

# Object- and Behavior-oriented Stereo Vision for Robust and Adaptive Robot Control

Volker Graefe

Institut für Meßtechnik  
Universität der Bundeswehr München  
85577 Neubiberg, Germany

## 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 and comprises an automatic adaptation to changing parameters. A key point of the concept is the newly proposed method of "object- and behavior-oriented stereo vision". Contrary to conventional stereo vision methods it uses an uncalibrated camera system and allows a direct transition from image coordinates to motion control commands of a robot.**

## 1. Introduction

Vision-based control of a manipulator typically requires a careful calibration of the optical subsystem, including the camera(s) and the lighting, and of the mechanical subsystem. Such a calibration tends to be rather cumbersome, and thus, expensive. Moreover, because neither subsystem is perfectly stable the calibration has to be repeated after relatively short time intervals, and also after any maintenance.

A robot not requiring any calibration of its subsystems would therefore be of great practical advantage. An approach to the realization of such a robot is presented in the sequel. It utilizes a close interaction between image interpretation and motion control to essentially perform a continuous implicit calibration as a side effect of normal operation. This self-calibration is automatically limited to those parameters which are actually necessary for motion control.

The continuous updating of the internal knowledge of system parameters may be considered a form of learning. It also adapts the control to changes of characteristics of the robot (e.g. mechanical wear or replacement of parts), of the sensors (e.g. camera mounting or focal length of the camera lens), and of the environment (e.g. positions of objects to be manipulated, or lighting).

Self-supervised learning for docking and target reaching has previously been described by Cooperstock and Milios (1993). They use a set of neural networks to realize a robot that is able to approach an object and grasp it without requiring a calibration. In contrast to them we use a novel stereo vision method

that provides a direct transition from image data to motion control commands, and does not require any training.

## 2. Experimental Setup

A specific setup has been used for developing and testing our approach to vision-based robot control in real-world experiments.

A 5-degree-of-freedom articulated arm (Mitsubishi Movemaster 2) is used for picking up objects. Of the 5 degrees of freedom, one refers to the rotation of the gripper around its axis. It will not be considered in the sequel. The remaining 4 degrees of freedom correspond to the joints  $J_0 \dots J_3$  (cf. Figure 1). Joints  $J_1 \dots J_3$  allow the gripper to be moved within a certain section of a vertical plane, the work plane. Joint  $J_0$  allows the arm to be rotated around a vertical axis, thus determining the azimuth of the robot's work plane. In our experiments joint  $J_3$  was always controlled in such a way that the gripper was in a vertical orientation, as indicated in Figure 1. Our arm has thus three independent degrees of freedom remaining, corresponding to the joints  $J_0$ ,  $J_1$  and  $J_2$ .

The task to be executed by the robot is to find a specific object that may be located somewhere in the robot's 3-D workspace, and pick it up.

Two video cameras are mounted on the arm. They participate in the rotation of the arm around its vertical axis, but relative to the work plane of the robot they are fixed. The cameras are mounted on opposite sides of the work plane. The location and orientation of each camera are somewhat arbitrary and not exactly known, but each camera should be mounted in such a way that its field of view covers that area within the work plane in which objects are to be manipulated.

The control of the robot is based entirely on an interpretation of the images of the two cameras. Uncontrolled ambient light is used for illuminating the scene.

## 3. The Classical Control Approach

A classical approach for controlling such a robot would evaluate the camera images according to the well-known methods for stereo evaluation. This evaluation requires all internal and external camera parameters (location, orientation, focal length, principal point, etc.) to be known with great accuracy. The stereo evaluation would then deliver the location of the object to be grasped relative to some coordinate system that is defined by the orientations and locations of the cameras. As an alternative, a carefully aligned projector illuminating the scene with a specific light pattern could be used in combination with one or two cameras. In both cases the coordinates of the object in a ground-based coordinate system would then be determined by an appropriate coordinate transformation.

Finally, the joints of the robot would be controlled in such a way as to move the gripper to that point in the ground-based coordinate system which was determined to coincide with the location of the object. This control implies a coordinate transformation between the joint angles and the ground-based coordinate system which requires accurate knowledge of the robot's dimensions, kinematics and joint angles.

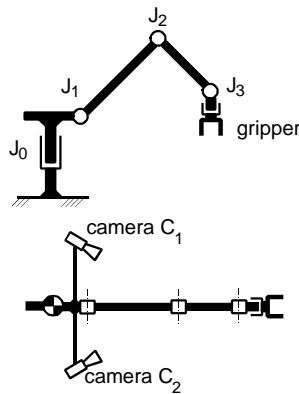
It is obvious that the classical approach requires exact knowledge of numerous system parameters. This knowledge must be provided by a careful, and usually cumbersome and expensive, calibration of the entire system. Moreover, the calibration has to be repeated frequently to minimize the effects of changes in the system parameters.

