), features are localized in each image with subpixel resolution, and the resulting data are smoothed with a recursive filter to reduce measurement noise.

The measurement accuracy improves quickly while approaching an initially distant object, even if it is low at the beginning. The absolute error decreases by the third power of the distance, and even faster, if the increasing number of measured values is fully utilized. This error behavior makes this method particularly suitable for mobile systems. (For a detailed discussion of errors see [Huber, Graefe 90]).

In many situations it is easy to find the necessary two points on an object that are equally distant from the camera. But there are exceptions; if the two points being observed are not equally distant from the image plane the value for Z computed according to equation (1) is not the true distance to either one of the points. The error caused by violating the assumption of equal distance depends in a complicated way on the locations of the two points; if the points are on opposite sides of the optical axis the method yields an intermediate value for the distance.

Geometric distortion of the lens is a source of additional systematic error. It impairs the measurements, especially if the image of an object is large, unless the distortion is known and the measured value for b is corrected accordingly prior to computing the distance.

Feature Selection

In earlier experiments it was assumed that two parallel straight edges of the object, whose distance was to be determined, were visible in the image. The distance between those edges was then used for b in equations (1) or (2).

Using edges, instead of discrete points, as features has both advantages and disadvantages. The advantages are that edge detectors are normally more robust against noise, and edges are easier to locate in images; a one-dimensional search path is sufficient. On the other hand, not all objects have visible edges known to be parallel to each other; and points yield, in principle, more information than edge elements (two coordinates instead of one).

Corners are among the image primitives that define discrete points. As an example for the applicability of point features in real-world images, the suitability of a specific corner detector proposed by [Kuhnert 88] was investigated.

For locating a corner in an image a prototype corner is correlated with a section of the image (search area). Figure 1 shows examples of such prototypes, or "masks". These particular masks match primarily corners composed of one horizontal and one vertical edge.

The masks shown in Figure 1 are rather small (only $4 \cdot 4$ pixels); therefore, they are rather tolerant with respect to the directions of the two edges forming the corner. If highly selective corner detectors are desired they may easily be designed in a similar way by using large masks closely resembling the shapes of the desired corners.

For computing the correlation function the selected mask is shifted relative to the image so that its reference point (indicated by a dot in each mask in Figure 1) coincides in turn with each pixel in the search area. A local maximum, or minimum, of the thus determined correlation function indicates the location of a corner point with a quantization error of up to $\pm 1/2$ pixel in each direction.

To reduce this error, a second-order polynomial is fitted to the correlation values in the vicinity of the extremum. The location of the local maximum, or minimum, of the polynomial is then calculated with subpixel resolution and taken as the precise location of the corner point.

It is not *a priori* clear that this interpolation procedure really yields the true location of a corner in an image. In fact, it is not even quite clear how the true location of a "corner" should be defined in a real image that is necessarily not quite sharp and has, in addition, been quantized. Therefore, experiments were performed to test the consistency and applicability of the approach.

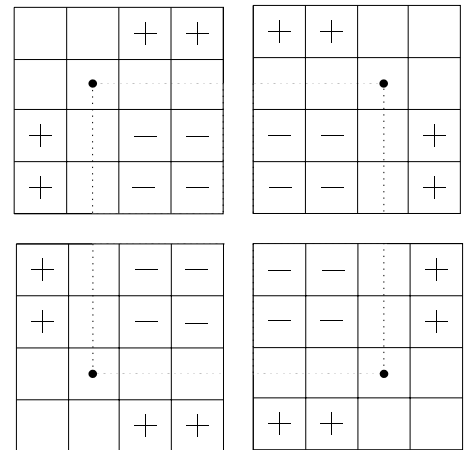


Figure 1
Correlation masks for detecting four types of corners; + stands for +1, - for -1. A dot marks the reference point of each mask, the shading indicates the type of corner for which each one of the masks is suitable.



