3-D Measurement of Moving Objects with an Active Stereo Vision System

Atsushi Yamashita*, Shinya Matsushita*, Takeshi Sonohara*
Toru Kaneko*, Kenjiro T. Miura* and Suekichi Isogai**
*Department of Mechanical Engineering, Shizuoka University
3-5-1 Johoku, Hamamatsu-shi, Shizuoka 432-8021, Japan
E-mail: yamashita@ieee.org
**Division of System Development, Denkosha

In this paper, we propose a new camera calibration and 3-D measurement method with an active stereo vision system for handling moving objects. To gain the extrinsic camera parameters in real time, a baseline stereo camera model and a projective transformation of stereo images are utilized by considering the epipolar constraints. To make use of 3-D measurement results of the moving object, the manipulator hand approaches the object. When the manipulator hand and the object are near enough for them to be in a single image, a very accurate camera calibration can be executed to calculate the manipulator size in the image. Our method does not need complicated image processing and can measure 3-D position and orientation of the object fast.

Keywords: Stereo Vision, Camera Calibration, 3-D Measurement, Epipolar Constraints
Research Areas: Computer Vision, Robotics, Image Processing

1. Introduction

Robot vision is very important to handle moving objects with a robot manipulator. In general object-handling tasks, at first the system recognizes the type of an object, calculates its 3-D position and orientation, decides the grasping points on its surface, and then the robot hand approaches it while measuring 3-D position and orientation. There are a lot of researches about the object-handling task with a camera and manipulator system. These studies asserted that the camera calibration was exactly done and if there existed some error about the calibration, the handling tasks could not be executed because the system recognized incorrect 3-D positions and orientations of the objects.

To measure the exact position and orientation of the objects with a stereo vision system, the camera calibration is very important. The calculation technique of a projective camera matrix mentioned in [1] is the most basic way. In [2], a fast calibration technique for a hand-eye system was proposed. However this technique cannot be applied to the moving object-tracking task. A self-calibration method for a moving camera have been proposed [3]. However these methods are generally weak for the reading errors of corresponding points and other image noise. Therefore a calibration method based on statistical model was proposed [4]. This method can calculate 3-D position and orientation precisely, but it is not suitable for real-time tasks because its computational cost is large.

In this paper, we propose a 3-D measurement method with an active stereo vision system for handling moving objects (Fig.1). Each camera can change its direction about the Y-axis to follow the moving objects because the field of view of the camera is not wide enough to see the movable range of the objects. Our real-time camera calibration method is developed for the active camera system.

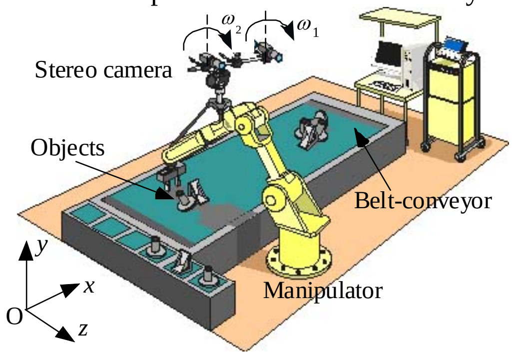


Fig.1 The manipulator handles moving objects.

2. Active Stereo Vision System

We propose a fast calibration method that utilizes the projective transformation and the epipolar constraints of the baseline stereo (parallel stereo) camera model. The epipolar constraints about the transformed pairs of images are considered. Moreover, the extremely exact 3-D measurement is not always required in object grasping tasks. When the cameras track the moving object, each camera never loses sight of the object by changing its direction to see the object on the center of its image. When the manipulator hand grasps it, high accuracy is needed. In that case, additional information can be obtained when the hand and the object is in the same image. Therefore, we propose a calibration technique that efficiently utilizes the manipulator's size in the images and relative relationship between the manipulator hand and the object.

Each camera changes its direction about the Y-axis to capture the image in which the moving object always locates on the center of each image plane. The characteristic points of the object's silhouette on two images are extracted and then corresponding points are searched. After that, the projective transformation that generates new pairs of baseline stereo camera images is calculated by using the evaluation functions. From these new pairs of images, the 3-D position and orientation of the object is measured and the manipulator moves forward to it. When the manipulator and the object are near enough to shot them in the same images, the more accurate calibration is performed to use additional information of the manipulator's size.

The axis of rotation must be set perpendicular to the X-Z plane (ground). But even when the great efforts are done about the camera system setting toward the world coordinate, the position and the direction of the rotation shafts necessarily deviates a little. Therefore, a deviation grows large according to the directional change of the active camera to follow the moving object. Unknown parameters of the active cameras increase because we must calculate the direction of each rotation shaft in addition to their initial relative position and initial direction between two cameras. This implies that an accurate calibration of each camera parameters becomes difficult.

