# Machine Vision for Intelligent Robots

**Rainer Bischoff and Volker Graefe***

Institute of Measurement Science
Bundeswehr University Munich

## Abstract

Vision is an ideal sensor modality for intelligent robots. It provides rich information on the environment as required for recognizing objects and understanding situations in real time. Moreover, vision-guided robots may be largely calibration-free, which is a great practical advantage. Three vision-guided robots and their design concepts are introduced: an autonomous indoor vehicle, a calibration-free manipulator arm, and a humanoid service robot with an omnidirectional wheel base and two arms. Results obtained, and insights gained, in real-world experiments with them are presented.

## 1 Introduction

### 1.1 Robots – Today and Tomorrow

Industrial robots are of great economic and technological importance. Until 1996 approximately 860,000 robots had been installed worldwide. At that time 680,000 of them were still being used, for the most part in automobile and metal-manufacturing [IFR 1997]. Typical applications include welding cars, spraying paint on appliances, assembling printed circuit boards, loading and unloading machines, and placing cartons on a pallet. Experts estimate that by the year 2000 about 950.000 industrial robots will be employed world-wide.

Although present robots contribute very much to the prosperity of the industrialized countries they are quite different from the robots that researchers have in mind when they talk about "intelligent robots". Today's robots

- are not creative or innovative,
- do not think independently,
- do not make complicated decisions,
- do not learn from mistakes,
- do not adapt quickly to changes in their surroundings.

They rely on detailed teaching and programming and carefully prepared environments. It is costly to maintain them and it is difficult to adapt their programming to slightly changed environmental conditions or modified tasks.

Although the vast majority of robots today are used in factories, advances in technology are enabling robots to automate many tasks in non-manufacturing industries such as agriculture, construction, health care, retailing and other services. These so-called "field and service robots" aim at the fast growing service sector and promise to be a key product for the next decades.

From a technical point of view service robots are intermediate steps towards a much higher goal: "personal robots" that will be as indispensable and ubiquitous as personal computers today. Personal robots must operate in varying and unstructured environments without needing maintenance or programming. They must cooperate and coexist with humans who are not trained to cooperate with robots and who are not necessarily interested in them. Advanced safety concepts will be as indispensable as intelligent communication abilities, learning capabilities, and reliability. It will be a long way of research to achieve this goal, but undoubtedly vision – the most powerful sensor modality known – will enable these robots to perceive their environments, to understand complex situations and to behave intelligently.

### 1.2 Scope and Outline of this Paper

This paper presents some of the underlying concepts and principles that were key to the design of our research robots. In brief, they are:

- **Vision** is the most powerful sensor modality for providing rich and timely information on a robot's environment.
- **Behavior** is the key to a powerful system architecture that enables a robot to construct complex actions by combining elementary behavior primitives.
- **Situation assessment** is the basis for the dynamic selection of the most appropriate behavior by a robot in its interactions with the outside world.
- **Perception** rather than measurement should be the basis for situation assessment and robot control.

We expect that these fundamental concepts are a strong basis for future generations of intelligent robots that combine locomotive and manipulative actions. In section 2 these concepts are explained in more detail.

Another fundamental principle has considerably influenced our research work: Every result has to be proved and demonstrated in practical experiments and in the real world. While this approach is rather demanding and often

*Address: Werner-Heisenberg-Weg 39, 85577 Neubiberg, Germany
E-Mail: {Rainer.Bischoff | Volker.Graefe}@unibw-muenchen.de
URL: http://www.unibw-muenchen.de/campus/LRT6/welcome.html

frustrating, compared to mere computer simulations it has the great advantage of yielding by far more reliable and valuable results. This will be shown in sections 3-5 where we introduce three of our vision-guided robots: an autonomous indoor vehicle (section 3), a calibration-free manipulator arm (section 4) and a humanoid mobile service robot (section 5).

# 2 Key Concepts for Intelligent Robots of the Future

## 2.1 Vision and its Potential for Robots

When a human drives a vehicle he depends mostly on his eyes for perceiving the environment. He uses his sense of vision not only for locating the path to be traversed and for judging its condition, but also for detecting and classifying external objects, such as other vehicles or obstacles, and for estimating their state of motion. Entire situations may thus be recognized, and expectations, as to their further development in the "foreseeable" future, may be formed.

The same is true for almost all animals. With the exception of those species adapted to living in very dark environments, they use vision as the main sensing modality for controlling their motions. Observing animals, for instance, when they are pursuing prey or trying to escape a predator, may give an impression of the performance of organic vision systems for motion control.

In some modern factory and office buildings mobile robots are operating, but almost all of them are blind. Their sensors are far from adequate for supplying all the information necessary for understanding a situation. Some of them have only magnetic or simple optical sensors, allowing them merely to follow an appropriately marked track. They will fail whenever they encounter an obstacle and they are typically unable to recover from a condition of having lost their track. The lack of adequate sensory information is an important cause making these robots move in a comparatively clumsy way and restricting their operation to the simplest of situations.

Other mobile robots are equipped with sonar systems. Sonar can, in principle, be a basis for powerful sensing systems, as evidenced by certain animals, such as bats or dolphins. But the sonar systems used for mobile robots are usually rather simple ones, their simplicity and low cost being the very reason for choosing sonar as a sensing modality. It is then not surprising that such systems are severely limited in their performance by low resolution, specular reflections, insufficient dynamic range, and other effects.

Nevertheless, even when comparing the most highly developed organic sonar systems with organic vision systems, it is obvious that in all environments where vision is physically possible animals endowed with a sense of vision have, in the course of evolution, prevailed over those that depend on sonar. This may be taken as an indication that vision has, in principle, a greater potential for sensing the environment than sonar. Likewise, it may be expected that advanced robots of the future will also rely primarily on vision for perceiving their environment, unless they are intended to operate in other environments, e.g. under water, where vision is not feasible.