Figure 2
An image recorded by the vision system in one of the outdoor-experiments

Outdoor-experiments

The actual performance of the motion stereo approach in outdoor scenes was investigated in experiments. (Experimental results obtained with a mobile indoor-robot were reported earlier; accuracies better than 1 % of the true distance were obtained.) Compared to indoor-experiments, conditions tend to be less controllable outdoors, especially regarding lighting and random motions of the camera.

In the outdoor-experiments the camera and the vision system were mounted in an experimental vehicle "VaMoRs" (5-ton van equipped with various sensors, computers, a power generator etc. [Dickmanns, Graefe 88]). An object, for instance a barrel about 1 m high and 50 cm in diameter, was placed on the road, and the distance from it was measured while the vehicle approached it. Figure 2 shows the scene as seen by the camera of the vision system. The size of the object, the width of the road, and all internal and external camera parameters were unknown to the system. In the experiments the initial distance was between 40 and 80 m, and the vehicle was driven at speeds of 20 to 30 km/h.

Figure 3 shows results of such an experiment. The initial distance was 75 m, and the vehicle was stopped about 9 m away from the object. The initial width b of the object's image was about 8 pixels. The calculated distance is plotted in the upper part of the figure, and the error in the lower part. Very briefly after the start of the measurement (after 7.5 m or 10 % of the initial distance) the error drops to less than 1.5 % of the true distance; after that it decreases even further. Vertical lines in Figure 3 mark those distances where b has increased by 1, 2, and 3 pixels, respectively, compared to the start of the measurement.

In a similar experiment corners were used as features (Figure 4). When seen from a long distance the image of the barrel seemingly has a rectangular shape. The lower left and right corners may thus be used for measuring the distance. But as the barrel is approached its image becomes larger and larger, and it becomes more and more obvious that actually there are no lower left and right corners in the image. Rather, the left and right sides of the image of the barrel are connected to each other smoothly by an elliptic arc.

Nevertheless, when correlating corner masks (as in Figure 1) with the image the resulting correlation functions still have maxima somewhere; these maxima were used as features for calculating distances in experiments, although they do not correspond to any specific points on the surface of the barrel.

Figure 4 shows the result of such an experiment. The accuracy obtained will be sufficient for many practical applications, but it is

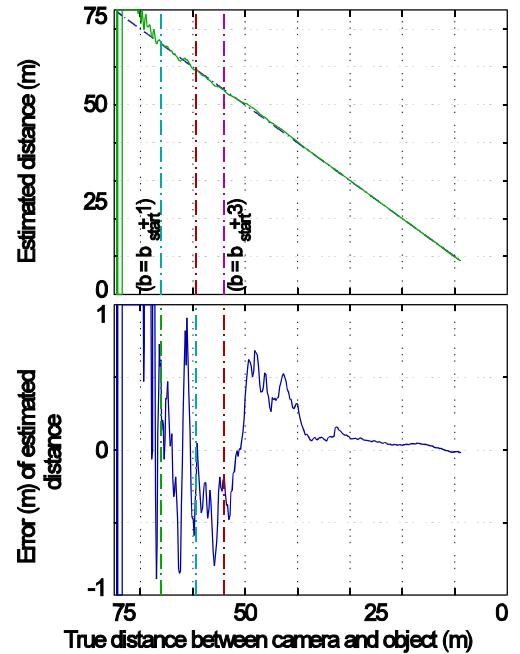


Figure 3 Distance calculated according to the motion stereo method using edge elements as features (upper graph); error of the calculated distance (lower graph; different scale for improving the visibility of the error)

