

Camera Calibration by Ray Tracing

Chiaki AOYAMA*

ABSTRACT

A highly accurate method of camera calibration which works by tracing the light rays incident on the camera has been developed. In this method, the relative position of the point light source is adjusted to position the point light source image in the center of each pixel. Afterwards, the relative distance to the point light source is changed and the relative position of the point light source is adjusted again. A three-dimensional locus of the incident rays to the two point light sources obtained is then estimated from the relative positions. This procedure can be used to determine the relative positions of multiple cameras. The accuracy of the method was verified by calibrating a stereo camera and estimating the three-dimensional position of the point light source. The camera calibration method displayed identical accuracy for an image across an irregularly warped shield, which is conventionally a difficult condition for calibration, as when no shield was present. In addition, it was confirmed that there was deviation from the pinhole camera model (no crossing of incident rays).

1. Introduction

The ability to recognize the environment is necessary for autonomous robots to perform actions in response to a changing external environment, and cameras mounted in robots are used as image sensors for this purpose⁽¹⁾⁻⁽⁴⁾. If relative three-dimensional (3D) positions can be accurately estimated, robots can not only move to avoid obstacles, but the recognition ability can also be applied to actions that involve manipulation of the environment, such as grasping objects.

Stereo vision based on the difference in the appearance of images taken with multiple cameras, as shown in Fig. 1, is often used to estimate relative 3D positions between the cameras and the environment.

Binocular stereo vision can be used to calculate the relative 3D position to the target using Eq. (1) if the incident ray from the corresponding position on the target to each camera is known.

For example, the procedure for estimating the relative 3D

position (P_x, P_y, P_z) of a headlight in Fig. 1 is as follows. First, the headlight position in the left image that corresponds to the same headlight in the right image is obtained. Next, the incident ray parameters ($\gamma_i, \alpha_i, \alpha_2, x_i, y_i, z_i, x_2, y_2$ and z_2 in Fig. 2) that correspond to the two-dimensional (2D) image position (pixel position) of the headlights in each image are derived from values measured beforehand. Finally, these parameters are applied to Eq. (1) to estimate the relative position of the headlight.



Fig. 1 Stereo images

* Wako Research Center (HGF)

$$\begin{aligned}
 P_x &= \frac{x_1 \tan \alpha_1 - y_1 - x_2 \tan \alpha_2 + y_2}{\tan \alpha_1 - \tan \alpha_2} \\
 P_y &= (P_x - x_1) \tan \alpha_1 + y_1 \\
 P_z &= (P_x - x_1) \tan \gamma + z_1
 \end{aligned} \tag{1}$$

In other words, estimation of relative position using stereo vision is divided into two processes. The first process is to estimate the pixel positions that show the same location between images taken from different positions; that is to say to find the correspondence relationship. The second process is to establish the correspondence between the images taken by the cameras and the real space. Various methods have been proposed to determine these two types of correspondences^{(5) - (10)}.

This paper describes newly developed camera calibration method to establish the correspondence relationship between images and the real space. The newly developed method increases calibration accuracy by tracing the rays incident on the lens. The method is presented below, together with the verification results.

2. Issues with Conventional Methods

The intensity value of a pixel in an image taken by a camera is proportional to the strength of the ray incident from a certain point on the target and traveled through space to reach the image sensor of the camera. Conversely, the intensity value can be said to be the brightness of the area struck by a straight line extending from the camera reference position if the ray is traced backwards from the image sensor. For example, the image of the robot appearing on the screen corresponds to the brightness of the rays from each position on the robot that pass through the pinhole and illuminate the screen, when using a pinhole camera model as shown in Fig. 3. That is to say, there is a point light source in the 3D space that corresponds to each pixel.

