The Sensor-Control Jacobian as a Basis for Controlling Calibration-Free Robots

Volker Graefe and André Maryniak

Institute of Measurement Science
Federal Armed Forces University Munich
85577 Neubiberg, Germany

Phone: +49-89-6004-3587, -3343
Fax: +49-89-6004-3074
E-mail: {Graefe | Andre.Maryniak}@UniBw-Muenchen.de

Abstract

A method for controlling the motions of robots is presented. It is based on the newly introduced sensor-control Jacobian matrix and avoids all quantitative modeling of the robot and the sensor system. The sensor-control Jacobian contains the coefficients that relate those changes in sensor data which are caused by a motion of the robot to the robot control words that caused the robot to move and, thus, the sensor data to change. A wide variety of tasks of robots can be reduced to minimizing the differences between actual sensor data and a set of hypothetical sensor data corresponding to some desired state. All these tasks can be solved by this method.

The method is especially useful for calibration-free robots, since neither quantitative models of the mechanical, kinematic and control characteristics of the robot, nor knowledge of the sensor characteristics are required.

The sensor-control Jacobian may be determined automatically in real time while the robot is operating. This yields a high degree of adaptability and flexibility against unforeseen changes in the robot's parameters. Because the concept has an open structure it allows further extensions and improvements, e.g., in terms of the utilization of sensor data redundancy and machine learning.

For the purpose of evaluation, the concept has been implemented on a calibration-free camera-manipulator system. Real-world grasping experiments have demonstrated the effectiveness of the method.

1. Introduction

In Robotics, manipulation tasks play an important role. There exist many applications for such tasks, and solutions as well. The basic idea for conventional approaches is to formulate and program the task using mathematical descriptions – quantitative models – of the manipulator and its kinematics, of the sensors (e.g., for cameras the inverse perspective transformation) and of the arrangement of the sensors with respect to the manipulator ("hand-eye calibration").

The necessary transformations between different coordinate systems, like world coordinates, sensor coordinates and joint coordinates, involving, e.g., inverse kinematics are computationally expensive and often ill-conditioned. Moreover, and being even more important, the mentioned models need to possess a high, and hard to obtain and to maintain degree of accuracy; otherwise the method would not work.

If quantitative models are relied upon, a task can be performed successfully only if the actual conditions of the robot and the environment match exactly those that have been modeled. If wear and tear apply, or if parts have to be replaced, or if the environmental conditions change, a new calibration often becomes necessary. Such calibration procedures are usually expensive, even disregarding costs of machine down time; but failing to perform a re-calibration in due time may cause a malfunction of the robot and lead to damages that are even more expensive.

Therefore, various work has been done to develop robot control concepts that avoid, or at least reduce, the necessity of modeling and calibration. [Yoshimi, Allen 1994] use an uncalibrated camera system to perform a peg-in-hole task utilizing rotational invariance in a specific setup. [Hollinghurst, Cipolla 1994] avoid using an exact camera model by performing a self-calibration at four known positions. Doing this, they can handle calibration errors of the cameras as well as linear modeling errors of the manipulator.

[Graefe, Ta 1995] developed a concept enabling calibration-free manipulation by a direct transition from visual input information to control commands. This development was independent of a similar work by [Conkie, Chongstitvatana 1990]; moreover, it differs from that work by using the motor control word space instead of the joint space, thus avoiding the need of calibrating the joints. [Vollmann, Nguyen 1996] added the capability of handling a further degree of freedom in order to grasp objects requiring a special alignment of the gripper in a vertical orientation. In order to speed up the grasping process [Xie et al. 1997] introduced a knowledge base enabling the robot to utilize knowledge already learned automatically during past grasping tasks.

Here, we introduce another approach to the control of a calibration-free camera-manipulator system. It is based exclusively on data that are internal to the robot (sensor data and control commands) and uses a newly introduced sensor-control Jacobian matrix to compute control commands directly from sensor data.

Although the method has been developed in the context of a grasping task, it is of a general nature and may easily be applied to other types of tasks, too.

2. Problem Statement

A robot is to manipulate an object that is resting at an initially unknown arbitrary location somewhere in the robot's work space. The only sensors used are two calibration-free video cameras in a calibration-free stereo arrangement. The key point is that almost no model knowledge, and in any case, absolutely no quantitative knowledge of the characteristics of the robot is being used by the system or in its design. This includes knowledge of the mechanical, kinematic and control characteristics of the robot, as well as knowledge of the optical characteristics (internal and external parameters) of the cameras.