## 4. The New Control Approach

### 4.1 Characteristics

In sharp contrast to the classical approach, our approach requires no knowledge regarding:

- ▶ the exact locations of the cameras  
(except that the cameras should be located some distance away from the work plane of the robot in an opposite arrangement)
- ▶ the exact viewing directions and the internal parameters of the cameras  
(except that both cameras should have the actual work space of the robot in their fields of view)
- ▶ the dimensions, kinematics and joint angles of the robot  
(except that, for practical reasons, we presently assume that the robot is of the articulated arm type, and that the general type of the gripper and the number of degrees of freedom of the system are known)
- ▶ the quantitative relationships between the control words sent to the motor controllers and the resulting motions



**Figure 1**  
The robot arm joints and the camera arrangement

(except that these relationships are assumed to be "smooth").

### 4.2 Concept

To be picked up, an object must be located in the work plane of the robot. The task of picking up an object may therefore be conceptually decomposed into two subtasks: rotating the robot around its vertical axis until the work plane coincides with the object to be picked up, and a subsequent motion of the gripper within the work plane. This decomposition only serves to simplify the description. In reality the robot should execute a single motion in three-space to approach the object and finally grasp it.

Since we are mainly interested in the control of the arm, and not in sophisticated grasp strategies, the objects used in our experiments are of simple cylindrical shape. To grasp such objects it is sufficient to make the center of the (open) gripper coincide with the center of the object, and close the gripper.

#### 4.2.1 Motion within the Work Plane

The control for the motion within the work plane will be described first. Let us assume for the moment that the work plane already coincides with the object. The task is then to modify the joint angles,  $\alpha_1$  and  $\alpha_2$ , corresponding to the joints  $J_1$  and  $J_2$  (Figure 1), in such a way that the gripper coincides with the object. Either one of the cameras may be used as a sensor for accomplishing this goal. Let us assume, as an example, that camera  $C_1$  is being used. It performs a one-to-one mapping between an area of the work plane and the camera's image plane. The images of the object and of the gripper coincide if, and only if, the object and the gripper are at the same location. To grasp the object it is, therefore, sufficient to make the *image* of the gripper coincide with the *image* of the object.

Let us assume that the joint angles,  $\alpha_0$ ,  $\alpha_1$  and  $\alpha_2$  of the robot are controlled by 3 internal control words,  $w_a$ ,  $w_b$  and  $w_c$ . Actually it is unnecessary to know which control word belongs to which joint angle, but for simplifying the subsequent explanation, let us assume that  $w_a$  and  $w_b$  belong to  $\alpha_1$  and  $\alpha_2$ . To make the gripper rendezvous with the object it is then necessary to assign appropriate values to the control words  $w_a$  and  $w_b$ . To determine these values we must know the gain coefficients,  $\delta x_1/\delta w_a$ ,  $\delta y_1/\delta w_a$ ,  $\delta x_1/\delta w_b$ , and  $\delta y_1/\delta w_b$ , relating the control words,  $w_a$  and  $w_b$ , to the coordinates,  $x_1$  and  $y_1$ , of the gripper in the image of camera  $C_1$ . ( $\delta$  symbolizes a numerical approximation to the partial derivative).

If the gain coefficients are still unknown for the present position of the gripper, the following *self-calibration procedure* may be used for determining them:

1. Modify control word  $w_a$  by a small amount  $\Delta w_a$ . (Modifying the control word by too large an amount might be dangerous. If a reasonable magnitude of the increment is completely unknown choose the smallest amount possible. If no visible motion of the gripper results double the increment and repeat this step.)
2. Measure the resulting displacement of the gripper in the image  $(\Delta x_1, \Delta y_1)^T$ .
3. Repeat steps 1 and 2 for the second control word,  $w_b$ .

Once the gain coefficients have been determined in certain locations in the image they may be stored in a table containing sets of these coefficients for each one of a number of areas within the image. For subsequent motions of the robot the stored values may then be used instead of measuring them again. If a motion based on the memorized coefficients turns out to be incorrect, the self-calibration may be repeated and the stored coefficients may be updated.