In this paper, a two-step calibration technique is proposed. In the first step, the rotation angles of the projective transformation about the X-, Y-, and Z-axes to make baseline stereo camera images are determined. The angles about the X- and Z-axes, and the change of baseline length are expressed as approximate functions of the change of the rotation angle about the Y-axis. This calibration is carried out before the grasping task. In the second step, the condition when the Y coordinate value of the corresponding points after transformation becomes equal is searched in real-time by using the evaluation functions. As for your being careful here, the real motor angle and the angle for the projective transformation are necessarily not the same. The reason why the original images are changed into those of the baseline stereo vision by using the projective transformation is to calibrate the camera parameters and to measure 3-D position faster compared with the method using the projective or the fundamental matrices. When the manipulator and the object are in the same image, more precise calibration is performed by using the manipulator's size and the epipolar constraints.

3. Calibration and 3-D Measurement

3.1 Pripciple of 3-D Measurement

The angles for the projective transformation for the baseline stereo vision images are searched with the epipolar constraints that the epipolar lines are horizontal, and the 3-D measurement is carried out from the information of corresponding points.

The pairs of baseline stereo images are generated from the ordinary images with the projective transformation about the X-, Y-, and Z-axes (Fig.2).

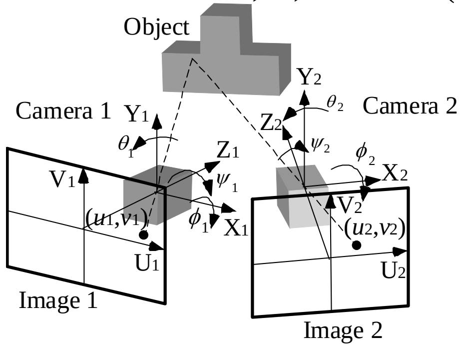


Fig.2 Image planes and coordinate systems.

The first projective transformation is about the Z-axis, secondly about the X-axis, and then about the Y-axis. The point (u_i, v_i) on each image (i=1: left camera image, i=2: right camera image) is projected to (x_i, y_i) with the projective transformation. Note that ψ_i, ϕ_i, θ_i are the angles of rotation about the Z_i-, X_i-, Y_i-axis. If ψ_i, ϕ_i and θ_i are known, the 3-D position of the corresponding point on each image plane is calculated as follows with the baseline stereo camera model (Eq.(1)-(3)).

$$X = \frac{b(x_1 + x_2)}{2d} \tag{1}$$

$$Y = \frac{b(y_1 + y_2)}{2d} \tag{2}$$

$$Z = \frac{bf}{d} \tag{3}$$

where b is the length of baseline, f is the image distance, and $d = x_1 - x_2$ is the disparity.

Ideally, each motor angle ω_i is equal to the angle θ_i for the projective transformation, and ψ_i and ϕ_i are always equal to 0. However there is a setting error of each motor's shaft and hence ω_i is not equal to θ_i. Hence, we use ψ_i, ϕ_i and θ_i for the projective transformation.

3.2 Evaluation of Projective Transformation

The epipolar lines of the baseline stereo camera images are horizontal. Therefore, two images are close to the baseline stereo vision images when the evaluation function E_{epi} has a small value (Eq.(4)).

$$E_{epi} = \frac{1}{N}\sum_{i=1}^{N}(y_{1,i} - y_{2,i})^2 \tag{4}$$

where $y_{1,i}$ and $y_{2,i}$ are the Y coordinate value of the corresponding points, and N is a number of the corresponding points.

Ideally, an good 3-D measurement can be done to use the angles ψ_i, ϕ_i, θ_i that minimize E_{epi}. However, a good calibration never be done with the influence of image noise. Therefore, we propose a two-step calibration method that considers not only the epipolar constraints but also the length information.

3.3 Calibration in Advance

In the fist calibration step, each camera's feature is measured by using a box whose surface is painted with the planner pattern (Fig.3).

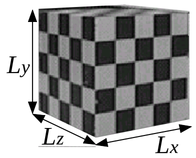


Fig.3 Box for calibration.

