

# Robot Vision without Calibration

*Volker Graefe*

Institute of Measurement Science  
Universität der Bw München  
85577 Neubiberg, Germany  
Phone: +49 89 6004-3590, -3587; Fax: +49 89 6004-3074

## Abstract

A novel concept for vision-based robot control is introduced. 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. An extension of the concept to the navigation of mobile robots is discussed.

**Keywords:** Robot Vision, Calibration-Free Robot Control, Machine Learning

## 1 Introduction

Conventional stereo vision methods for grasping objects require repeated calibration of the manipulator and the vision system. The necessity of repeated calibrations is a major impediment to the practical application of intelligent robots in industrial and other less than perfectly structured environments. Not only does it constitute a significant cost factor, but a dependence on the accuracy of models and their parameters also prevents systems from behaving robustly in the real world where continuous and unforeseeable changes are the rule rather than the exception. In contrast to conventional robots, animals and humans do not depend on any calibrations for controlling their motions. Also, they can effortlessly adjust to changes in their sensory system, e.g. when putting eye glasses on or off, and in the environment, e.g., when driving different cars of unknown characteristics.

To avoid the necessity of calibrations for robots Graefe and Ta [1995] have proposed an approach to adaptive and calibration-free robot control. The method was later named “object- and behavior-oriented stereo vision” [Graefe 1995]. It was initially tested and evaluated in experiments involving the grasping of objects by a vision-guided manipulator arm; the objects to be grasped were in those experiments limited to cylindrical objects with a vertical axis of symmetry. With a newer version of the algorithm it is possible, however, to manipulate objects of nearly any shape and in an arbitrary orientation, in addition to flat cylindrical objects [Vollmann, Nguyen 1996]. Additional degrees of freedom of the robot are used to accommodate the

arbitrary object orientation. The concept underlying object- and behavior-oriented stereo vision will be described in section 3.

It may seem at a first glance that methods not depending on world coordinates and on accurate internal models obtained by a careful calibration can yield only qualitative and inaccurate results. However, this is not correct. One counter-example is the motion stereo method introduced by [Graefe 1990] and [Huber, Graefe 1991]. It allows quite accurate distance measurements (relative errors less than 1 %) in the real world without any camera calibration. Moreover, nature provides an abundance of examples for systems that operate and survive in a wide variety of unstructured environments without ever needing an explicit calibration.

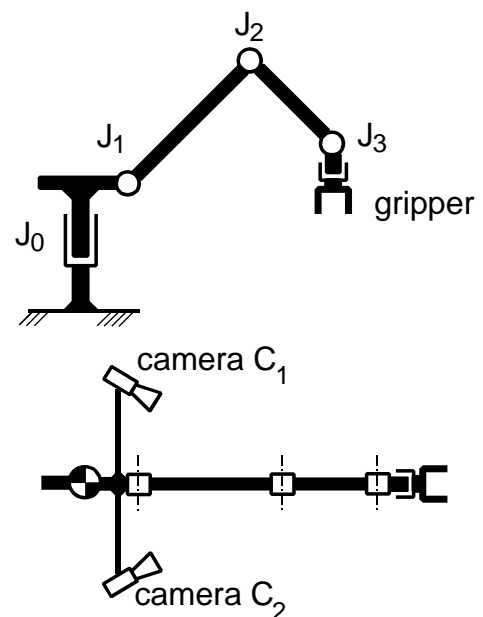
Two classes of calibration-free systems are conceivable, systems that do not need any model knowledge at all, and systems that acquire whatever model knowledge they may need in the course of their normal operation. Systems of both types have in common that they do not need any explicit calibration and that they automatically adjust to changes of their own parameters and of the environment. In the systems that we have realized so far we have employed a mixed approach: we have strived to minimize the number and required accuracies of model parameters that the system must know, and we have designed the systems in such a way that they acquire and maintain the necessary knowledge by continuous learning (and forgetting) as part of their ordinary operation.