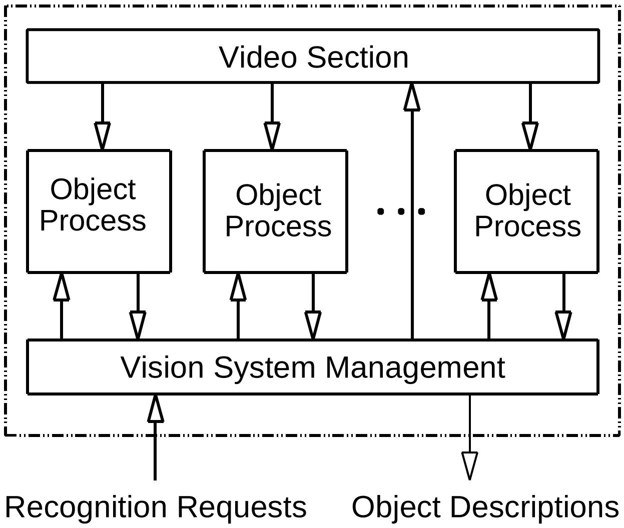


**Figure 1:**
Conceptual structure of object-oriented robot vision systems.

One apparent difficulty in implementing vision as a sensor modality for robots is the huge amount of data generated by a video camera: about 10 million pixels per second, depending on the video system used. Nevertheless, it has been shown (e.g., by [Graefe 1989]) that modest computational resources are sufficient for realizing real-time vision systems if a suitable system architecture is implemented.

As a key idea for the design of efficient robot vision systems the concept of object-oriented vision was proposed. It is based on the observation that both the knowledge representation and the data fusion processes in a vision system may be structured according to the visible and relevant external objects in the environment of the robot (Figure 1). For each object that is relevant for the operation of the robot at a particular moment the system has one separate "object process". An object process receives image data from the video section (cameras, digitizers, video bus etc.) and generates and updates continuously a description of its assigned physical object. This description emerges from a hierarchically structured data fusion process which begins with the extraction of elementary features, such as edges, corners and textures, from the relevant image parts and ends with matching a 2-D model to the group of features, thus identifying the object.

This concept is practical because it was found that in any given moment only a small number of objects are relevant and that, consequently, only a small number of processes need to be active simultaneously. In the next moment, however, different objects may be relevant; therefore, the ability to switch the system's focus of attention quickly is crucial. The switching of attention and the control of the cameras is performed by a vision system management process that dynamically generates appropriate object processes upon request.

The potential of object-oriented vision systems was first demonstrated in high-speed autonomous highway driving applications [Graefe, Kuhnert 1988], [Graefe 1992]. Later the same concept has proved its value in mobile and stationary indoor robots.

## 2.2 Behavior

Biological behaviors could be defined as anything that an organism does involving action and response to stimulation, or as the response of an individual, group, or species to its environment. Behavior-based robotics has become a very popular field in robotics research because biology proves that even the simplest creatures are capable of intelligent behavior: They survive in the real world and compete or cooperate successfully with other beings. Why should it not be possible to endow robots with such an intelligence? By studying animal behavior, particularly their underlying neuroscientific, psychological and ethological concepts, robotic researchers have been enabled to build intelligent behavior-based robots according to the following principles:

- complex behaviors are combinations of simple ones, complex actions emerge from interacting with the real world
- behaviors are selected by arbitration or fusion mechanisms from a repertoire of (competing) behaviors
- behaviors should be tuned to fit the requirements of a particular environment and task
- perception should be actively controlled according to the actual situation

Many system architectures and control methods exist to realize behavior-based robots. As one realization out of this class of architectures we propose the concept of situation-oriented behavior-based control (section 2.3). Its main characteristics are active perception of the robot’s dynamically changing environment, recognition and evaluation of its current situation, and dynamic selection of behaviors appropriate for the actual situation. Animals simplest capabilities, i.e., to perceive and act within an environment in a meaningful and purposive manner, can thus be imitated by our robots to a certain degree.

## 2.3 Situation Assessment

According to the classical approach, robot control is model-based. Numerical models of the kinematics and dynamics of the robot and of the external objects that the robot should interact with, as well as quantitative sensor models, are the basis for controlling the robot’s motions. The main advantage of model-based control is that it lends itself to the application of classical control theory and, thus, may be considered a straight-forward approach. The weak point of the approach is that it breaks down when there is no accurate quantitative agreement between reality and the models. Differences between models and reality may come about easily; an error in one of the many coefficients that are part of the numerical models suffices. Among the many possible causes for discrepancies are initial calibration errors, aging of components, changes of environmental conditions, such as temperature, humidity, electromagnetic fields or illumination, maintenance work and replacement of components, to mention only a few. Consequently, most robots work only in carefully controlled environments and need frequent recalibrations, in addition to a cumbersome and expensive initial calibration.

Organisms, on the other hand, are robust and adapt easily to changes of their own conditions and of the environment. They never need any calibration, and they normally do not know the values of any parameters related to the characteristics of their “sensors” or “actuators”. Obviously, they do not suffer from the shortcomings of model-based control which leads us to the assumption that they use something other than numerical models for controlling their motions. Perhaps their motion control is based on a holistic assessment of situations and the selection of behaviors to be executed on that basis, and perhaps robotics could benefit from following a similar approach.

According to Webster’s Third New International Dictionary [Babcock 1976] the term “situation” describes among others “the way in which something is placed in relation to its surroundings”, a “state”, a “relative position or combination of circumstances at a given moment” or “the sum of total internal and external stimuli that act upon an organism within a given time interval”. We define the term “situation” in a similar way, but with a more operational aim, as the set of all decisive factors that should ideally be considered by the robot in selecting the correct behavior pattern at a given moment. These decisive factors are:

- perceivable objects in the environment of the robot and their suspected or recognized states;
- the state of the robot (state of motion, presently executed behavior pattern, ...);
- the goals of the robot, i.e. permanent goals (survival, obstacle avoidance) and transient goals emerging from the actual mission description (destination, corridor to be used, ...);
- the static characteristics of the environment, even if they cannot be perceived by the robot’s sensors at the given moment;
- the repertoire of available behaviors and knowledge of the robot’s abilities to change the present situation in a desired way by executing appropriate behavior patterns.

Figure 2 illustrates the definition of the term “situation” by embedding it in the action-perception loop of a situation-oriented behavior-based robot. The actions of the robot change the state of the environment, and some of these changes are perceived by the robot’s sensors. After assessing the situation an appropriate behavior is selected and executed, thus closing the loop. The role of a human operator is to define external goals via a man machine interface and to control behavior selection, e.g., during supervised learning.