If the coefficients are known the following *rendezvous procedure* may be executed to make the gripper rendezvous with the object:

1. Assume, as an approximation, that the relationships between control word increments and the resulting displacements of the gripper are linear in a certain image area around the present location; compute those control word increments which will cause a motion of the gripper in the image of a magnitude of not more than, say, 10 pixels in the direction to the object. (Solve a system of 2 linear equations.)
2. Execute the motion.
3. Observe the resulting displacement of the gripper in the image, and compare it with the expected one. If the difference is significant use it for updating the local gain coefficients.
4. Repeat steps 1 to 3 until the image of the gripper coincides with the image of the object.

#### 4.2.2 Three-dimensional Motion

Up to this point it has been assumed that the object was initially located in the work plane. Therefore, only one of the two cameras was necessary, and it did not matter which one of them was actually used. Their images, and the mappings between control words and image coordinates of the gripper, are different, but the control word increments computed in step 1 of the rendezvous procedure above are the same, regardless of the camera whose image is used. The reason is that there is only one correct motion in three-space and, thus, only one correct set of control word increments.

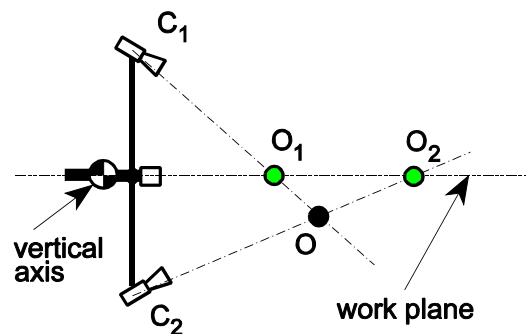


Figure 2

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

Now let us assume that the object is not located in the work plane. In this case, the joint angle  $\alpha_0$ , corresponding to joint  $J_0$ , must be modified. Whether such a situation exists may be determined by executing step 1 of the rendezvous procedure (and possibly the self-calibration procedure) twice, once with the image from each camera.

Figure 2 shows why the two results will be *different* if the object is *not* located in the work plane. Figures 3 and 4 show the images obtained by the two

cameras in such a situation. In the rendezvous procedure it is assumed implicitly that any object seen is located in the robot's work plane. Controlling the joints,  $J_1$  and  $J_2$ , based on the image of camera  $C_1$  would, therefore, generate a motion of the gripper towards  $O_1$ , the projection of  $O$  onto the work plane as seen by  $C_1$ . Similarly, basing the motion control on the image of camera  $C_2$  will cause a motion of the gripper towards  $O_2$ . The two sets of motion control words will be identical if, and only if,  $O$  is located in the work plane, because only then will  $O_1$ ,  $O_2$ , and  $O$  be at the same location.

If the two sets of control words computed on the basis of the two images are different it may be concluded that the object is not located in the work plane. In this case, the control word for joint  $J_0$  should be modified, causing the azimuth angle,  $\alpha_0$ , of the work plane to change. (If it were initially unknown which control word controls joint  $J_0$  it may easily be determined by modifying each control word in turn. Modifying  $\alpha_1$  or  $\alpha_2$  makes the gripper move in the camera images, while varying  $\alpha_0$  makes the object move in the images.) The correct sign and magnitude of  $\Delta w_0$ , the increment of the corresponding control word which makes the work plane coincide with the object, would be determined basing on the disparity of the two sets of control words,  $w_a$  and  $w_b$ , computed from the images of the two cameras. A simple trial-and-error method may be used for this.

#### 4.3 Object- and Behavior-oriented Stereo Vision

The described concept amounts to a novel realization of stereo vision. While conventional stereo vision measures the disparity between **corresponding features** in two images in order to determine the coordinates of these features in **Euclidian space**, in our realization of stereo vision we measure the disparity between **corresponding objects** in two images in order to determine the coordinates of the objects in the **control word space** of a robot. By this approach we avoid abstract coordinate transformations; instead, we use image data directly to control a behavior of the robot, or interactions of the robot with physical objects. This "object- and behavior-oriented stereo vision" has two important advantages:

- ▶ The correspondence problem, which is known to be an extremely hard problem is much more tractable when the correspondence between images of physical objects is sought than when the correspondence between images of features is sought.
- ▶ The direct transition from image coordinates to motion control words makes the knowledge of many hard-to-measure optical and mechanical system parameters unnecessary; moreover, it lends itself to the realization of learning and adaptive robots.

#### 4.4 Summary of the Concept

By the concept described above, all three joint angles of the robot may be controlled in an appropriate way without any previous knowledge of system parameters. In the interest of speed the rotation of the arm around the vertical axis and the motion of the gripper in the work plane may be executed simultaneously. An average of the control words computed on the basis of the data from the two cameras may be used for controlling the joints,  $J_1$  and  $J_2$ , even while the azimuth angle is still being adjusted.