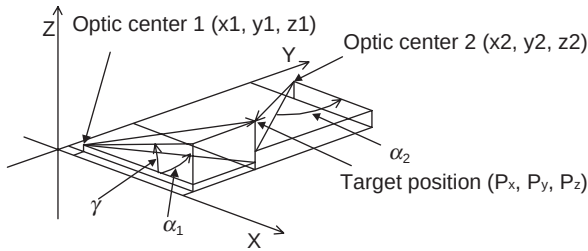


Fig. 2 Coordinate system

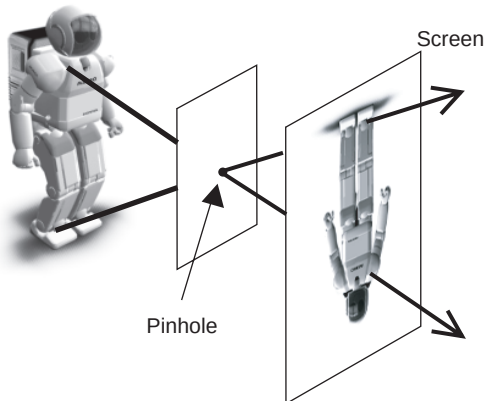


Fig. 3 Pinhole camera model

Camera calibration is the process of establishing the relationship between the positions on this screen, and the path of the light rays passing through the 3D space. The issue of calibration is to estimate these paths.

In a real camera, an image is created by the light passing through the lens assembly. Using a lens normally produces the phenomenon of image distortion whereby straight lines appear as curved lines in the image on a screen. Some types of image processing require an image for which this distortion has been corrected. In these cases as well, distorted image can be corrected to a distortion-free image if the relationship between the positions on screen and the rays is known.

Conventional camera calibration methods estimate the relationship of all pixels by taking a pattern of known size and shape, extracting the positions of several dozen locations with distinguishing features from the image, and then applying a polynomial fit to the relationship between the image position and the actual spatial position. It has been assumed thus far that distortion-free images corrected by the polynomial fit are obtained by the same type of projection as images obtained by pinhole cameras.

In addition, the cameras used by robots must sometimes use a shield in consideration of weather resistance or design factors. These shields are a factor producing irregular and significant image distortion such as that shown in Fig. 4. In these cases, the image distortion cannot be adequately corrected by polynomial-based function approximation, making it impossible to obtain sufficient position estimation accuracy.

3. Calibration Principle and Method

In order to estimate the 3D path of incident rays, this research utilized the principle that if any two points in the space through which a ray passes can be determined, the straight line connecting those two points is the 3D path of that incident ray. Strictly speaking, the ray incident on the lens is an expanding luminous flux, so the path through which the ray passes is not a line. However, in consideration of the fact that the luminous flux is narrowed to an extreme by reducing the aperture of the lens, the luminous flux becomes a single straight line called a principal ray. Here, the principal ray can be thought of as an incident luminous flux that has become a straight line. In addition, as shown in Fig. 2, the path of the ray from the target to the lens is important for image-based 3D measurements, but the internal path of the ray between the lens and the image plane is fixed. Therefore, only the principal ray before entering the lens ("pre-incident principal ray") is taken into account.

Initially, a simple method for estimating the 3D path of the pre-incident principal ray was considered. Estimation of a ray trajectory utilizes the phenomenon where the position of a light source in the image does not change when the point light source is moved along the path of the ray. First, a point light source is positioned within a 3D space so that the image of the point light source appears in the



Fig. 4 Example of distorted image

center of a certain pixel, assuming the camera is fixed horizontally. Next, the point light source is moved farther away while maintaining a constant image position on the pixel. The straight line connecting the initial and subsequent positions of the point light source is the path of the pre-incident principal ray. The paths of the pre-incident principal rays for all pixels can be estimated using the procedure above for the entire screen.

Figure 5 shows a 2D representation of this procedure.

However, moving the point light source for the entire screen becomes a challenging issue when this procedure is used for a camera with a wide angle of view, since the point light source must be moved through an enormous space, which makes it impractical.

3.1. Measurement Principle

Instead of moving the point light source, this research used the method of changing the relative 3D position of the point light source by placing the camera on a rotation stage as shown in Fig. 6, and rotating the stage so that the point light source appears at the center of each pixel in the image. The method shown in Fig. 5 moved the point light source vertically while moving it farther away. In contrast, the newly developed method produces relative vertical movement by rotating the camera. As a result, the space needed for this operation is only the narrow range connecting the distant point light source and the camera.

Figure 7 shows a 2D representation of the new method. Here, the camera is rotated so that the light sources located at different distances appear in the center of the same pixel.