Although situation-oriented robot control has proven much more robust and flexible under real-world conditions than classical model-based control it is not perfect. One reason is that, obviously, the robot cannot base its behavior selection on a “true” or “real” situation, but only on an internal

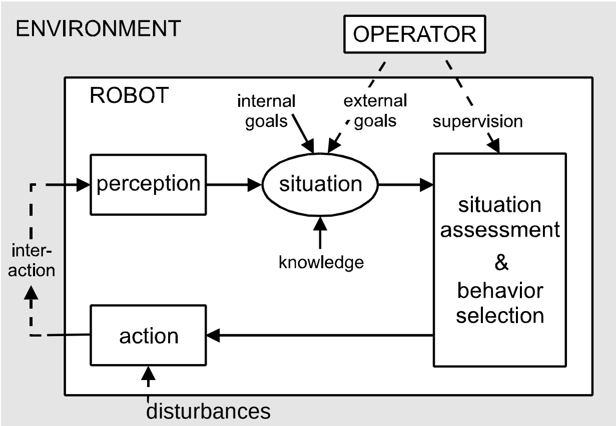


**Figure 2:**
The role of "situation" as a key concept in the perception-action loop of a situation-oriented behavior-based robot.

image of the situation as created by the robot according to its sensor information and its – always imperfect – knowledge of the world and of its own characteristics. Also, disturbances during the behavior execution can lead to non-expected situations. Although the disturbances may be corrected by either adjusting behavior-immanent parameters or selecting a different behavior, they will usually cause the robot to move in a non-ideal way.

## 2.4 Perception

Model-based robot control depends on a continuous flow of numerical values describing the current state of the robot and its environment. These values are derived from measurements performed by the robot's sensors. One problem here is that the quantities that are needed for updating the numerical models may be difficult to measure, e.g., the distance, mass and velocity of some external object that is posing a collision danger. Also, there are certain important decisions that cannot be made on the basis of measurements alone; the hypothetical decision whether in a particular situation a collision with a parked car should be brought about in order to avoid a collision with a pedestrian is an example.

Humans and other organisms, on the other hand, do not depend on measurements for controlling their motions. If, for instance, we want to sit down on a chair or pass through an open door, we do not first measure the size of the chair, the door, or our body; rather, we make a qualitative judgement whether the chair is high or low, or whether the door is wide or narrow, and then execute a sequence of motions that is adequate for the situation. In short, we substitute perception for measurement.

According to Webster's Dictionary "perception" is:
- a result of perceiving;
- reaction to sensory stimulus;
- direct or intuitive recognition;
- the integration of sensory impressions of events in the external world by a conscious organism;
- awareness of the elements of the environment.

"To perceive" means, according to the same source,
- to become aware of something through the senses;
- to become conscious of something;
- to create a mental image;
- to recognize or identify something, especially as a basis for, or as recognized by, action.

Typical questions to be answered by perception are:
- Which objects exist?
- What is the relationship between objects?
- Is it necessary to react? How?

Perception, rather than measurement, is thus a prerequisite for, and a complement of, situation assessment. Vision is the ideal sensing modality for perception because it is capable of supplying very rich information on the environment.

The actual design and implementation of a behavior pattern and of related perceptual processes depend on the robot's environment and task. A mobile robot navigating in a network of passageways needs different behaviors and recognition modules than a walking robot intended to explore rough terrain. In the sequel we will explain how we implemented visually controlled behaviors for an autonomous indoor vehicle navigating in a network of passageways, a stationary calibration-free manipulator grasping objects, and a service robot docking to tables and eventually intended to combine locomotive and manipulative actions of the former mentioned robots.

# 3 Vision-Based Navigation

Autonomous navigation of robots in complex environments like, e.g., factories, warehouses, or office buildings, is of great practical interest. Three system architectures based on different navigation concepts were investigated with our vision-guided mobile robots:

- **coordinate-based navigation**, relying on accurate odometers and precise maps that are difficult to provide;
- **behavior-based navigation** with the need of artificial landmarks, entered into a topological map that could eventually be generated automatically;
- **situation-oriented behavior-based navigation** without needing any artificial landmarks, only relying on a topological map which is extended by suitable attribute lists to facilitate reorientation and coordination of behavior patterns; attributes may consist of, e.g., visible "natural" landmarks (i.e., fixed visible objects) along the robot's path or geometric properties of the robot's environment that have been "measured" with its own (inaccurate) odometers.

Basing our robots' navigation on coordinate systems, exact kinematic models of the drive system and precisely measured maps (with help of a theodolite!) was time consuming and did not yield reliable results. In sharp contrast, behavior-based navigation in combination with an attributed topological map allows a large tolerance for errors regarding the odometry and the exact metric of the modeled network. It suffices to represent the exact topology, since navigation could rely on visual cues only.

## 3.1 Realized System Architecture

Figure 3 gives an overview of the system architecture that has been realized for our situation-oriented behavior-based robot. In this section we give only a short introduction to the different modules and their interaction (see [Bischoff et al. 1996] for details):

- a *situation module* taking into account all the decisive factors explained in section 2.3 and basing thereupon the dynamic selection of behaviors
- a *sensor module* comprising an object-oriented vision system as the main sensor and a proprioceptor system that provides auxiliary information needed by certain behavior patterns
- an *actuator module* executing commanded behaviors by activating a sequence of control laws for the drives
- an extendable *knowledge base* providing information about the static characteristics of the environment and the actual mission and goals
- a *man machine interface* for operator intervention and status display

Reasoning about the current situation is realized within the situation module which interacts with the sensors, the actuators, the data base management, and the man machine interface in a bidirectional way. In Figure 3 an arrow pointing away from the module indicates a request or a command, and an arrow pointing towards the module indicates a data and information flow. Fusing the data and information from the different modules makes situation assessment and behavior selection possible. Acting as the core of the whole system this module is responsible for system management and behavior coordination.

Each cooperation between actuators and sensors is supervised by a specific coordination process within the situation module. By activating and deactivating these coordination processes the management process realizes the situation-dependent concatenation of elementary behavior patterns to complex and elaborate action.

## 3.2 Object Recognition and Behavior Control

The basis for behavior-based navigation is a robot possessing a certain repertoire of built-in behaviors and cognitive abilities, e.g.,

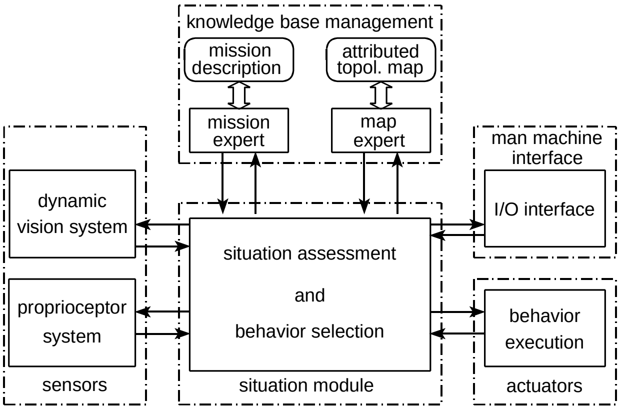


**Figure 3:**
System architecture of a situation-oriented behavior-based robot.

- to recognize and follow a passageway,
- to recognize and turn at a junction, if necessary,
- to recognize objects and avoid collisions while moving,
- to recognize goal objects/locations in the environment.