The calibration is executed by changing each camera angle ω_i and the position of the box along Z-axis. The surface of the box is set perpendicular to the optical axes of two cameras in the initial angle (ω_i = 0). At first, the image distance *f* and the baseline length b_0 in the initial angle is calculated by changing the distance between the box and the camera system. After that, the left camera angle ω_1 is changed discretely and the pairs of images are gained while ω_1 is always set 0. The discrete value of the angle change is constant. The right camera angle ω_2 is changed in the same way while $\omega_1 = 0$. The same procedure is repeated by changing the distance between the box and the camera system. To minimize the disparity of length error between 3 axes, E_{dis} is introduced (Eq.(5)).

$$E_{dis} = \frac{1}{3}\{(l_x - l_y)^2 + (l_y - l_z)^2 + (l_z - l_x)^2\} \quad (5)$$

where l_x, l_y, and l_z are the length of the box's edge along X, Y, Z-axis from the result of 3-D reconstruction of the box shape.

An accurate calibration can be done to use the angles ψ_i, ϕ_i, θ_i that minimize E_1 (Eq.(6)).

$$E_1 = \alpha_1 E_{epi} + \beta_1 E_{dis} \quad (6)$$

where α_1 and β_1 are weight coefficients.

When ω_2 is settled 0 and ω_1 is changed, the optimal θ_1, ψ_1 and ϕ_1 are calculated discretely according to the ω_1 value to minimize E_1. The change of baseline length Δb_1 is calculated in Eq.(7). When ω_1 is settled 0, the optimal θ_2, ψ_2, ϕ_2, and Δb_2 is calculated discretely, too. The baseline length is calculated in Eq.(8).

$$\Delta b_i = \left(\frac{L_x + L_y + L_z}{l_x + l_y + l_z} - 1 \right) b_0 \quad (7)$$

$$b = b_0 + \Delta b_1 + \Delta b_2 \quad (8)$$

where L_x, L_y, L_z are the real length of the box, and l_x, l_y, l_z are measured length.

The discrete values of θ_i, ψ_i, ϕ_i, and Δb_i are fitted on approximate functions. To find optimal angle for the projective transformation fast in the second-step real-time calibration, ψ_i, ϕ_i, and Δb_i are expressed as the functions of θ_i (Eq.(9)-(11)).

$$\psi_i = f_i(\theta_i) \quad (9)$$

$$\phi_i = g_i(\theta_i) \quad (10)$$

$$\Delta b_i = h_i(\theta_i) \quad (11)$$

3.4 Calibration in Real Time

To do a more accurate calibration, the length information is utilized. The marks are attached to the manipulator hand and the distance between two marks are known (Fig.4(b),(c)). The evaluation function of the difference between the real length and the measured length E_{len} is set as Eq.(12). To search θ_i that minimizes E_2, the real-time 3-D measurement is executed. The computation costs of the optimal conditions become small because ψ_i, ϕ_i, and Δb_i are expressed as function of θ_i, and all that the system must do is to decide θ_i that minimizes E_2 (Eq.(13)).

$$E_{len} = \frac{1}{M}\sum_{i=1}^{M}(L_i - l_i)^2 \quad (12)$$

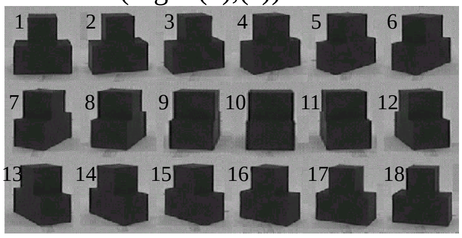
$$E_2 = \alpha_2 E_{epi} + \beta_2 E_{len} \quad (13)$$

where L_i is the real distance between two marks, l_i is the calculated distance by image processing, M is a number of mark distances, and α_2 and β_2 are weight coefficients.

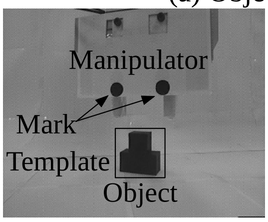
In the early time when each camera tracks the moving object, the manipulator is not in the images. In that case, the information about the mark distance cannot be used and β_2 is set 0. The manipulator is guided close to the moving object by the result of 3-D measurement only considering E_{epi}. When the manipulator and the object are in the same image, β_2 becomes a positive number. The accuracy of 3-D measurement becomes good and the relative distance between the manipulator and the object in the camera coordinate is utilized to control the position and the orientation of manipulator hand at the same time.

3.5 Object Recognition

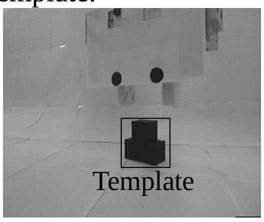
We use a template matching technique for detecting the types of the objects and the corresponding points between two images. The sequential similarity detection algorithm (SSDA) is adopted as the template matching because it can work fast. There are two processes in our template matching using the object template and the point template respectively. At first, the type of the object and the rough orientation recognition is done by using the object templates (Fig.4(a)). In each object template, the type and the orientation of the object change discretely. The most fitting template is selected with SSDA by searching the center position of each camera image. For example, object templete 17 is selected in the left image (Fig.4(b)) and object templete 15 is selected in the right image (Fig.4(c)). Secondly, exact corresponding points are searched with the point templates of each object template. Each point template has small template ranges in which characteristic area of object's silhouette exists (Fig.4 (d),(e)).