In actual measurement, when the light source is moved farther away in a straight line, the measurement of continuous intermediate positions are not necessary, and only the first and last positions are sufficient. Therefore, instead of varying the distance for each pixel,

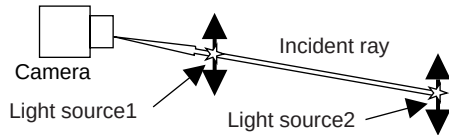


Fig. 5 Simple ray tracing

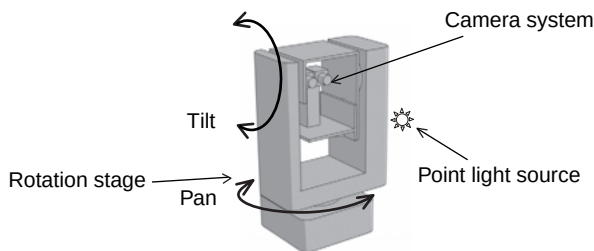


Fig. 6 Relative positions of camera, rotation stage, and point light source

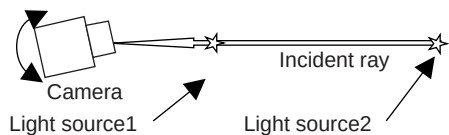


Fig. 7 Ray tracing by new method

the newly developed method obtains the relative positions for all necessary pixels with the fixed point light source, and then repeats the measurements after changing the position of the point light source.

The measurement method for a single measurement cycle is described below.

First, the camera is placed at an appropriate position on a dual axis rotation stage as shown in Fig. 6. The direction in which a light source appears when the camera is rotated by $-\theta_{\text{tilt}}$ and $-\theta_{\text{pan}}$ corresponds to the direction in which the light source appears when the camera is fixed and the light source is rotated by θ_{pan} and θ_{tilt} around the rotation stage center. (Fig. 8)

The position of the light source at this time (x_1, y_1, z_1) has the following relationship to the position of the light source before rotation $(0_{x1}, 0_{y1}, 0_{z1})$.

$$\begin{pmatrix} x_1 \\ y_1 \\ z_1 \end{pmatrix} = \begin{pmatrix} \cos \theta_{\text{tilt}} \cos \theta_{\text{pan}} & -\cos \theta_{\text{tilt}} \sin \theta_{\text{pan}} & \sin \theta_{\text{tilt}} \\ \sin \theta_{\text{pan}} & \cos \theta_{\text{pan}} & 0 \\ -\sin \theta_{\text{tilt}} \cos \theta_{\text{pan}} & \sin \theta_{\text{tilt}} \sin \theta_{\text{pan}} & \cos \theta_{\text{tilt}} \end{pmatrix} \begin{pmatrix} 0_{x1} \\ 0_{y1} \\ 0_{z1} \end{pmatrix} \quad (2)$$

This position estimation is repeated for all prescribed pixel.

Next, the light source is moved to a position with a different distance to the camera $(0_{x2}, 0_{y2}, 0_{z2})$, and the orientation of the rotation stage is adjusted so that the light source appears in the center of the pixel.

When the respective light source positions are (x_1, y_1, z_1) and (x_2, y_2, z_2) as shown in Fig. 9, the straight line connecting these

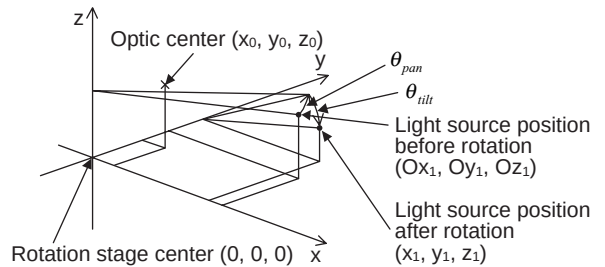
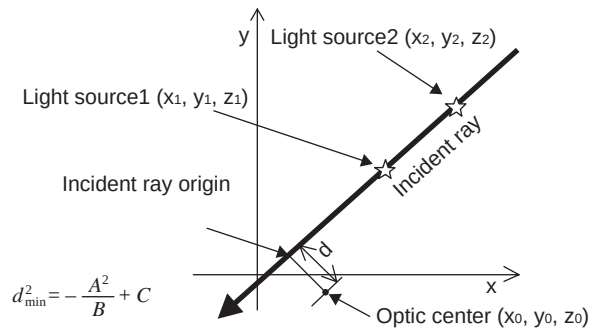


Fig. 8 Relation of positions before and after rotation of camera center and light source



$$d_{\min}^2 = -\frac{A^2}{B} + C$$

$$t_{\min} = -\frac{A}{B}$$

$$A = (x_2 - x_1)(x_1 - x_0) + (y_2 - y_1)(y_1 - y_0) + (z_2 - z_1)(z_1 - z_0)$$

$$B = (x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2$$

$$C = (x_1 - x_0)^2 + (y_1 - y_0)^2 + (z_1 - z_0)^2$$

Fig. 9 Relation between incident light and light source

two positions matches the pre-incident principal ray, and it can be expressed by Eq. (3).