Vision is an ideal sensing modality to realize such cognitive abilities. Combined with the robot’s awareness of its actual situation a very good prediction of what objects are likely to be visible in the next moment, what they will look like, and where and how to find them in the image. This knowledge is a key factor for reducing the search space, thus saving valuable processing time.

Four basic cognitive abilities enabling the robot to recognize walls, junctions, certain goal points, and rectangular landmarks have been implemented and allow the robot to execute the behaviors “moving along corridor”, ”turning at corners and junctions”, and “driving towards goal points or landmarks” (Figure 4). Details of the vision and control algorithms that are necessary for these behaviors have been published elsewhere [Bischoff et al. 1996].

## 3.3 Experiments and Results

**Experimental robot *ATHENE II*.** The described concepts for controlling a mobile robot by object-oriented vision and behavior-based navigation were implemented as a final state machine and tested on the mobile robot *ATHENE II* (Figure 5). A PC-AT 80486, 50 MHz, is used

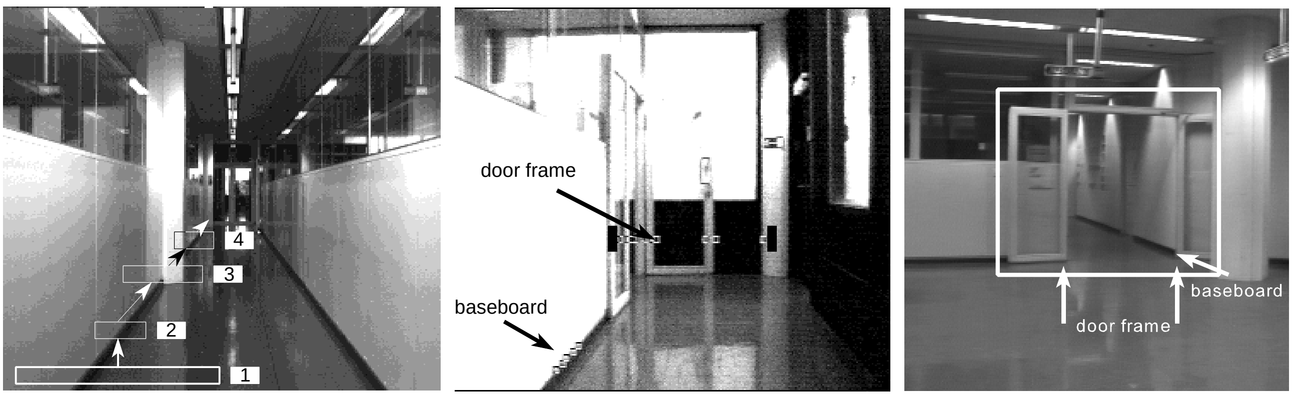


**Figure 4:**
Cognitive Abilities. **Left**: baseboard finder (initialization phase). **Center**: elementary behavior pattern following a baseboard; at the same time distance measurement based on motion stereo to determine the right moment for turning to the left (squares are marking detected object features). **Right**: traversing open area by detecting goal points (door frame and baseboard).


**Figure 5:**
*ATHENE II*, a vision-guided mobile robot.

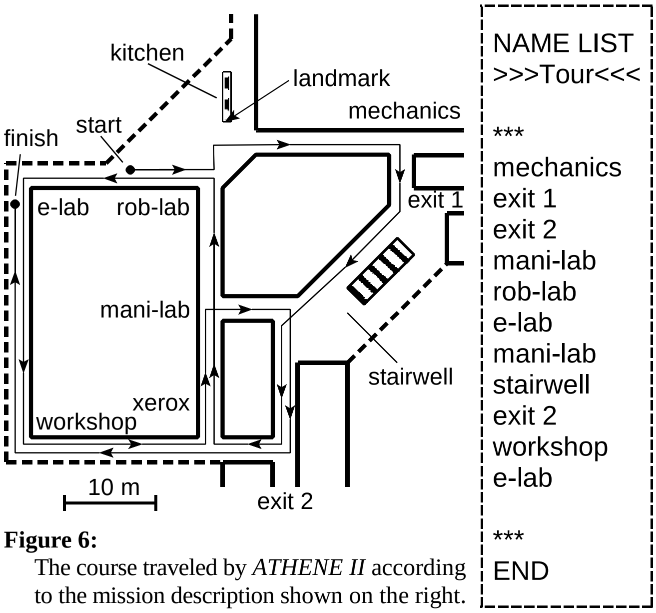


**Figure 6:**
The course traveled by *ATHENE II* according to the mission description shown on the right.

as system controller and hosts the image processing system based on a single T805 transputer frame grabber (cycle time 100 ms).

**Laboratory test course environment.** Several different experiments were carried out in a part of a general purpose laboratory building [Wershofen 1996] (Figure 6). It is illuminated by both artificial light and natural light shining through the glass walls causing space and time variant conditions. Due to the dimensions of the vehicle all corners must be considered narrow and, therefore, require a relatively accurate maneuvering (clearance 10-15 cm).

**Complex robot motion control.** The situation-oriented behavior-based navigation concept was verified in a number of real-world experiments. The velocity of the vehicle was up to 0.8 m/s while driving along corridors and 0.2 m/s at turns. The conducted experiments confirmed a large tolerance for map errors (about 2 m for distances and 10° for angles), even if the course included points where high accuracy is needed, e.g., to correctly turn around corners into narrow passageways.

**Missions.** Mission descriptions may be given to the robot either as a list of behaviors to be executed, or, much more conveniently, as a list of locations to be passed. Figure 6 shows an example of a course that was correctly followed by the robot after it had received the mission description in the form of a list of locations. A more convenient human-robot communication and interaction is discussed in [Graefe, Bischoff 1997].