As an example for a typical manipulation task we study the grasping of an object by a robot.

We make the following assumptions:

- The robot's gripper and the object to be grasped are visible for both cameras.
- The internal and external camera parameters (optical characteristics, locations, and viewing directions) are unknown.
- The motions of the cameras are unknown. The cameras may either be at rest (located somewhere outside the robot), or they may move together with some part of the robot (not necessarily the end effector); they do not move independently, and they do not move relative to each other.
- There is no kinematic redundancy.
- The robot's motions are controlled by numerical control words, one for each degree of freedom. The functional dependencies of world coordinates of the gripper on control words, as well as of image coordinates on world coordinates of the gripper and the object, are unknown since we do not use quantitative models. However, we assume that a zero control word causes no motion, a word of small magnitude causes a small motion, and a word of larger magnitude causes a larger motion.

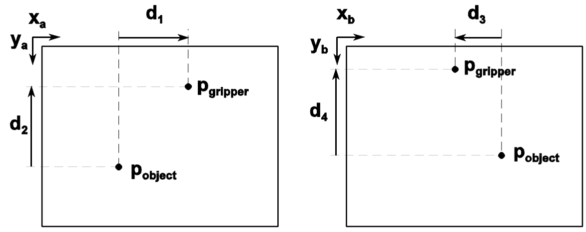

Figure 1

The gripper and the object (modeled as points) in the images of two cameras, a and b

In the context of this particular work, we are only interested in the problem of robot control without model knowledge, not in grasping strategies, collision avoidance or complex vision problems. Therefore, we have structured the environment for our experiments in such a way that both the gripper and the object are easy to locate in the images, and that the object is easy to grasp by a gripper approaching it from above without any risk of collisions with other objects.

3. The Concept for a Solution

For the purpose of describing the concept for our approach we model both the gripper and the object as visible points. The robot's vision system continuously delivers the image coordinates of both points in both cameras (Figure 1). The grasping may be considered to have been accomplished when the two points, gripper and object, coincide in the real world. Regardless of any camera characteristics, this is the case if, and only if, they coincide in the images of both cameras. The rendezvous between the gripper and the object may, therefore, be brought about by any sequence of robot motions that makes the two points coincide in the images(!) of both cameras simultaneously.

Since we want to manipulate point-like objects in 3-D space without kinematic redundancy, we assume that the robot has exactly 3 degrees of freedom. Thus, our control word vector, $\underline{c}$, has 3 control words as its elements[1], each control word controlling the motions of one degree of freedom.

[1] We denote vectors by underlining and matrices by capital letters.

Generally, when the robot moves, the image coordinates of the object or the gripper will change. Therefore, also the elements, $d_1 ... d_4$ (Figure 1), of the 4-D distance vector, $\underline{d}$, representing each the signed distance between the gripper and the object in one dimension of the two images, will change.

For any given configuration of the robot the relationship between a control word vector, $\underline{c}$, that is transmitted to the robot's actuators and the resulting change in the 4-D distance vector, $\underline{\Delta d}$, may be written as

$$\underline{\Delta d} = f(\underline{c}) \qquad \textbf{\textit{(1)}}$$

where the components of $\underline{\Delta d}$ are known (because they are measured by the vision system) and f is an unknown 4-D function of 3 variables. We assume that all partial derivatives of f exist; moreover, since a zero control word vector does not cause any motion of the robot, $f(\underline{0}) = \underline{0}$. Approximating f by the first-order terms of its Taylor expansion, therefore, yields

$$f(\underline{c}) \approx f(\underline{0}) + J \cdot \underline{c} = J \cdot \underline{c} \ . \qquad \textbf{\textit{(2)}}$$

J is the Jacobian matrix of $f(\underline{c})$; it contains all its partial derivatives and indicates how strongly each component of a vector of sensor data – in our case the 4-D image distance between the object and the gripper – is affected by each control word.

[Jägersand, Nelson 1995] use a similar matrix and call it "image Jacobian" or "visual-motor Jacobian". However, unlike us, they use joint angles as the independent variables, as [Hosoda, Asada 1994] did in a similar approach, too. [Yoshimi, Allen 1994] use Cartesian world coordinates as independent variables, and they call their matrix "image Jacobian". In order to avoid a conflict with these and other definitions we shall keep referring to our matrix of linear dependencies of sensor data on motor control commands as "sensor-control Jacobian".