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = t \begin{pmatrix} x_2 - x_1 \\ y_2 - y_1 \\ z_2 - z_1 \end{pmatrix} + \begin{pmatrix} x_1 \\ y_1 \\ z_1 \end{pmatrix} \quad (3)$$

This study measured incident rays using the above method, and the results confirmed a deviation from the pinhole camera model in that the rays do not cross at a single point in the manner described by the pinhole camera model. This made it necessary to newly define the principal point that corresponds to the pinhole. This research defined the point at which the sum of the square of the distance d between all incident rays is a minimum as the principal point. In addition, the point at which the distance d from the principal point is a minimum was defined as the base point of the pre-incident principal rays, and the rays are assumed to extend to the light source from this base point.

These definitions make it possible to think of the pre-incident principal ray in separate terms, as those of the base point position (x , y and z in Fig. 2) and the direction (α and γ in Fig. 2). Through this definition, the model is equivalent to the pinhole camera model in case all the base point positions are the same.

3.2. Relative Rotation Method of the Light Source for Each Pixel

The method of displaying the light source in the center of each pixel is described below.

The rotation stage is rotated for each pixel as shown by the processing flow in Fig. 10. The imaging conditions are such that the captured images only show the point light source; that is to say the images show a white spot on a black image. The peak position is estimated as the center of the white spot. When this peak position matches the prescribed pixel with error of constant ϵ or less, the angle of rotation of the rotation base at that time is recorded, and the algorithm moves on to the next pixel. When there is an offset, the offset amount is converted into the corresponding angular value and the stage is rotated by that amount.

The peak position is estimated from the strengths of the adjacent pixels in the top, bottom, left and right directions. When measurements for all prescribed pixels have been finished, the position of the point light source is changed and the process is repeated. After measurement has been performed at different distances, the parameters of the pre-incident principal ray are calculated by the method described in section 3.1, and these results are used as the camera calibration data. When multiple cameras

are present, the relative relationship between the cameras is also determined at the same time.

4. Measurement Results of the Pre-Incident Principal Rays

4.1. Measurement Conditions

The cameras used were XC-EI30 (Type 1/3 CCD, 768×494 pixels) made by Sony Corporation, and the lenses used were TL-4M (4 mm F2.0) made by TESCO JAPAN CO., LTD. This combination provides a horizontal view angle of approximately 60° . A frame grabber that performs 8-bit A/D conversion in sync with the pixels was newly developed and used. A double-layer plastic shield was placed in front of two cameras, and the distance between the cameras was approximately 79 mm.

The position of the point light source was varied so that the x coordinate was set at 800 mm and 1 600 mm with the center of rotation of the rotation stage as the origin. The azimuth estimation accuracy was ± 0.05 pixels or less.

4.2. Pre-Incident Principal Rays

Figure 11 shows the pre-incident principal rays for the one line in the center of the screen projected onto the X-Y plane. It can be observed that the rays incident from the right hand side of the figure have widening characteristics and thus they do not converge perfectly onto a single point. Widening produced by error in estimating the light source position is approximately 0.3 mm, and the incident rays can be seen here to converge with widening of approximately 0.8 mm, thus verifying deviation from the pinhole camera model. However, note that the straight lines in Fig. 11 are extensions of the pre-incident principal rays, and the actual rays that pass through the lens undergo complex refraction inside the lens. Also, the lens section shown in the figure is a model drawing, and does not represent the actual lens geometry.

The lens used showed a tendency for the base point of the pre-incident principal rays to be located farther back for rays incident from a wider outside angle. Figure 12 shows even greater widening for a super-wide angle lens (TS2V114E made by Pentax Corporation).