**(Re-)orientation.** At the beginning of another experiment the robot was commanded to drive to the e-lab, starting from an unknown position somewhere in the corridor between the workshop and the xerox machine. In order to reorient itself it first randomly concatenated wall following and turning behaviors, simultaneously searching distinctive landmarks that could be matched to landmark descriptions stored in the attributed topological map. When the robot was near the rob-lab it found the landmark in front of the kitchen and eventually completed its mission.

**Learning.** Two forms of learning were investigated, supervised and unsupervised learning. In both cases the goal was to have the robot generate, or extend, an attributed topological map of the environment. During supervised learning the robot explores an unknown environment together with a human teacher. The geometric information is provided by the robot's dead-reckoning system, and relevant location names as well as attributes relating to landmarks, are entered by the teacher. When the cognitive abilities of the robot are not sufficient to safely navigate at critical points the teacher provides the correct behavior or action pattern to solve this navigational problem.

Experiments with unsupervised learning are still in their early stages. To achieve results comparable to those obtained in supervised exploration runs the sensory abilities of the robot have to be further improved.

# 4 Vision-Based Calibration-Free Manipulation

## 4.1 Concept

A novel concept for vision-based manipulator control has been introduced previously [Graefe 1995]. It eliminates the need for a calibration of the robot and of the vision system, it uses no world coordinates, and it comprises an automatic adaptation to changing parameters. The concept is based on the utilization of laws of projective geometry that always apply, regardless of camera characteristics, and on machine learning for the acquisition of knowledge regarding system parameters. Different forms of learning and knowledge representation have been studied, allowing either the rapid adaptation to changes of the system parameters or the gradual improvement of skills by an accumulation of learned knowledge.

The concept is simple. While the robot watches its end effector with its cameras, like a playing infant watches his hands with his eyes, it sends more or less arbitrary control commands to its motors. By observing the resulting changes in the camera images it "learns" the relationships between such changes in the images and the control commands that caused them. After having executed a number of test motions the robot is able to move its end effector to any position in the images that is physically reachable. If, in addition to the end effector, an object is visible in the images the end effector can be brought to the object in both images and, thus, in the real world.

Based on this concept a robot can localize and grasp objects without any knowledge of its kinematics or its camera parameters. In contrast to other approaches with similar goals, but based on neural nets, no training is needed before the manipulation is started.

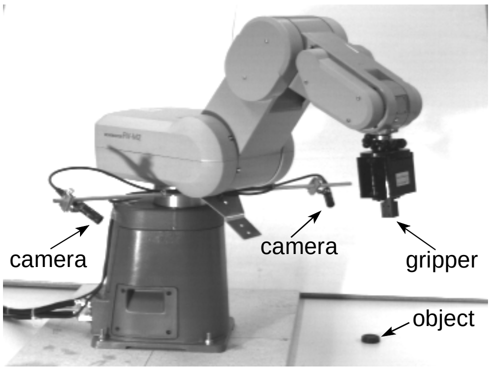


**Figure 7:** Manipulator robot consisting of an articulated arm (5 degrees of freedom) with a two-finger gripper and a stereo vision system.

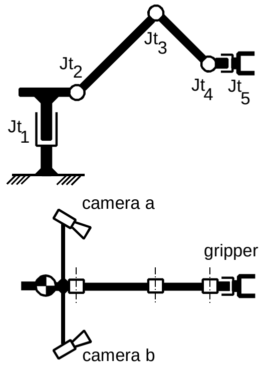


**Figure 8:** Schematic diagram of the articulated arm and camera arrangement.

## 4.2 Experimental Results

We have implemented the concept on an articulated arm robot (Mitsubishi Movemaster RV-M2) with 5 degrees of freedom, corresponding to the 5 joints $Jt_1$ to $Jt_5$ (Figures 7 and 8). The cameras have been attached to the robot on metal rods at the first link so that they rotate around the (vertical) axis of $Jt_1$ together with the arm. They are mounted in a rather unstable way to make the impossibility of any calibration or precise adjustment obvious, and to allow easy random modifications of the camera arrangement.

Objects of various shapes were successfully grasped, although no knowledge regarding the optical or kinematic parameters of the robot was used. Even arbitrary rotations of the cameras were tolerated while the robot was operating. Key to this unusual robustness is the renunciation of quantitative model knowledge (i.e., camera model and kinematic model of the arm) and the direct transition from image coordinates to motor control commands, without using any inverse perspective or kinematic transformations.

The gripper approached the object in a series of steps with some test motions in-between ("look-think-move" approach). The grasping process is, therefore, relatively slow (about 50 s).

In a first implementation the relationships between control commands and the resulting motions in the camera images were forgotten after the end of each grasping experiment. This made the system insensitive to parameter changes between experiments, but it also prevented the robot from learning from experiences and from improving its motion skills.

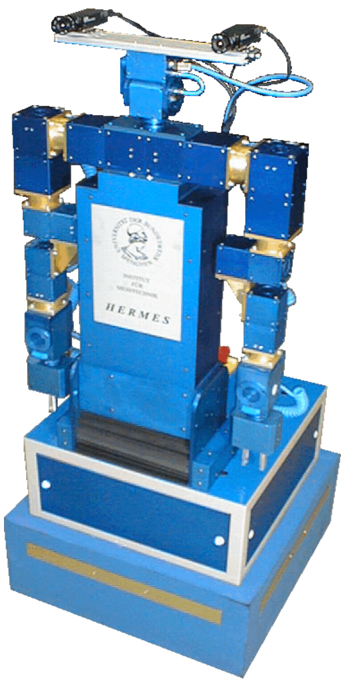

**Figure 9:** The humanoid robot *HERMES* with an omnidirectional wheel base, two arms with 6 degrees of freedom each and a stereo-vision system on a pan-tilt head; dim.: 70 cm x 70 cm x 170 cm.

## 4.3 Long-Term Memory

We are currently working on the realization of a long-term memory for the purpose of accumulating knowledge; it should be capable of learning, re-learning and forgetting, and of course, of real-time operation. The key idea is to store the image coordinates of the gripper in both images (4 dimensions!) and the control commands that have brought the gripper to that particular location, in 4-D tables. When later a similar location is to be reached suitable control commands may be recalled from the table.

As [Xie et al. 1997] have shown it is possible to realize permanent learning and a gradual skill improvement: With each experiment the motions of the robot become faster and smoother, and after a few successful grasps the time required reduces to only a few seconds. Open questions are the optimal form of knowledge representation, the acceleration of learning in more complicated situations, the realization of re-learning, and the grasping of arbitrarily shaped objects in arbitrary poses.

