능동 양안시 장치의 보정

Calibration of Active Binocular Vision Systems

도용태, 대구대학교 정보통신공학부
Yongtae Do, School of Computer and Communication Engineering, Taegu University

ABSTRACT

로봇의 지능적 작업을 위해서는 3 차원 공간에 대한 감각기능이 필수적이며, 이러한 목적으로 시각장치가 사용되었을 경우 보정은 여타 절차에 선행하게 된다. 실제 보정 과정중 중요한 것은 제어점들과 그들의 영상점을 획득하는 일과 이를 이용한 광학적, 기하학적 카메라의 파라메터 결정에 있다. 본 논문에서는 빔프로젝터와 컴퓨터에 의해 제어되는 양안시 장치를 활용하여 많은 수의 3 차원 제어점들과 이들의 영상좌표값들을 간편하게 획득하는 방법을 서술한다. 또, 위치가 고정된 카메라의 경우와 달리 능동 카메라 장치는 그 파라메터의 일부가 변수가 되는데, 이 경우에 유용한 선형의 보정 기법을 제안한다.

I. Introduction

Obtaining accurate three dimensional (3D) positional information is of great importance in many robotic applications. An effective approach is to use stereo cameras to infer the 3D position from stereo images. For this purpose, two critical problems must be solved: *camera calibration* and the *stereo feature matching* (*correspondence problem*). While much of the research has been focused on the latter problem, camera calibration is a prerequisite of most vision-related computational problems including stereo matching [1].

The camera calibration problem consists of two sub-problems: *data collection* and *camera modeling*. Training and testing data samples with known world coordinates must be identified to be used for calibration and verification purposes. A popular method to collect data is to use a flat panel printed with calibration patterns (grids). By positioning this panel in front of camera using precise equipment like robot, accurate world coordinates can be obtained. For example, in Tsai [2], the micrometers are used for positioning with a resolution of 0.00254mm. Martins *et al.* [3] used a high precision robot with three linear axes, of which each could be positioned within 0.00254mm.

The accuracy of the calibration pattern is another important factor. For example, Weng *et al.*[4] used diamond-shaped patterns printed on an ultra flat optical glass by means of a high-precision photographic process that kept the positional error of the pattern within 0.05mm.

The locations of image points at stereo image planes corresponding to the pattern in 3D space must also be determined accurately. In [4], these points are determined as the intersection of the edges and are claimed to achieve sub-pixel accuracy.

Finally, the paring of 3D points and corresponding image points was done manually. Developing and utilizing such high-precision calibration equipment, however, is very costly and may not be practical in many applications.

In this paper, the data collection procedures are automated using a pair of computer-controlled pan-tilt cameras and a LCD video projector [5]. Red and green blobs are projected onto the wall by the projector with their centers at pre-defined world coordinates. Two cameras mounted on a mobile robot are automatically controlled by a computer to aim their optical centers at the red blob, or *fixate*. The image coordinates of the center of the green blob, which is not on the optical center ray of both cameras but near the red blobs, are recorded in addition to the pan and tilt angles of these cameras. The process is automated and a large number of data can be collected easily.

Camera modeling is directly related to the calibration and 3D reconstruction processes. Most existing camera calibration techniques are based on the pin-hole camera model [6] that describes a linear mapping between 3D points and corresponding 2D image points. Parameters to be calibrated are those related to the geometric position (extrinsic parameters) and optical features (intrinsic parameters) of a camera. If the viewing angles of a camera are controllable, the extrinsic parameters can be divided further into the fixed and variable parts. In this paper, a linear method to actively determine the fixed extrinsic parameters in addition to the intrinsic parameters is described. When a camera is mobile rather than being fixed, the analysis of factors contributing to calibration errors is even more complicated. Mechanical errors may be more significant than optical non-linearity of the vision system.