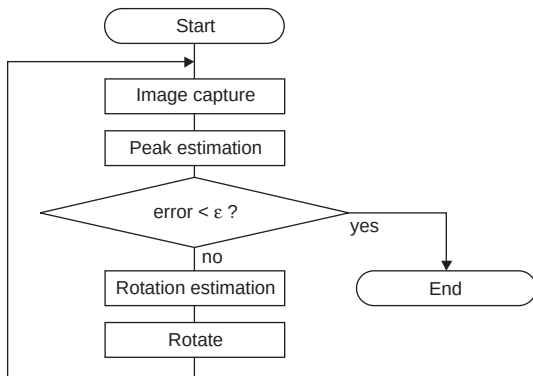


Fig. 10 Direction estimation processing flow

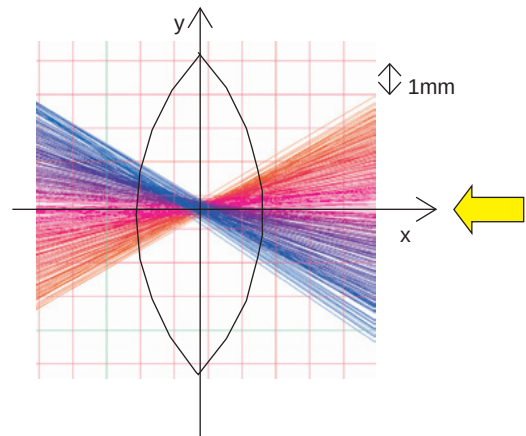


Fig. 11 Incident ray

4.3. Comparison of Accuracy by 3D Position Calculation

In order to compare and verify calibration accuracy with the conventional method, a stereo camera system was configured and calibrated, a point light source was moved to the prescribed positions, and the estimated 3D position of the light source was compared with the actual moved light source position. Comparison was performed for each of the three types of calibration data noted below, both with and without a shield.

Measured values: Measured value data obtained using the newly developed method that takes into account deviation from the pinhole camera model

Approximation values 1: Data approximated using the pinhole camera model, with changes in the base point position eliminated from the measured values, and the principal point matched

Approximation values 2: Data approximated by applying a 5th-order polynomial to the measured values

In addition, the position of the point light source was estimated as follows. The 2D image positions of the point light source were estimated with subpixel accuracy from each of the right and left images. The parameters shown in Fig. 2 that correspond to these image positions were then obtained from the calibration data and applied to Eq. (1) to obtain the 3D position.

The point light source was fixed to 0 in the Z direction, and moved at 0.01 m intervals from 0.8 to 1.6 m in the X direction, and at 0.01 m intervals from -0.4 to 0.4 m in the Y direction.

Generally speaking, the X direction error increases proportionally to the square of the distance, and the Y and Z direction errors increase proportionally to the distance in stereo vision based range estimation methods. Therefore, the error was evaluated by normalization with the square of the distance, and the distance, respectively. For example, when measurement error in the X direction is 12 mm at a position of 2 m, the normalized error is $12/(2^2) = 3$. The measurement is m^{-1} in the X direction, and unitless in the Y and Z directions.

Table 1 summarizes the normalized error (2σ) over the measurement range when there is no shield. The X direction error using measured values with calibration data obtained by the newly developed method was reduced by 50% and 20%, compared to measurement results obtained using calibration data derived from the approximation values 2 and the approximation values 1, respectively.

When the estimation accuracy of the 2D image position of the point light source had error of ± 0.5 pixels, the theoretical errors

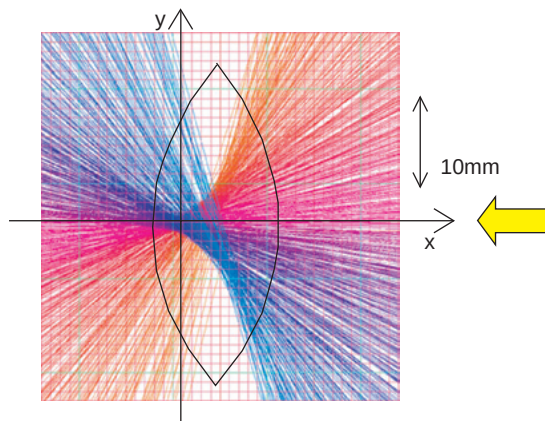


Fig. 12 Ray incident on wide angle lens

Table 1 Difference in error with no shield

	Measured data	Approximation data	
		Pinhole camera	Polynomial
X	± 4.2	± 5.2	± 8.5
Y	± 0.23	± 0.34	± 0.79
Z	± 0.22	± 0.24	± 0.90

Table 2 Difference in error with shield

	Measured data	Approximation data	
		Pinhole camera	Polynomial
X	± 4.6	± 5.4	± 18.5
Y	± 0.30	± 0.49	± 1.28
Z	± 0.72	± 0.71	± 2.80

produced by 3D value estimations were 8.7, 0.75 and 1.80 in the X, Y and Z directions, respectively. The measured value errors were half of these amounts, which shows that subpixel estimation in the 2D image works effectively. The reason that a large difference is not seen between Z direction error for the measured values and the approximation values 1 is that only the base point offset contributes to the error.