# 5 Vision for Service Robots

## 5.1 Introduction

A first generation of service robots already assists or rationalizes manipulation, transportation and processing tasks in various service domains. Possible application areas include cleaning, transport and handling of goods, surveillance and protection of buildings, construction and maintenance. All robots working in these domains have been designed to provide special solutions for specific problems. They are difficult to adapt to slightly modified tasks and environments.

To develop the next generation of service robots that could be used in many different environments (domestic, public and industrial) for a variety of tasks (e.g., elderly care, helping handicapped people, co-workers in factories) much research is still needed to improve considerably design and safety concepts, locomotion and manipulation capabilities, cooperation and communication abilities, reliability, and – probably most important – adaptability, learning capabilities and (visual!) sensing skills. To advance research in these fields we have developed the versatile experimental robot *HERMES* (Figure 9) based on our previously gained experiences.

## 5.2 Design and Realization of *HERMES*

Future service robots will interact much closer with humans and work in environments which are typically co-inhabited by humans and especially designed for humans. These environments will not be modified for the benefit of robots in any way. If a robot is to operate in such surroundings it should be designed according to an anthropomorphic model and should have sensory and motor skills comparable to those of a human. Such a humanoid design also facilitates the interaction between humans and robots, making it ideally as easy and natural as between humans.

In designing our humanoid experimental robot we placed great emphasis on modularity and extensibility [Bischoff 1998]. All drives are realized as modules with compatible mechanical and electrical interfaces; each drive module consists of 2 cubes containing a motor-transmission combination, power electronics, sensors, a micro-controller, and a communication interface. A standardized CAN bus connects all drive modules with the main computer. *HERMES* runs on 4 wheels, arranged on the centers of the sides of its base. The front and rear wheels are driven and actively steered, the lateral wheels are passive. Two drive motors of 500 W each are sufficient for an acceleration of 1 m/s² and a maximum speed of 2 m/s.

The manipulator system consists of two articulated arms with 6 degrees of freedom each on a body that can bend forward (130°) and backward (-90°). The work space extends up to 120 cm in front of the robot. The heavy base guarantees that the robot will not loose its balance even when the body and the arms are fully stretched. Currently each arm is equipped with a two-finger gripper that is sufficient for basic manipulation experiments.

Main sensors are two video cameras on a pan-tilt platform; numerous proprioceptors are integrated in the motor modules, additional sensors may be connected via available interfaces. A hierarchical multi-processor system is used for information processing and robot control. The control and monitoring of the individual drive modules is performed by the sensors and controllers embedded in each module. The main computer is a network of digital signal processors (DSP, Texas Instruments TMS 320C40) embedded in a standard industrial PC. Sensor data processing (including vision), situation recognition, behavior selection and high-level motion control are performed by the DSPs, while the PC provides data storage and the human interface. The developer and user interfaces have been realized under Windows NT 4.0. A radio Ethernet interface allows to control the robot remotely. Separate batteries for the motors and the information processing system allow a continuous operation of the robot for several hours without recharging.

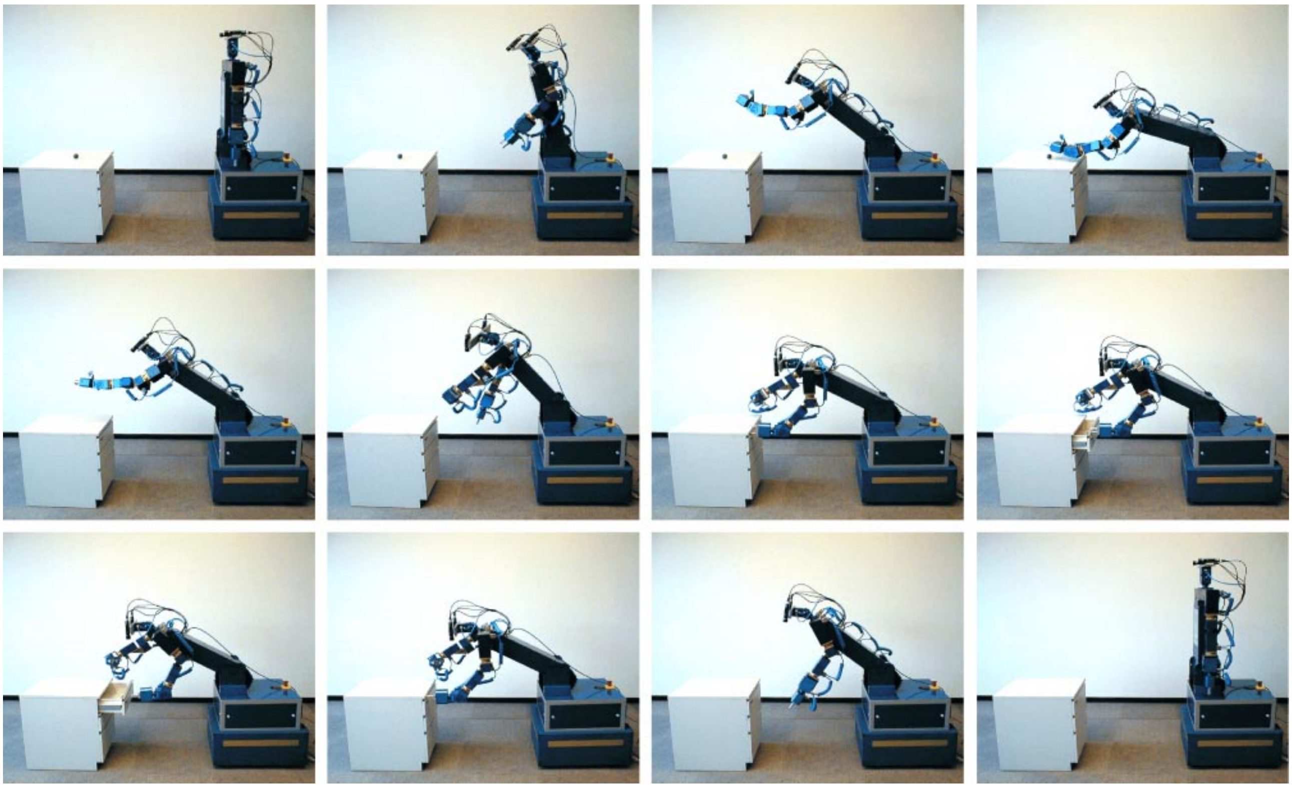

**Figure 10:**
*HERMES* executing a complex pre-programmed motion sequence (duration: 120 s): picking up a ball from a table with its left gripper, handing it over to the right gripper, grasping a drawer and pulling it open with its left gripper, dropping the ball into the drawer, and closing drawer by pushing it. This motion sequence demonstrates how effectively *HERMES'* articulated body extends the reach of its arms while the cameras remain in a favorable position for close supervision of the manipulation tasks.