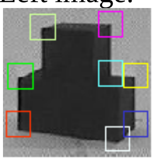
(a) Object template.

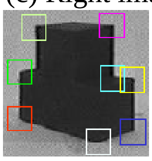


(b) Left image.



(c) Right image.





(d) Point template (Left). (e) Point template (Right).

Fig.4 Template matching.

4. Experiments

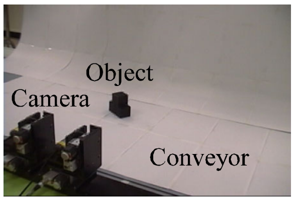
The results of 3-D reconstruction are shown in Table 1. In Table 1, X_e, Y_e, Z_e indicate the average length error along each axis and η_e indicates the average angle error between two axes. The real length and angle are 60mm and 90deg. In Strategy 1, the 3-D measurement is done by only using motor angle ω_i for the projective transformation. In this case, $\omega_i = \theta_i$ and ψ_i, ϕ_i, and Δb_i are always equal to 0. This means that the calibration is not done in this strategy. In Strategy 2, the 3-D measurement is done by using Eq.(13) under condition when $\alpha_2 = 1$ and $\beta_2 = 0$. It is the case where the manipulator and the object are not in the same image. In Strategy 3, $\alpha_2 = 1$ and $\beta_2 = 1$. It is the case where they are in the same image and the length information is utilized for the calibration.

The system often failed the handling task when Strategy 1 was adopted. The main reason why the manipulator failed to grasp the object was that the accuracy of 3-D measurement was not high and the manipulator could not approach the object appropriately. To adopt Strategy 2 when the manipulator and the object are not in the same image and adopt Strategy 3 when they are in the same image, the manipulator could smoothly approach the object and grasp it (Fig.5).

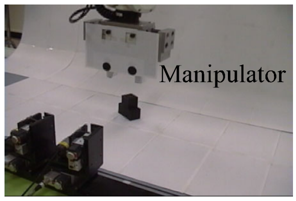
From these experimental results, our proposed method is for practical use when the manipulator handles the moving objects in real time.

Table 1 The results of 3-D reconstruction.

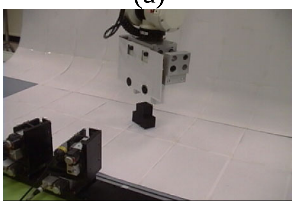
	X_e	Y_e	Z_e	η_e
1	8.9mm (14.8%)	6.4mm (10.7%)	9.7mm (16.1%)	2.9deg (3.2%)
2	3.8mm (6.3%)	1.4mm (2.3%)	3.4mm (5.7%)	2.4deg (2.7%)
3	0.1mm (0.2%)	0.8mm (1.3%)	1.8mm (3.1%)	0.7deg (1.0%)



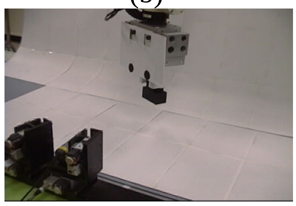
(a)



(b)



(c)



(d)

Fig.5 Experimental results.

5. Conclusions

In this paper, we have proposed a new calibration and 3-D measurement method for handling moving objects. By using the projective transformation and expressing the relationship between angles for the transformation as the approximate functions, the fast calibration can be done. Experimental results show that 3-D measurement has high accuracy.

References

[1] R.Y.Tsai: "A Versatile Camera Calibration Technique for High-Accuracy 3D Machine Vision Metrology Using Off-the-Shelf TV Cameras and Lenses," IEEE Journal of Robotics and Automation, Vol.RA-3, No.4, pp.323-344, 1987.

[2] R.Y.Tsai and R.K.Lenz: "A New Technique for Fully Autonomous and Efficient 3D Robotics Hand/Eye Calibration," IEEE Trans. on Robotics and Automation, Vol.5, No.3, pp.345-358, 1989.

[3] S.J.Maybank and O.D.Faugeras: "A Theory of Self-Calibration of a Moving Camera," Int. Journal of Computer Vision, Vol.8, No.2, pp.123-151, 1992.

[4] K.Kanatani: "Gauge-Based Reliability Analysis of 3-D Reconstruction from Two Uncalibrated Perspective Views," Proc. 15th Int. Conf. on Pattern Recognition, pp.76-79, 2000.