Table 2 shows the case when a warped shield is attached. This table shows that using measured values reduces the X direction error by 75% and 15% compared to measurement results obtained using the approximation values 2 and the approximation values 1, respectively. The measured value error is greater than when no shield is present, but still less than the approximation values 1 error with no shield.

These results show that the newly developed method achieves calibration with approximately the same accuracy, regardless of whether or not a shield is present. In addition, even when a shield is present, accuracy is higher than when using the pinhole camera model with no shield. Furthermore, there is an obvious improvement in accuracy over the polynomial approximation when irregular distortion is present.

These measurement results showed a reduction in the X direction error of approximately 15 to 20%. However, when using a lens with large aperture, low longitudinal symmetry and complex configuration such as a bright, wide angle lens, the deviation from the pinhole camera model increases as shown in Fig. 12, so this difference is thought to be even greater.

Figures 13 to 15 show 3D graphs of the normalized error in the X direction for the cases when no shields were present. The measured values are shown in Fig. 13, the approximation values 1 in Fig. 14, and the approximation values 2 in Fig. 15. Here, the center of the right camera is located on the x axis of the X-Y coordinate.

The error when using the pinhole camera model increased relative to the measured values at close distances. This is because fluctuation in the base point position due to deviation from the pinhole camera model increases at close distances, where the incident ray angles to the right and left cameras of a stereo camera system increase.