## 5.3 First Experimental Results

The robot's hardware has been completed, and a few driving and manipulation experiments have been conducted to evaluate its performance.

**Pre-programmed and taught motion patterns.** To allow simple high-level control of *HERMES*, basic functions were written in C to control groups of modules as functional units: the mobile wheel base, the left and the right arm, the camera platform and the body. These simple functions enable us to drive *HERMES* manually and to teach it simple pick and place operations (via a wireless keyboard), but also to program more sophisticated tasks, e.g., handing over a ball from one hand to the other, or opening a drawer. Figure 10 shows a series of 12 snapshots from a 120 s motion sequence in which *HERMES* is picking up a ball from a side table and dropping it into a drawer to be opened and closed. The positions of the ball and the table had been taught in a previous test run. These snapshots illustrate how effectively the bendable body enlarges the robot's work space, and that it also allows the cameras to be kept in a favorable position for closely supervising all kinds of object manipulation.

**Vision-guided behaviors.** Methods for object manipulation that do not depend on any calibration of the robot or its vision system, as well as situation-oriented behavior-based control and navigation methods using visual feedback are currently being implemented on *HERMES*. First implementations of vision-guided behaviors enable *HERMES* to fixate known objects during navigation (e.g., landmarks), and to approach docking stations (e.g., tables).

## 5.4 Vision-Guided Docking Behavior

The main task of a mobile service robot with manipulator arms is to manipulate various objects at different locations. Prerequisite for manipulating objects is to bring them within the working range of arms and grippers, i.e., to navigate the robot sufficiently close to the object to be manipulated. We propose a new vision-guided method enabling the robot to approach objects that are in its field of view and to stop in front of them at a pose (position and orientation) that is suitable for subsequent manipulation. It is based on the continuous tracking of a predefined reference line that corresponds to a physical edge of the docking object (e.g., a table's front edge) and the fixation of a

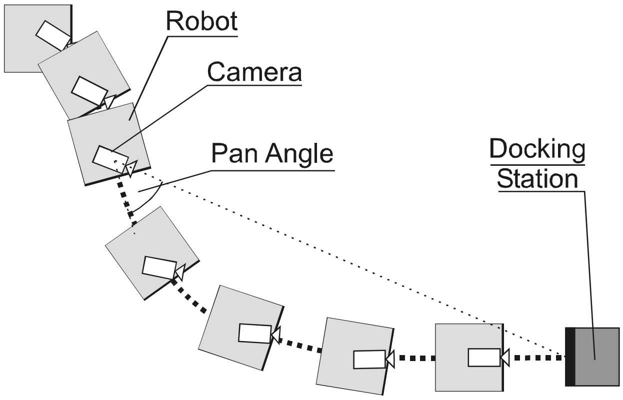


**Figure 11:**
Typical trajectory followed by the robot during the docking maneuver; it is directly derived from sensor readings and not pre-calculated (see text). The start pose may be arbitrary, the end pose is reached with sufficient accuracy towards the middle of the docking station's front side and in parallel to it.

reference point of the object (e.g., a table corner). This fixation or gaze control yields encoder values (angles) for the pan ($\alpha$) and tilt ($\gamma$) axis of the camera platform. Together with the slope $m = tan\ \beta$ of the tracked reference line in the image they allow to directly control the robot's trajectory during the docking maneuver (Figure 11).

The reference point together with the camera's fixation point determine the final docking position laterally in front of the docking object, and the reference line helps to adjust the final orientation of the robot with respect to the object. A third constraint is necessary to define the final (longitudinal) distance to the object. Although this constraint could also be derived from image primitives (e.g., edges of the robot's front at a certain image distance to the reference line of the docking object) we have chosen to stop the vehicle, for simplicity reasons, at a certain tilt angle $\gamma_{end}$ corresponding to a sufficiently small distance to the front of the docking station (Figure 12c).

In the current implementation of this docking behavior the following assumptions have been made: laws of projective geometry apply and the lines of the cameras' CCD chips are always in parallel to the even floor; they are also in parallel to the robot's front if the pan angle of the camera platform $\alpha = 0°$. The camera's pan angle $\alpha$ and its tilt angle $\gamma$ can be directly obtained from encoder readings of its platform. Reference line and reference point of the docking object are both visible until the end of the docking

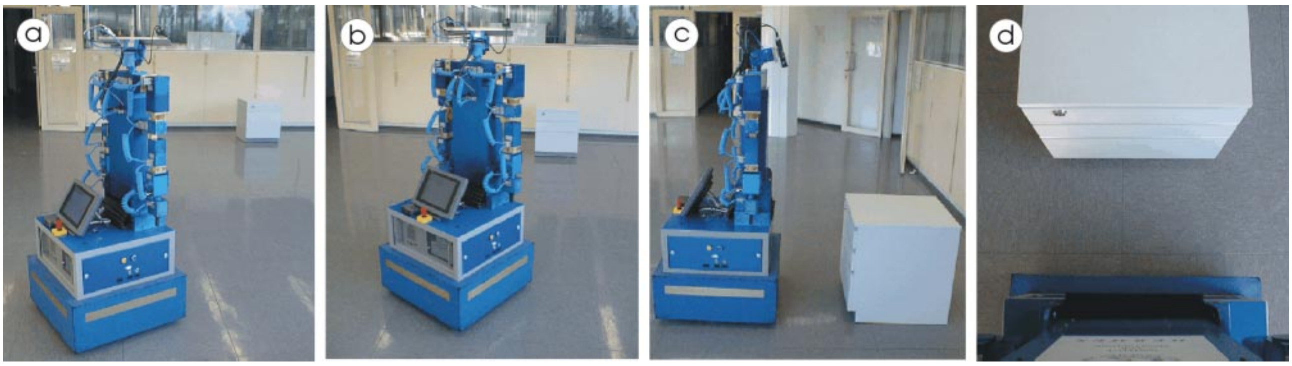


**Figure 12**:
Vision-guided docking behavior: (a), (b) *HERMES* approaches the docking station while fixating it with its vision system; (c) end pose is reached; (d) *HERMES'* view of the final docking pose

maneuver, and the reference line must be in parallel to the floor.