Other researchers, e.g., [Cooperstock, Milios 1993], have addressed similar problems. They have used neural networks to represent the functions necessary for controlling a manipulator that was mounted on a movable platform. An advantage of neural networks is that they are, generally speaking, well suited for representing arbitrary nonlinear functions, which is exactly what is needed here. On the other hand, neural networks require an – often lengthy – training phase before they can begin to do useful work, and it is difficult to design neural networks that can continue to learn, and relearn, while they perform their duty after the initial training phase. Such an ability is, however, needed if the system is to adapt to continuous changes.

## 2 Experimental Setup

Figure 1 shows schematically the manipulator that we have used to develop and demonstrate the concept of object- and behavior-based stereo vision. The task is to grasp and pick up an object that is located at an initially unknown position in the work space of the manipulator. The only sensors are two cameras at unknown locations and of unknown optical characteristics; the cameras rotate together with the robot around its vertical axis as they are fixed relative to the link that connects  $J_0$  with  $J_1$ . Each camera is mounted in such a way that it overlooks a section of the work plane that is accessible to the gripper; the exact viewing directions are unknown.

The objects we have used in our experiments are simple cylindrical disks lying on a horizontal support. Therefore, grasping the object is easy once the center of the (open) gripper coincides with the center of the



**Figure 1**

The robot arm joints and the camera arrangement

object. The motion that is necessary to bring the gripper to the object consists of a rotation of the robot around its vertical axis (joint  $J_0$ ) to bring the object into the robot's work plane, and a 2-dimensional motion of the gripper within the robot's work plane (joint  $J_1$  and  $J_2$ ;  $J_3$  is controlled in such a way that the gripper always maintains a vertical orientation).

### 3 Object- and Behavior-Based Stereo Vision

#### 3.1 Motion within the Work Plane

Let us assume, for a moment, that the object is already located in the robot's work plane, and discuss the 2-dimensional planar motion first. Each camera performs a one-to-one mapping of the work plane onto its image plane. Since the object is known to be located in the work plane the image of the gripper coincides with the image of the camera if, and only if, the gripper coincides with the object. This holds true, regardless of any particular characteristics of the camera or the robot; therefore, we base the control of the robot's motions on this fact.

The task is then to make the **image** of the gripper coincide with the **image** of the object to be grasped. The key idea is here that we are not at all concerned with distances, coordinates, or any other relations in the real world, but only with the image of either one of the cameras. The robot can influence the location of its gripper **in the image** by modifying the contents of two control words,  $w_a$  and  $w_b$ . We humans know that the robot has an articulated arm with motors moving the joints and each control word governing one of the joint angles; we even know the lengths of the links and which control word is associated with which joint. The robot knows nothing of all this; it only knows that there are two control words, each one of them influencing the location of the gripper in the camera image in its own way, and it assumes that the relationships between control word increments and resulting image motions are, in some sense, smooth.

The Robot accomplishes the rendezvous between the gripper and the object by first modifying the control words, one at a time, by a small amount and observing the effects in the image. It then estimates the gain coefficients relating image motions to control word increments and computes by linear extrapolation those control word increments that would bring the gripper to the object. Knowing that linearity is not guaranteed and that collisions should be avoided it executes only a fraction of the computed motion which brings the gripper closer to the target than it was before. After two or three iterations the rendezvous is accomplished (for more details cf. [Graefe, Ta 1995]).

#### 3.2 Rotation around the vertical axis

Although the images of the two cameras will normally look rather different, either one of them should lead to the correct control word increments; therefore, evaluation of both images should lead to the same result. However, as Figure 2 shows, this holds true only if the object is, indeed, located in the work plane of the robot. If it is not, an evaluation of the two images according to the method sketched in the previous paragraph with its implicit assumption that the object is located in the work plane

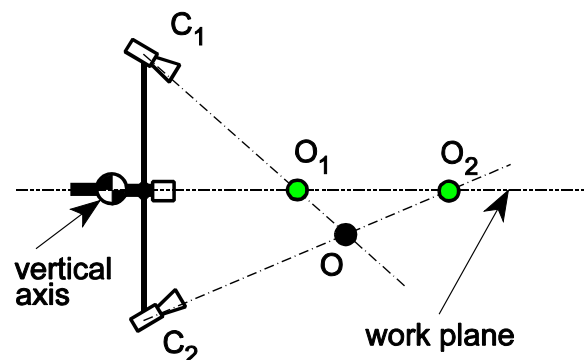


Figure 2

Disparity of apparent object locations,  $O_1$  and  $O_2$ , corresponding to an object,  $O$ , outside of the robot's work plane

yields two different sets of motion control words: one that would bring the gripper to position  $O_1$ , and one that would bring it to a different position,  $O_2$ .