It should be noted that in our approach we are neither concerned with the positions of the gripper or the object in the real world, nor with any configuration variables of the robot, such as joint angles or positions. We consider all of these to be unknown, hard to obtain, and – in a sense – irrelevant. Instead of them, our sensor-control Jacobian describes directly the relationship between quantities that are internal to the robot and, thus, easily available: the control words that have been transmitted to the actuators and the image coordinates that are being delivered by the vision system.

The sensor-control Jacobian may be determined experimentally by executing test movements, i.e., by making the robot move by a small amount in each one of its degrees of freedom and observing the resulting effects in the images. It may not even be necessary to execute specific test movements; according to [Jägersand, Nelson 1995] who refer to [Broyden 1965] a Jacobian may also be estimated by observing the normal operation of a robot.

If the linear approximation in (2) is valid, the rendezvous between the gripper and the object may be accomplished by computing and transmitting a control word vector $\underline{c°}$ that causes $\underline{d}$ to become $\underline{0}$. Computing $\underline{c°}$ is equivalent to the task of making $\underline{\Delta d} = -\underline{d}$, which may be done by solving the equation

$$-\underline{d} = f(\underline{c°}) \qquad \textbf{\textit{(3)}}$$

or, according to the approximation (2), the equation

$$-\underline{d} = J \cdot \underline{c°} \qquad \textbf{\textit{(4)}}$$

for $\underline{c°}$.

Equation (4) is equivalent to m scalar equations (m is the dimension of the sensor data vector) for n unknown control words (n is the number of degrees of freedom of the robot). Depending on the robot and the sensors being used, equation (4) may, thus, be under- or over-determined. If (4) is under-determined the sensor data are inadequate for an unconstrained task, and a solution cannot be found. If (4) is over-determined more sensor data than necessary are available. Such a redundancy may, in principle, be utilized for improving the robustness of the system.

In our case equation (4) is equivalent to 4 scalar equations for the 3 unknown control words; it is, thus, over-determined. In a separate paper ([Maryniak, Graefe 1998]) it is discussed how the redundant sensor data may be utilized to improve the robustness of the system but here we simply remove the redundancy by arbitrarily discarding one of the 4 equations. Instead of $\underline{d}$ we use a 3-D vector, $\underline{d}^*$, that contains 2 horizontal (x-) and 1 vertical (y-) component of the image distance vector. Instead of (4) we then have to solve

$$-\underline{d}^* = J^* \cdot \underline{c°}, \qquad \textbf{\textit{(5)}}$$

where J^* is the correspondingly truncated 3×3 sensor-control Jacobian.

In reality, the relationship between control words and image coordinates is non-linear because of the non-linearity of imaging and kinematics, and the matrix J is not known with perfect accuracy. Transmission of the control word vector $\underline{c°}$ as computed by solving (5), therefore, will not make the gripper coincide with the object exactly. However, it will reduce the distance between the two by some amount. (For a detailed discussion on the convergence, please refer to [Zeidler 1996], [Bronstein, Semendjajew 1990], [Halmos 1967], [Hutson, Pym 1980] and [Ljusternik, Sobolew 1975].)

Iterating the process and starting each iteration with the configuration of the robot that has resulted from the previous step will lead to a continued distance reduction. The iteration may be terminated when the distance has finally become small enough for practical purposes. (Here we assume implicitly that it will be possible to determine from an observation of the images that the distance between the gripper and the object in the real world has, indeed, become small enough. This question will be addressed in section 4.)

Depending on the degree of non-linearity and on the size of the previous step, it may be necessary to determine the sensor-control Jacobian, J, before each step (which consumes time), or it may be sufficient to use the same values for the elements of J as for the previous step.

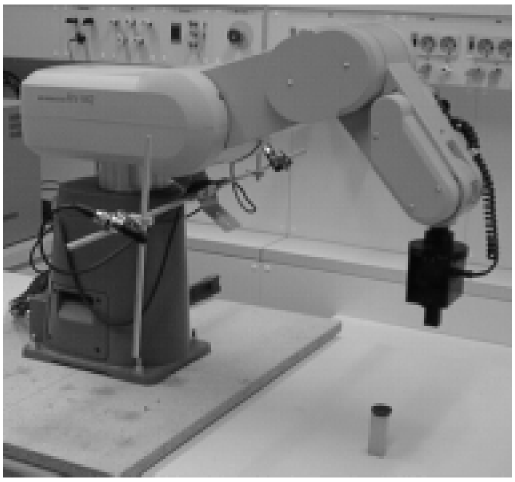

Figure 2

The robot used for the experiments; the black object below the gripper is a cylindrical disk that is to be grasped.