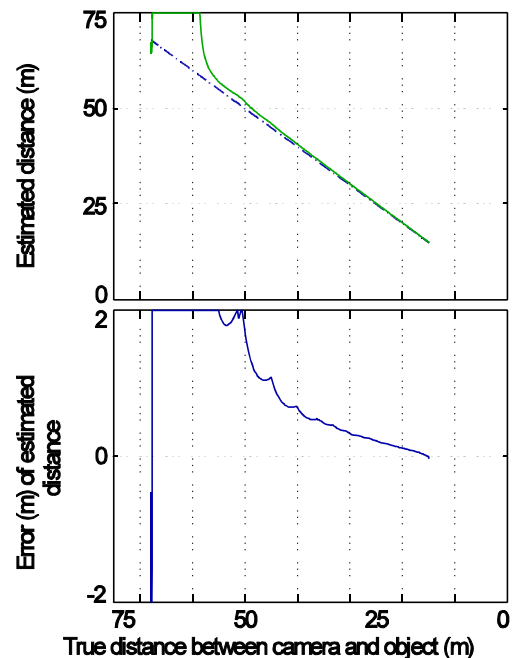


Figure 4 Distance measurement based on corners otherwise similar to Figure 3

not as good as when basing the measurement on edge elements. The relative error is 4 % at a distance of 50 m, and it drops to 1 % only when the distance is reduced to 30 m.

The initial setting is much slower than in the case of Figure 3; this could possibly be corrected by adjusting the parameters of the smoothing filters. But also the subsequent decrease of the error is much less pronounced, which is to be expected, when considering the increasingly noticeable incorrectness of the underlying corner model as explained above.

Mobile Robot Navigation

In addition to being a means for measuring distances precisely, vision may also be used for perceiving the environment as a whole in a more qualitative way. Autonomous navigation of robots in complex environments like factories or warehouses is an example of great practical interest.

Behavior-based Navigation

A framework for mobile robot navigation is provided by the behavior-based navigation, which was introduced in [Wershofen, Graefe 92]. The main principle of behavior-based navigation is the achievement of a desired navigation task by activating an appropriate sequence of elementary behavior patterns in a situation-dependent way. The combination and selection of behavior patterns is performed active and knowledge-based by the mobile robot itself, while it is operating. Suitable behavior patterns for a mobile robot navigating in an office building are for example "going along a corridor", "turning around at an intersection", or "moving toward a landmark".

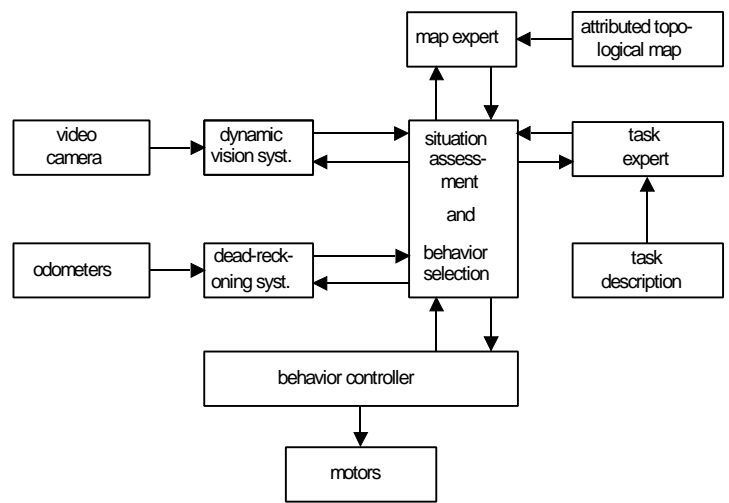


Figure 5:
Behavior-based navigation

Figure 5 shows the principle of behavior-based navigation. At the core of the system is a module called situation assessment and behavior selection.

This module generates a description of the current situation by fusing information provided by the sensor system (dynamic vision system, dead-reckoning system) and information about the static characteristics of the environment (stored in the attributed topological map). Depending on the prevailing situation and the current task, the behavior pattern to be executed next is determined. It is selected from a repertoire and commanded to the behavior controller. The behavior controller produces appropriate control signals for the motors by using interpreted sensor signals.

Object-oriented Vision

Vision with its ability to perceive many aspects of an environment is the ideal sensing modality for behavior-based navigation.

The environment of a mobile robot consists of a variety of individual physical objects. These objects, however, have different significance for the robot. Only some of them, for example pathways, intersections, landmarks, or obstacles are relevant for the proper operation of the robot. So, only those objects must be observed and modeled by the robot. A dynamic vision system, based on the concept of object-oriented vision as introduced in [Graefe 89], is optimally suited for recognizing task-relevant objects.