Consequently, if an evaluation of the two camera images yields two significantly different sets of control words, the third control word, associated with joint  $J_0$ , should be modified first until the resulting rotation of the robot has brought the object into the work plane. The magnitude and sign of the appropriate increment of the control word may be determined in a similar way as described for the two other control words. All three degrees of freedom necessary for making the gripper rendezvous with an object in three-dimensional space may, thus, be controlled. In experiments with a real manipulator arm objects were, indeed, grasped successfully.

### 3.3 Discussion

Key points of the concept described above are

- ▶ a direct transition from image coordinates to motion control commands
- ▶ the complete absence of world coordinates
- ▶ online determination of gain coefficients by observing the results of motion control commands, either during special test motions or during the normal operation of the robot

While the implementation was sufficient for proving the concept, and especially the ability of goal-directed motion without previous knowledge of system parameters, it has shortcomings in at least two respects:

- ▶ The object to be grasped is represented by its center point, and it is implicitly assumed that in the images of both cameras the center point of the object image corresponds to the same point within the physical object. This assumption is approximately correct if the objects are circular disks lying on a horizontal support, as in our experiments, but not in general. Circular disks also simplify the task of grasping because they do not require the gripper to be rotated into any particular orientation.
- ▶ The robot starts moving without any knowledge of its system parameters and acquires such knowledge while it is moving, but it cannot accumulate any experience. At the end of each experiment it completely forgets everything it has learned. This makes the system rather robust against sudden changes of parameters between experiments (even major changes of the orientation of a camera between successive experiments are tolerated without any degradation of the system's performance), but, more importantly, it also prevents the robot from improving its skill in repeated experiments.

Work is currently performed to overcome these shortcomings. It will be briefly sketched in the next section.

## 4 Further Developments

### 4.1 Objects of General Shape and Orientation

An extension of the described implementation allowing horizontally oriented elongate objects, such as ballpoint pens to be grasped, too, was reported by Vollmann and Nguyen [1996]. Such objects require the gripper to be oriented parallel to the axis of the object before the actual

grasping is performed. For this purpose a fourth degree of freedom of the manipulator, rotation of the gripper, was activated and controlled in a similar way as the other degrees of freedom.

Grasping a ballpoint pen without restricting its orientation requires five degrees of freedom of the robot to be controlled. This has been accomplished in a number of cases. Grasping objects of an arbitrary shape in an arbitrary orientation would require an arm with at least six degrees of freedom. Moreover, recognizing suitable points for grasping such objects and a suitable reference point is difficult. If an object has two opposite surfaces certain areas of which are parallel to each other these surfaces may define a grasping point and a reference point. We have had some success with such an approach if the orientation of the objects in space was such that it facilitated both recognition and grasping, but the work on such problems is still going on.

## 4.2 Learning, Remembering, Forgetting

Ideally, a calibration-free robot should be able to start working immediately after it has been switched on, without requiring a training phase. Since it does not yet know its own characteristics its movements will not be optimal. The robot should then learn from experience while it is performing its task and improve its skills over time. For this purpose it must have some kind of long-term memory that the robot described above is lacking. However, the characteristics of the robot or of the environment may change, for instance due to the aging of parts or to some maintenance that is performed on the robot. In such a case the robot should be flexible enough to replace or modify the contents of its long-term memory, either gradually or in large steps, depending on the nature of the changes, in order to adapt to the new situation.