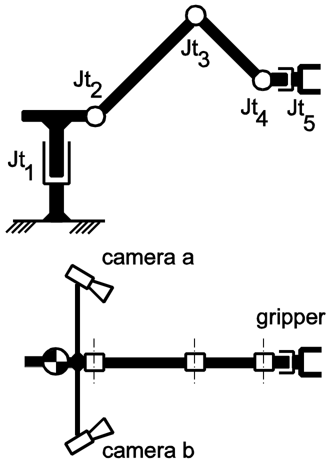

Figure 3

Schematic diagram of the robot and the camera arrangement

The concept introduced here may be extended in several ways. In the context of this paper it is assumed that the robot forgets after each grasping what it has learned about the relationship between the distance vector, $\underline{d}$, the state of the robot, and the sensor-control Jacobian, *J*. If the robot would remember such relationships beyond the end of a grasping process, it could build up some skill and eventually grasp objects much faster, possibly even in just a single step. Obviously, the concept introduced here may also be extended to situations requiring more than 3 degrees of freedom.

4. Implementation

To test the concept in real-world experiments we have implemented it using an articulated arm robot (Figures 2 and 3). The robot has 5 degrees of freedom (Figure 3), corresponding to the 5 joints, Jt_1 to Jt_5. Out of these, Jt_4 is controlled in such a way that the axis of the gripper remains always vertical. Jt_5 allows the gripper to rotate around its axis; it is not used here because our objects are flat circular disks that can be grasped with any orientation of the gripper around its vertical axis. The robot has, thus, 3 active degrees of freedom remaining. The cameras have been attached to the robot 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.

As reference points for the object and the gripper we use the centers of their images. These center points are relatively easy to determine in the images, but they do not necessarily correspond to any particular point on, or in, the physical objects. Therefore, it is not guaranteed that image coordinates extracted from the images of both cameras refer to the same point in the real world. Experiments must be performed to find out whether this is a serious problem.

If we actually tried, according to the previously described concept, to make the real gripper (in a closed condition) coincide with the object, they would collide with each other before the actual grasping even started. To avoid such a collision we define an intermediate position a few cm above the object and move the closed gripper in a first phase of the grasping process to the intermediate position. Once the gripper has arrived there, it is opened and then moves downward to grasp the object.

Approaching an intermediate position first, before actually grasping the object, has the additional advantage that small errors in the motion control as they may be caused, e.g., by the non-linearity of the system or by measurement noise do not lead to a collision of the gripper with the object. There are, however some problems with this idea:

- for a calibration-free robot the terms “a few cm” and “above the object” are meaningless, and
- the intermediate point to which the gripper is to be moved is invisible.

We have not yet been able to find a theoretically well-founded solution to these problems, but some ad-hoc solutions that allow our concept to be tested in experiments are available. The easiest one is to modify the *y*-coordinate of the object in the image by a certain number of pixels to make the object appear higher up in the image than it really is. The gripper is then moved to this apparent object position. In doing so we assume implicitly that the cameras are in a “reasonable” orientation; this means that „up” in the world corresponds approximately to „up” in both images. The magnitude of a suitable height offset is determined heuristically.

A better approach for defining an intermediate position would be to utilize the fact that the gripper is always in a vertical orientation and has a fixed size. The intermediate point could then be defined, e.g., as one apparent gripper diameter away from the object in the direction of the gripper axis. We have not yet implemented this second approach.

The iteration of movements is terminated, and the intermediate position is considered to have been reached with sufficient accuracy, when the sum of the absolute values of all 4 components of the image distance vector has become

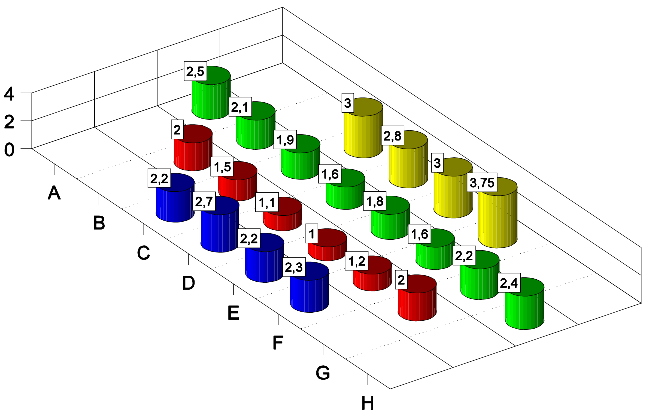

Figure 4

Average number of control steps necessary for approaching an intermediate position just above the object; ten measurements per field; the manipulator socket is situated at bottom left.

smaller than 16 pixels.