II. Systems for Data Collection

The system consists of two Directed Perception pan-tilt units with color Sony XC-999 CCD cameras and two Meteor frame-grabbers. The cameras and pan-tilt units are mounted on an RWI-B21 mobile robot that contains an Intel Pentium Pro-200 processor and an Intel Celeron 533 processor. A video LCD projector with 960 pixels by 720 pixels resolution is hung from the ceiling in the back of the robot (the LCD projector is in place of a monitor on the computer running the display software). The distance between the projector and the wall is about 4 meters. The image of each pixel on the wall has a size about 4 mm^2.

The calibration patterns consist of red and green circular blobs projected by the LCD projector to specific locations on the wall. The centroid of each blob can be estimated accurately up to sub-pixel accuracy. At each fixed position of the robot, a series of projected patterns were presented to the robot (we refer to such a series as a "sheet"). Within each sheet, the red blobs were presented at 160 (10 vertical, 16 horizontal) regularly-spaced positions with 40 pixels increments in each direction.

The green blob position was defined relative to the red blob. For each red blob position on the sheet, the green blob had 12 possible positions relative to the red blob. There are a total of 1920 (160 × 12) different combinations within each sheet. A sheet was presented every three centimeters over 15 centimeter total displacement of the robot.

After each sheet was completed the robot was moved 3 centimeters. The position of the robot was re-measured manually after every movement due to poor position control of the robot. The measurement was done using two tape measures and two plumb weights. This allowed for accurate positioning of the robot and control of the rotation of the robot turret. This process was repeated over a 15-centimeter total distance of the robot.

The error in positioning of the blobs and robot was minimized as much as possible but could not be eliminated entirely. Obvious sources of errors include:

- Assumed planarity of the wall
- Assumed fixed lateral robot position
- Distance between the wall and robot
- Angular resolution of PTU
- Image location of the blob centroids

III. System Calibration

To compute the 3D position of an arbitrary point from corresponding stereo image points, traditional methods dictate that the system parameters should be computed first. There are many calibration techniques available to do this. Most of them determine the linear transformation matrix first and then correct the errors due to the system non-linearity assuming that lens distortion is the major cause of non-linearity. This approach, however, is not effective to the mobile vision systems of this paper because the transformation matrix is not fixed but variable and lens distortion is no longer the most significant error source. We use a linear method assuming the pin-hole camera model. The transformation matrix is divided into intrinsic, variable extrinsic, and fixed extrinsic parts. The fixed extrinsic part is determined from the data collected by actively controlling a camera. The intrinsic part is then obtained using the fixed extrinsic part and variable extrinsic part set by pan-tilt angles of a camera.

When a pin-hole camera model is used, a 3D point at (x_w, y_w, z_w) in world coordinate system $\{W\}$ can be transformed into the corresponding 2D point at (u,v) on the image plane by

$$\begin{bmatrix} su \\ sv \\ s \end{bmatrix} = [INT][EXT] \begin{bmatrix} xw \\ yw \\ zw \\ 1 \end{bmatrix} \tag{1}$$

where s is a scale factor; $[INT]$ is a (3×4) matrix describing the perspective transformation definable by intrinsic camera parameters, (u_o, v_o) and (k_1, k_2), the center of image plane and scaling factors respectively; $[EXT]$ is a (4×4) homogeneous transformation matrix describing the geometric relation between the camera coordinate system $\{C\}$ with the world coordinate system $\{W\}$;

$$[INT] = \begin{bmatrix} k_1 & 0 & u_o & 0 \\ 0 & k_2 & v_o & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \tag{2.1}$$

$$[EXT] = \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_x \\ r_{21} & r_{22} & r_{23} & t_y \\ r_{31} & r_{32} & r_{33} & t_z \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{2.2}$$

For the pan-tilt camera system, the [*EXT*] can be considered as a multiplication of two transformation matrices like

$$[EXT] = [V][F] \tag{3}$$

where [*F*] describes the homogeneous transformation between {*W*} and *the home of {C}*, on which the pan and tilt angles of a camera can be defined resulting a rotation matrix [*V*]. From the definition it is clear that elements of [*F*] are fixed while those of [*V*] are varying when a camera is rotated to fixate a target point on a sheet. Figure 1 shows the transformations between frames.