A long-term memory for the robot that picks up cylindrical disks may be built in the form of a four-dimensional table. The table is addressed by the coordinates of a point in the images of both cameras ( $2 * 2$  dimensions) and contains the set of control words that causes the gripper to move to that location and the gain coefficients that are valid at that location. We have built such a system. The image of each camera is covered by  $9 * 9$  equidistant support points, therefore, the table has  $9^4$  fields. When the robot is supposed to move the gripper to a certain location it first determines the support point that is closest to the target point in 4-D space. If the corresponding field in the table contains data the gripper is immediately moved to the support point by using the stored data; if the table field is still empty the gripper is moved to the support point according to the try-and-iterate method described above, and the control word set thus determined is stored in the table. From the support point the gripper is then moved to the target point by an extrapolation based on the stored gain coefficients.

This approach was successful in the sense that the could function right from the beginning and could work faster and faster as it learned more and more support points. It is also efficient in the sense that only those table fields are filled with data that belong to support points that have actually been visited. The ability to adapt to changing system parameters is, however, still limited.

## 5 Conclusions and Outlook

Robot vision without calibration is possible. It has been demonstrated under somewhat restricted conditions, but it seems that the presented concepts may be extended to more general cases. The approach we have taken has two key elements:

- ▶ we design the robot in such a way that it utilizes certain laws that always apply, regardless of system parameters; an example is the fact that when the image of the gripper coincides with

the image of an object in the images of two cameras, the object is at the same location as the gripper in 3-D space, regardless of any camera parameters;

- ▶ we provide the robot with the ability to learn its own characteristics to the extent necessary for motion control.

The problems of learning and knowledge representation have not yet been fully solved. The dimensionality and, thus, the size of the table we are using must be increased as the number of degrees of freedom that must be controlled is increased. As the table gets larger the learning speed decreases because the number of table fields that should be filled with data increases, and the likelihood of revisiting any specific support point decreases when the number of potential support points gets larger.

The contents of a table field that has already been filled with data are slightly modified every time the gripper is again moved to the same support point if the contents seem to be incorrect. We might be able to realize an improved relearning and adapting ability if we used a Kalman filter for updating the table. We are currently working on an implementation of this idea.

Automatic recognition of sudden drastic changes of the system parameters that make the previously accumulated experience useless has not been implemented yet, although it may be desirable. In such a case forgetting everything that has been learned and starting from scratch may be more efficient than gradual relearning.

The concept presented here is not limited to manipulator arms. A similar approach may be used for guiding mobile robots and autonomous vehicles. The key idea is again to exploit certain laws that apply regardless of the characteristics of the camera used for guiding the vehicle. For instance, if a camera is translated parallel to a straight guide line the location of the guide line in the image does not change. By measuring the motions of selected features in the image of a camera carried by the vehicle it is possible to generate motor control commands that make the vehicle follow a desired course relative to visible objects. Again, no calibration of the camera is needed, and the dynamic and kinematic characteristics of the vehicle may be largely unknown. Moreover, the approach should make it possible to modify the dynamic characteristics of the system at any time in a simple and straight-forward way, and to place greater weight either on the agility of the vehicle or on the smoothness of its motions depending on the requirements of the situation. Work to evaluate these ideas in experiments with a mobile robot is currently in progress.

## References

- Cooperstock, J. R.; Milios, E. E. (1993):** Self-supervised learning for docking and target reaching. *Robotics and Autonomous Systems* 11 (1993), pp 243-260.
- Graefe, V. (1990):** Precise Range Measurement by Monocular Stereo Vision. Japan-USA Symposium on Flexible Automation. Kyoto, pp 1321-1324.
- Graefe, V. (1995b):** Object- and Behavior-oriented Stereo Vision for Robust and Adaptive Robot Control. International Symposium on Microsystems, Intelligent Materials, and Robots, Sendai. pp 560-563.

**Graefe, V.; Ta, Q. (1995):** An Approach to Self-learning Manipulator Control Based on Vision. IMEKO International Symposium on Measurement and Control in Robotics, ISMCR '95. Smolenice, pp 409-414.

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

**Vollmann, K.; Nguyen, M.-C. (1996):** Manipulator Control by Calibration-Free Stereo Vision. In D. Casasent (ed.): Intelligent Robots and Computer Vision XV, Proceedings of the SPIE, Vol. 2904. Boston, pp. 218-226.