5. Experiments and Results

In order to evaluate the concept described in section 3 we performed a series of real-world experiments. A further goal was to find out whether our modeling of the gripper and the object by points used in our implementation holds and how the cost for determining the Jacobian can be reduced.

On an equally spaced grid of four times eight squares with a lateral length of 3.5 cm a light support carrying the dark cylindrical object was placed arbitrarily, and a grasping process was performed. We observed whether an intermediate position somewhere above the object where the second subtask should start could be reached, and how many control steps were necessary for this. Further, we observed whether the grasping was successful or not. This process was repeated ten times per square, while the position of the object to be grasped was changed arbitrarily each time.

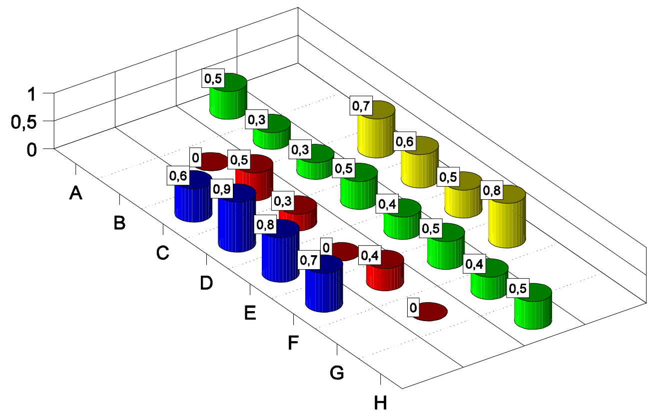

Figure 5

Standard deviation of the number of control steps

Figure 2 shows a typical situation, and Figures 4 to 7 the results.

Figure 4 shows that the nearer the object is situated to the borders of the work area of the robot the more control steps are necessary. Also, the standard deviation of their number increases (Figure 5). Both facts are due to the greater distance the gripper has to travel and the associated growing influence of non-linearities in the system. The average number of necessary control steps varies between 1.0 and 3.75.

After the intermediate position above the object has been reached the actual grasping phase is initiated. In our current implementation, visual feedback is not possible in this phase. This led initially to wrong grasps in about a third of all cases. The malfunctioning was eliminated by newly determining the Jacobian at the beginning of the grasping phase. The results are shown in Figure 6.

The good results that we have obtained in the central part of the work area indicate that our approach is correct and that

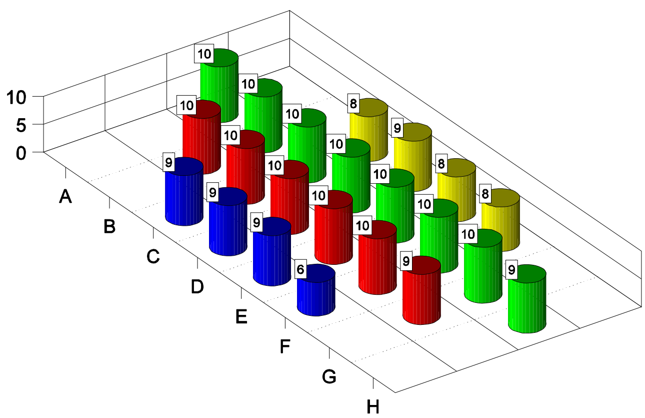

Figure 6

Number of successful grasping processes of 10 tests per field

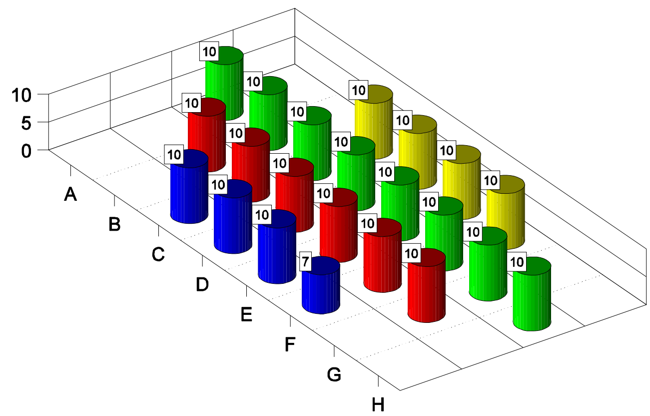

Figure 7

Number of successful approaches to the intermediate position just above the object of ten tests per field the modeling of the object and the gripper is justified. The inferior results near the borders of the work area are caused by limitations of the vision system, such as a limited search area that sometimes causes objects to be seen only partly.