Figure 6 shows the structure of an object-oriented vision system.

A separate object process corresponds to each object relevant for solving the robot's task. The task of an object process is to recognize a related object and to generate a continuously updated description of it. The object description includes a classification of the object and information about the spatial relationship (distance, orientation) of the object relative to the robot. The object description is generated by first fusing pixels (as provided by the video section) to elementary features; then those features are fused into objects by using 2-D models resembling the visual appearance of objects. The spatial relationship may then be computed by applying the motion stereo method described in the previous section of this paper.

Each object process may be adjusted depending on the situation and the task to be solved. This is managed by the situation assessment; it starts and terminates object processes depending on the prevailing situation and the resources of the system. (For a detailed discussion of specific object processes see [Albert, Meier 92].)

Implementation of the Behavior-based Navigation

The mobile robot ATHENE [Kuhnert 90] is a test-bed for experiments with the behavior-based navigation. ATHENE is a 3-wheeled vehicle, about 1.5 m long and 0.7 m wide, propelled and steered by the front wheel. Main sensors are a CCD-camera mounted on a platform, and shaft encoders on both rear wheels.

A heterogenous multiprocessor-system serves as hardware-platform for the behavior-based navigation. It consists of a control computer for the undercarriage, a so-called manager computer, and a real-time vision system. The control computer (INTEL 80188) for the undercarriage reads the sensor signals (encoders, bumpers, etc.) and controls the motors. The interface to the human operator, the situation assessment module, the behavior controller, and the attributed topological map are implemented in the manager computer (PC-AT 386/33 MHz). An i960-BVV [Albert, Meier 92] is used as real-time vision system. This robot vision system is a custom-designed board within the PC, based on the INTEL 80960 CA RISC processor, coupled with a commercial frame grabber card serving as A/D converter and for visualizing the results of image processing.

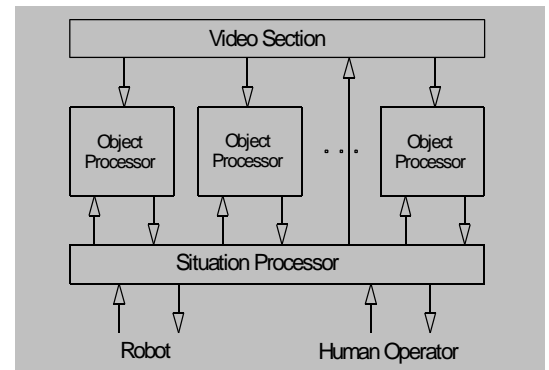


Figure 6:
Object-oriented vision

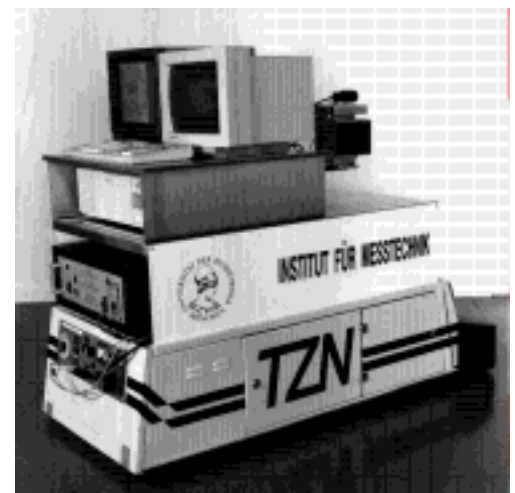


Figure 7:
Mobile robot ATHENE

Experiments

First experiments with the behavior-based navigation were carried out in a network of pathways consisting of hallways (Figure 8). Therefore, two basic behavior patterns were implemented. They are "following a baseboard", and "turning around the corner at the end of a hallway". These behavior patterns are sufficient to enable the mobile robot ATHENE to follow a test course with a speed up to 1 m/s. This test course consists of four hallways with a total length of about 100 m and a width of 1.80 meters. The hallways are linked by 90 degree corners.