When a camera is fixated to a 3D point, the corresponding image point lies at the center of the image plane. This means that both x_c and y_c coordinates of the point in {*C*} are zero yielding the following equations

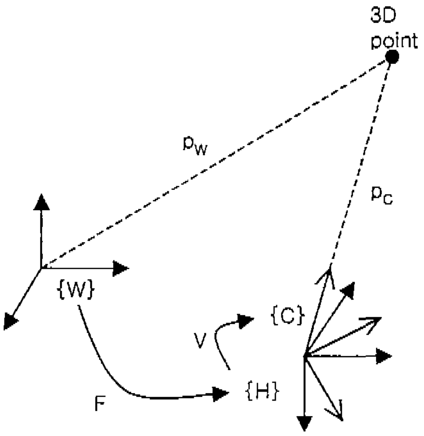


Figure 1. Transformations between coordinate systems.

$$\begin{aligned} &v_{11}(f_{11}x_w + f_{12}y_w + f_{13}z_w + f_{14}) + \\ &v_{12}(f_{21}x_w + f_{22}y_w + f_{23}z_w + f_{24}) + \\ &v_{13}(f_{31}x_w + f_{32}y_w + f_{33}z_w + f_{34}) = 0 \end{aligned} \tag{4.1}$$

$$\begin{aligned} &v_{21}(f_{11}x_w + f_{12}y_w + f_{13}z_w + f_{14}) + \\ &v_{22}(f_{21}x_w + f_{22}y_w + f_{23}z_w + f_{24}) + \\ &v_{23}(f_{31}x_w + f_{32}y_w + f_{33}z_w + f_{34}) = 0 \end{aligned} \tag{4.2}$$

where $\{v_{11}, v_{12}, \ldots, v_{33}\}$ are elements of rotation matrix [*V*] definable from pan and tilt angles of the camera. The unknown twelve elements of matrix [*F*], $\{f_{11}, f_{12}, \ldots, f_{34}\}$, can be determined using more than six control points fixated by least error squares.

Once the fixed part of the [*EXT*] is determined, a 3D point can be represented in {*C*} with any given pan and tilt angles. Then, [*INT*] can be determined from

$$\begin{aligned} uz_c &= k_1 x_c + u_o z_c \\ vz_c &= k_2 y_c + v_o z_c \end{aligned} \tag{5}$$

with more than two 3D points and their image points. Since we use the centroids of red blobs and those of green blobs as the targets to be fixated and their neighbors respectively, data from red blobs are used for determining [*F*] while data from green blobs are used for determining [*INT*].

Calibration and 3D reconstruction techniques based on ideal pinhole camera model including the technique described in this section are fast and simple, which are required in many robotic applications, because only linear equations in closed-form are needed to be solved. On the other hand, the disadvantage of the methods is also clear; they may not be accurate because the non-linear terms are ignored. In most existing techniques, the major non-linear term is lens distortion, which is a function of image coordinates. In our system, however, other terms such as camera fixation error and robot location error seem like more significant as shown in experimental results in section 4.

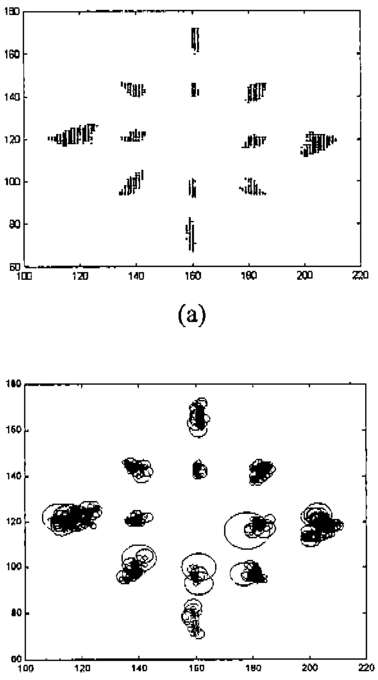
IV. Experimental Setup and Results

In this experiment, we deliberately use large amount of data samples so that effectiveness of the proposed method can be analyzed with less affect of statistical sampling error.