**Image processing.** Prerequisite for tracking and fixating objects is a reliable and fast image processing, preferably at video rate. In our driving experiments the lower edge of the side table has been chosen as reference line, its lower right corner as reference point. Only the right camera of the vision system is used. The image processing is divided into two stages: One, detection of the table: It is based on a separation of the floor from all other objects, thus getting a contour line of objects standing on the floor; then, matching this contour line with a generic 2-D model of cubic objects leads to the identification of possible docking stations. Two, tracking of the reference features is based on simple edge and corner detectors. The angle $\beta$ of the reference line is obtained by linear regression. Adaptive thresholds are used to detect and track the chosen features reliably despite specular reflections and difficult illumination conditions. The tracking cycle time is 15-25 ms depending on the size of the object in the image.

**Robot control**. The main idea of the docking behavior is to directly derive the steering angle $\lambda$ of the robot's driven wheels at each moment from the values of $\alpha$, $\beta$ and $\gamma$ in order to maneuver the robot into its predefined final docking pose (e.g., $\alpha = 0°$, $\beta = 0°$, $\gamma = \gamma_{end}$). The vehicle's speed is set to a constant value during the maneuver (ca. 0.2 m/s). Figure 11 shows a typical trajectory that could be generated by either of the two different docking behaviors which have been developed and tested with *HERMES*:

1) The steering angle $\lambda$ is a linear combination of $\alpha$ and $\beta$, $\lambda = \alpha + g(\gamma)\,\beta$, with the weight function $g = k\,(\gamma_{end} - \gamma)$, i.e., $\beta$ is weighted more in the beginning of the docking maneuver and less in the end.

2) A rule-based method with two phases: In the first phase the steering angle $\lambda$ is controlled to keep $\alpha = constant$ (e.g., *30*°), which brings the robot closer to the front of the docking object on a trajectory that points away from it; the end of this phase is reached when *ß* is below a certain limit (e.g., *1*°). Then, in the beginning of the second phase, $\lambda$ is set to $\alpha$ which causes the robot to turn almost on spot and eventually approach the docking station from the front. The first phase is omitted if the robot is already standing somewhere in front of the docking station.

**Results.** Both implementations of the docking behavior yield good results. The robot may start its approach from arbitrary poses and stops sufficiently close in the middle of the docking station's front, while both front sides (robot and table) are in parallel (Figure 12). It is important to notice that both the robot's trajectory and the final docking pose are directly derived from sensor data (image features and encoder readings). They are not calculated from distance measurements, kinematic models and inverse perspective transforms with respect to a fixed reference frame (world coordinate system). Another advantage is that the method is generic and suitable for all kinds of docking or goal objects where a reference line and a specific point can be defined. In addition, it may be implemented in a calibration-free or self-calibrating way.

## 6 Conclusions

In this paper our fundamental concepts and principles for designing and building intelligent robots have been presented. We strongly believe that vision – the sensor modality that predominates in nature – is also an eminently useful and practical sensor modality for robots. It provides rich and timely information on the environment and allows real-time recognition of dynamically changing situations. Situation-dependent perception and behavior selection rather than measurement and control based on quantitatively correct models are additional key factors for advanced robots. Motor control commands should be derived directly from sensor data, without using world coordinates or parameter-dependent computations, such as inverse perspective or kinematic transforms.

Building robots according to these rules and testing them intensively in the real world lead to robust and intelligent robots with the ability to adapt themselves to modified environmental conditions and tasks. Three generations of experimental robots, and particularly their visual and behavioral capabilities, have been introduced: an autonomous vehicle, a calibration-free manipulator, and the prototype of a versatile service robot that will eventually combine the abilities of both and bring us closer to our long-term goal, the intelligent, adaptable and universally employable service robot.

## References

**Babcock, P. (1976).** Webster's Third New International Dictionary of the English Language. G. & C. Merriam Company, Springfield, MA, U.S.A., 1976.

**Bischoff, R. (1998).** Design Concept and Realization of the Humanoid Service Robot *HERMES*. In A. Zelinsky (ed.): Field and Service Robotics. Springer, Berlin, 1998, pp. 485-492.

**Bischoff, R.; Graefe, V.; Wershofen, K. P. (1996).** Object-Oriented Vision for a Behavior-Based Robot. In D. Casasent (ed.): Intelligent Robots and Computer Vision XV. Boston, November 1996, Proc. SPIE, Vol. 2904, pp. 278-289.

**Graefe, V. (1989).** Dynamic Vision Systems for Autonomous Mobile Robots. Proc. IEEE/RSJ International Workshop on Intelligent Robots and Systems, IROS '89. Tsukuba, pp. 12-23.

**Graefe, V. (1992):** Visual Recognition of Traffic Situations by a Robot Car Driver. Proceedings, 25th ISATA; Conference on Mechatronics. Florence, pp 439-446.

**Graefe, V. (1995).** Object- and Behavior-oriented Stereo Vision for Robust and Adaptive Robot Control. Int. Symp. on Microsystems, Intelligent Materials, and Robots, Sendai, pp. 560-563.

**Graefe, V.; Bischoff, R. (1997).** A Human Interface for an Intelligent Mobile Robot. 6th IEEE Int. Workshop on Robot and Human Communication. Sendai, Japan, Sept. 1997, pp. 194-197.

**Graefe, V.; Kuhnert, K.-D. (1988):** Towards a Vision-based Robot with a Driver's License. Proc. IEEE Int. Workshop on Intelligent Robots and Systems, IROS '88. Tokyo, pp. 627-632.

**IFR International Federation of Robotics (1997).** Key data for the world robot market. http://www.ifr.org/stat.htm.

**Wershofen, K. P. (1996)**. Zur Navigation sehender mobiler Roboter in Wegenetzen von Gebäuden – Ein objektorientierter verhaltensbasierter Ansatz. Dissertation, Fakultät für Luft- und Raumfahrttechnik der Universität der Bundeswehr München.

**Xie, Q.; Graefe, V.; Vollmann, K. (1997).** Using a Knowledge Base in Manipulator Control by Calibration-Free Stereo Vision. IEEE Intern. Conference On Intelligent Processing Systems. Beijing, China, Okt. 1997, pp. 1307-1311.