Following the tests in the laboratory the vehicle was transferred to a factory environment. Also in this environment ATHENE demonstrated its ability to follow the determined course with great precision at all critical points while using only a small number of landmarks as references. The factory test course included a narrow door with only 4 cm of space on either side of the vehicle. After completing the course of 85 m length, the vehicle returned to its starting position with an error of merely a few centimeters.

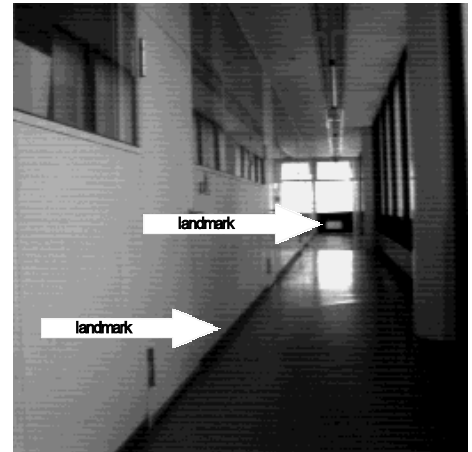


Figure 8:
Part of the laboratory test course

Conclusions

Vision has an extremely high potential as a universal sensing modality for advanced mobile robots. Experimental evidence shows that vision may enable a robot to measure precisely its distance from external objects. The same approach may enable other robots (or the same one) to find their ways in a network of passageways or in a factory environment, and to move with dexterity in crowded quarters.

A dynamic motion stereo method for precise range measurements was implemented and tested experimentally. Errors below 1 % of the true distance were achieved with the method even in uncontrolled outdoor environments and with an uncalibrated camera.

Experiments with a vision-guided autonomous vehicle (mobile robot) showed that vision is an excellent sensing modality for the concept of behavior-based navigation. By equipping the mobile robot with a dynamic vision system based on object-oriented vision the performance of the system, especially the speed of a robot, is mainly limited by the dynamics of the robot itself.

Despite the remarkable results obtained in the reported experiments, extensive research work still lies ahead before vision-based mobile robots will be an everyday occurrence.

References

- Albert, M.; Meier, H. (1992):** Dynamisches Rechnersehen für verhaltensbasierte Landmarkennavigation. 8. Fachgespräch über autonome mobile Systeme. Karlsruhe, pp 15-33.
- Dickmanns, E.D.; Graefe, V. (1988):** Dynamic Monocular Machine Vision. Machine Vision and Applications 1 (1988), pp 223-261.
- Graefe, V. (1989):** Dynamic Vision Systems for Autonomous Mobile Robots. Proceedings, IEEE/RSJ International Workshop on Intelligent Robots and Systems, IROS '89. Tsukuba, pp 12-23.
- Graefe, V. (1990):** Precise Range Measurement by Monocular Stereo Vision. Japan-USA Symposium on Flexible Automation. Kyoto, pp 1321-1324.

Huber, J.; Graefe, V. (1990): Subpixelauflösung und Bewegungsstereo für die räumliche Deutung von Bildfolgen. Fachgespräch Autonome Mobile Systeme. Karlsruhe, pp 173-184.

Huber, J.; Graefe, V. (1991): Quantitative Interpretation of Image Velocities in Real Time. IEEE Workshop on Visual Motion. Princeton, pp 211-216.

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

Kuhnert, K.-D. (1990): Dynamic Vision Guides the Autonomous Vehicle ATHENE. Japan-USA Symposium on Flexible Automation. Kyoto, pp 507-510.

Lee, D. N. (1976): A theory of visual control of braking based on information about time-to-collision. Perception, Vol. 5, pp. 437-459.

Wershofen K.P., Graefe, V. (1992): An Intelligent Autonomous Vehicle Guided by Behavior-based Navigation. IFToMM-jc International Symposium on Theory of Machines and Mechanisms. Nagoya, pp 244-249.