A robot based on this concept will be able to begin working with almost no knowledge of its system parameters; later it will become more efficient by learning from experience gained in the course of normal operation. By continuously updating its stored knowledge it will be able to adapt to changes in the environment and in system parameters.

### 5. Experimental Results

The approach was tested in a series of grasping experiments. The object to be grasped had a cylindrical shape, about 2.5 cm in diameter and 8 mm in height, and dark color (Figure 3 and 4). The object was placed on a support of unknown height (0 to about 10 cm) somewhere on a table where it could be reached by the robot arm.

In the experiments the object was located and grasped reliably, regardless of its initial location in the workspace of the robot. The adaptability of the system was tested by moving one of the

cameras by an arbitrary amount. As expected, the grasping operation could immediately continue without any recalibration whatsoever.

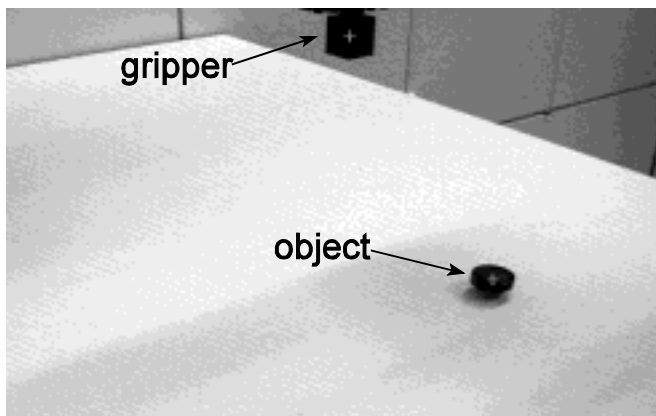
The motion is rather slow; approaching the object and grasping it requires about 60 s. This slowness is mainly due to the fact that, for reasons of simplicity, in this initial implementation of the concept the control of the robot is handled in a rather static way: a pause is made after each motion command to wait until the robot has completely stopped; the result of the previous command is then evaluated, and the next command is issued. Also, the two motions, motion of the gripper within the work plane and rotation of the work plane, are not yet executed simultaneously.

It would be much more efficient to issue velocity commands to the robot, observe the resulting image velocities, e.g. by the methods proposed by (Huber, Graefe 1991), and issue corrective velocity commands to the robot while it is still moving. It may, however, be difficult to implement such a control strategy on the presently used robot and its associated control box.

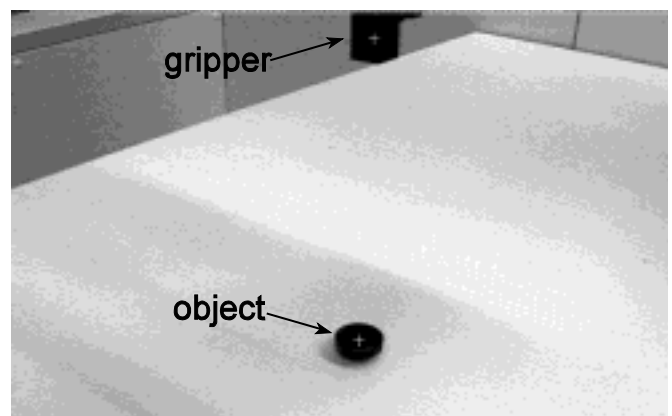
### 6. Conclusions

A novel method for controlling a robot manipulator by vision has been introduced. It does not require the robot, or the vision system, to be calibrated, and it provides an automatic adaptation to changing system parameters. The key point of the method is a direct transition from image data to motion control commands. This direct transition avoids not only cumbersome computations of coordinate transformations and inverse kinematics, but it also makes it unnecessary to know numerous system parameters that would otherwise have to be determined by expensive calibration procedures.

The validity of the concept has been demonstrated in real-world experiments where an articulated arm robot grasped objects that were located at arbitrary positions in 3-D space. The speed and efficiency of the system may be improved in the future by a more dynamic control strategy and by including a long-term memory for accumulating knowledge of the mapping between



**Figure 3**  
Scene from the left camera



**Figure 4**  
Scene from the right camera

image data and motion control gain coefficients.

## References

- Cooperstock, J. R.; Milius, E. E. (1993):** Self-supervised learning for docking and target reaching. *Robotics and Autonomous Systems* 11 (1993), pp 243-260.
- 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.
- Huber, J.; Graefe, V. (1991):** Quantitative Interpretation of Image Velocities in Real Time. *IEEE Workshop on Visual Motion. Princeton*, pp 211-216.