Five sheets of data samples were collected. The red blob data samples (samples on the center ray) were processed separately from green blob samples (sample off the center ray). The data from each sheet was then put into one of two categories. The data from first(z_w=162.20cm), third(z_w=168.15cm),, and fifth(z_w=174.00cm) sheets was used as calibration data, and the data from the second(z_w=165.10cm), and fourth(z_w=171.00cm), sheets was used for testing.

To find the home position of the a camera, the elements of [*F*] matrix, 480 red blobs on three calibration sheets are used. Then, all of the green blobs available are used to determine [*INT*]. Table 1 shows the result of using the linear technique described in section 3 for 3D reconstruction. As shown in the table, the error in z_w direction is much larger than those in other directions and this can be easily expected considering only three z_w values are used for calibration.

When the projection error of a camera is checked on image plane after the linear calibration, the relation between the position of an image point and the magnitude of its error is not significant as shown in figure 2 unlike the assumption of lens distortion, which is dependent on the image position, made in many existing camera calibration techniques. This may be explained with the two facts: (i) the measurement errors associated with the system such as camera rotation angles and robot movement are larger than that by lens distortion, (ii) since only the center portion of the image plane was used for the experiment by fixating the red blob at the center and locating green blobs near the red blob, the lens distortion can not be large naturally.



(a)

(b)

Figure 2. (a) The distribution of centroid images of green blobs on right camera's image plane, (b) the mean squared error circle for each image location corresponding to (a)

V. Conclusions

red	σ	0.84	0.72	3.12
	μ	0.58	0.49	2.43
green	σ	0.89	0.89	3.25
	μ	0.62	0.58	2.53

Table 1. Experimental results of calibration of the robotic vision system, where μ and σ refer to the mean and standard deviation of the error.

A linear calibration technique for an active binocular vision system is proposed with data collected automatically. The technique performed well in the x and y direction with a sigma and average error within a centimeter. An error of this magnitude may be acceptable in many cases with data points in the range between 1.6 and 1.8 meters away. Since we can not assume the lens distortion to be a major source of measurement error, the correction methods of most camera calibration techniques available are not applicable here.

Acknowledgement

The robot vision systems at the mechanical engineering department, University of Wisconsin at Madison, were employed for this paper.

References

[1] O.D.Faugeras and G.Toscani, "The calibration problem for stereoscopic vision," NATO ASI Series, Vol.52, Sensor Devices and Systems for Robotics, Eds. A.Casals, Springer-Verlag, Berlin, 1989.

[2] R.Y.Tsai, "A versatile camera calibration technique for high accuracy 3d machine vision metrology using off-the-shelf TV cameras and lenses," IEEE J.Robotics & Automation, Vol. RA-3, No.4, pp.323-344, 1987.

[3] H.A.Martins, J.R.Birk, and R.B.Kelley, "Camera models based on data from two calibration planes," Computer Graphics and Image Processing, Vol.17, pp.173-180, 1981.

[4] J.Weng, P.Cohen, and M.Herniou, "Camera calibration with distortion models and accuracy evaluation," IEEE Trans. Pattern Analysis and Machine Intelligence, Vol.14, No.10, pp.965-980, 1992.

[5] N.Guse, A neural network for visual kinematics, M.Sc. Thesis, Department of Mechanical Engineering, University of Wisconsin, Madison, 1999

[6] Y.Yakimovsky and R.Cunningham, "A system for extracting three-dimensional measurements from a stereo pair of TV cameras," Computer Graphics and Image Processing, Vol.7, pp.195-210, 1978.