However, due to the visual feedback, even such cases can often be treated; in 98.6 % of the experiments the intermediate position was reached, as shown in Figure 7.

Initially the Jacobian was determined at each control step. Later, the time-consuming determination of the Jacobian was repeated only when a control step did not cause the magnitude of the image distance vector, $\underline{d}$ (Figure 1), to decrease. Figure 7 shows that this shortcut that significantly improves the speed of operation is, indeed, permissible.

6. Conclusions and Outlook

We have presented a method, based on the sensor-control Jacobian matrix introduced here, for controlling the motions of calibration-free robots. It does not depend on quantitative models and calibrations and is, thus, robust with respect to changes of parameters of the robot or its vision system.

The feasibility of the method was proven in real-world grasping experiments with a calibration-free vision-guided robot where the cameras partially move with the manipulator. Objects have been grasped successfully in almost all cases with between 2 and 5 iteration steps executed.

The robustness of the system may be further improved by utilizing the redundancy inherent in the available sensor data [Maryniak, Graefe 1998]. The speed of the operation may be improved by including a long-term memory for storing once determined values of the sensor-control Jacobian, so that they may be used in later actions of the robot. We are currently working on these topics. Also, we will study whether it is possible to determine the sensor-control Jacobian without executing any test movements; this, too, would help in improving the operating speed of the robot.

Acknowledgments

We thank Mr Minh Chinh Nguyen for making his image processing software available for our use, and especially Mr Karl Vollmann for the many fruitful discussions we had.

References

Bronstein, I. N., Semendjajew, K. A.: Ergänzende Kapitel zum Taschenbuch der Mathematik. BSB B. G. Teubner, Leipzig 1970; Verlag H. Deutsch 1990.

Broyden, C. G. (1965): A Class of Methods for Solving Nonlinear Simultaneous Equations. Mathematics of Computation, v 19, pp 577-593.

Conkie, A., Chongstitvatana, P. (1990): An Uncalibrated Stereo Visual Servo System. DAI Research Paper No. 475, University of Edinburgh.

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.

Halmos, P. R. (1967): A Hilbert space Problem Book. Van Nostrand Comp. Princeton.

Hollinghurst, N., Cipolla, R. (1994): Uncalibrated stereo hand-eye coordination. Image and Vision Computing, Volume 12/3, pp 187-192.

Hosoda, K., Asada M. (1994): Versatile Visual Servoing without Knowledge of True Jacobian. IEEE/RSJ/GI International Conference on Intelligent Robots and Systems, IROS '94, pp 186-193.

Hutson, V. C. L., Pym, J. S. (1980): Applications of Functional Analysis and Operator Theory. Academic Press, London.

Jägersand, M., Nelson R. (1995): Visual Space Task Specification, Planning and Control. IEEE International Symposium on Computer Vision, pp 521-526.

Ljusternik, L. A., Sobolew, W. I. (1975): Elemente der Funktionalanalysis. Akademie-Verlag, 4th edition 1968; reprint Verlag H. Deutsch 1975.

Maryniak, A., Graefe, V. (1998): Utilizing Sensor Data Redundancy to Gain Robustness in the Control of Calibration-Free Robots. IEEE International Conference on Industrial Electronics, IECON ’98. Aachen, Sept. 98.

Vollmann, K., Nguyen, M.C. (1996): Manipulator control by calibration-free stereo vision. Intelligent Robots and Computer Vision XV, Proceedings of the SPIE, Vol. 2904. Boston, pp. 218-226.

Xie, Q., Graefe, V., Vollmann, K. (1997): Using a Knowledge Base in Manipulator Control by Calibration-Free Stereo Vision. IEEE International Conference on Intelligent Processing System. Beijing, pp 1307-1311.

Yoshimi, B.H., Allen, P. K. (1994): Active, uncalibrated visual servoing. IEEE International Conference on Robotics and Automation. Vol. 4, pp 156-161.

Zeidler, E. (1996): Teubner-Taschenbuch der Mathematik. BSB B. G. Teubner, Leipzig.