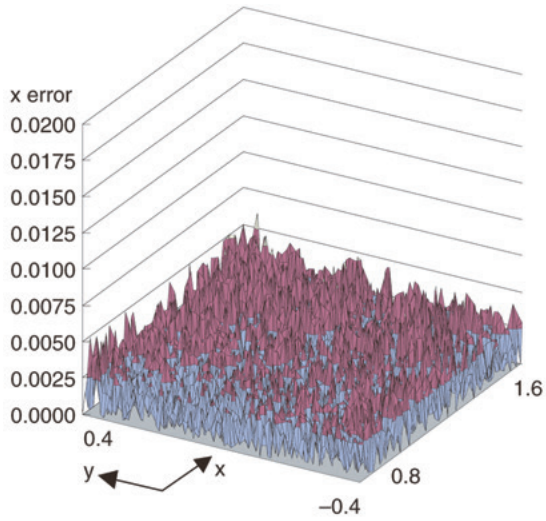


Fig. 13 Measured data with no shield

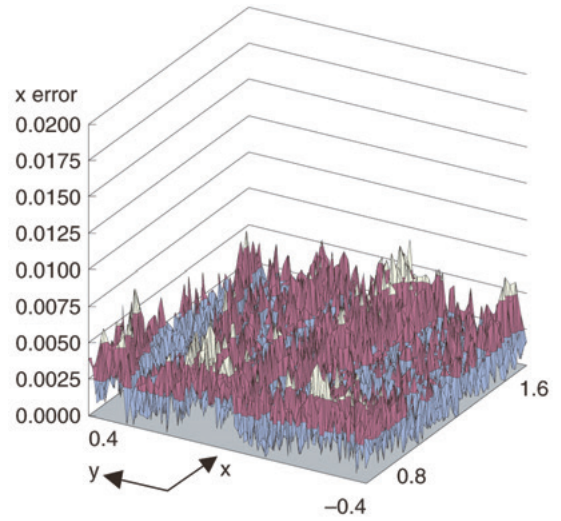


Fig. 16 Measured data with shield

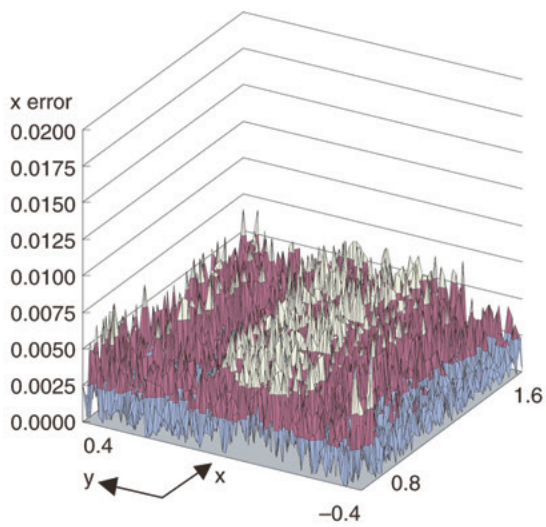


Fig. 14 Pinhole camera model with no shield

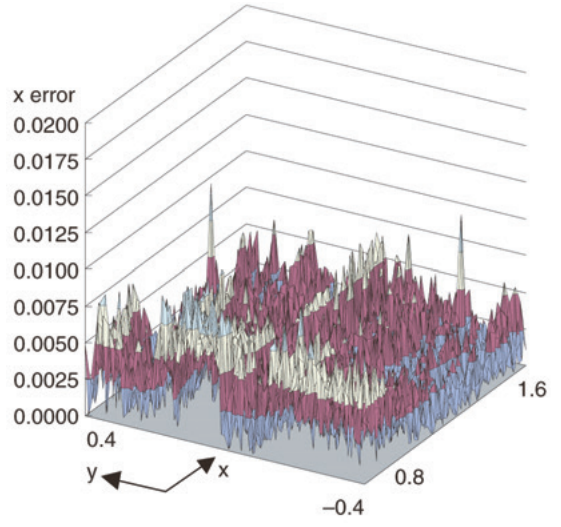


Fig. 17 Pinhole camera model with shield

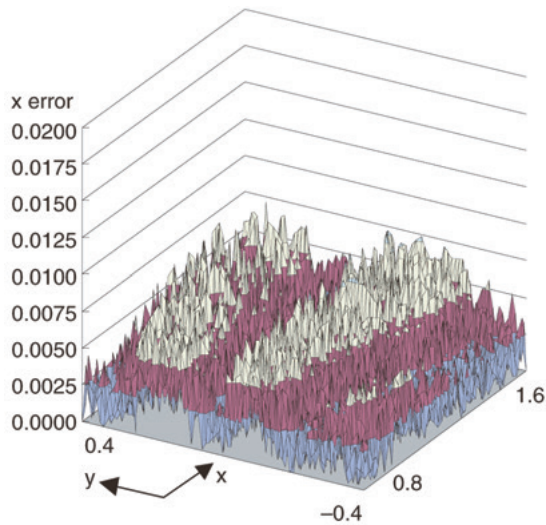


Fig. 15 Polynomial approximation with no shield

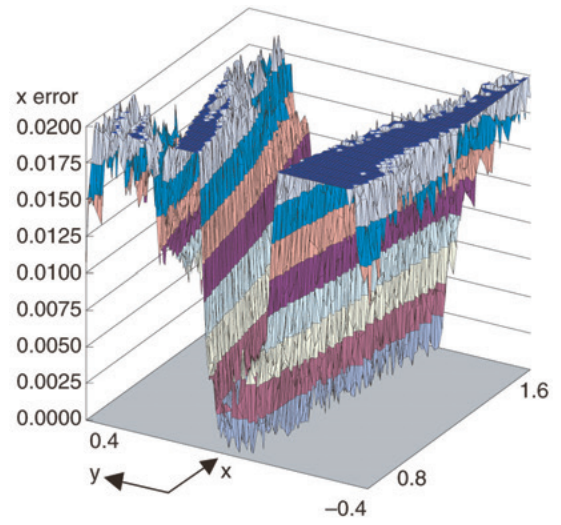


Fig. 18 Polynomial approximation with shield

Figures 16 to 18 show error for the cases when shields were present. Figure 16 shows the measured values, Fig. 17 shows the approximation values 1, and Fig. 18 shows the approximation values 2.

Comparison of Fig. 16 and Fig. 17 shows that the error occurring with the approximation values 1 is also greater at close distances. Figure 18 shows that the distortion cannot be sufficiently corrected over the entire measurement area for the approximation values 2.

As a result, an autonomous robot with two legs and two arms with a shield was able to perform a grasping task that requires accuracy of approximately 1 cm or less including control of the robot system, using this stereo camera system.

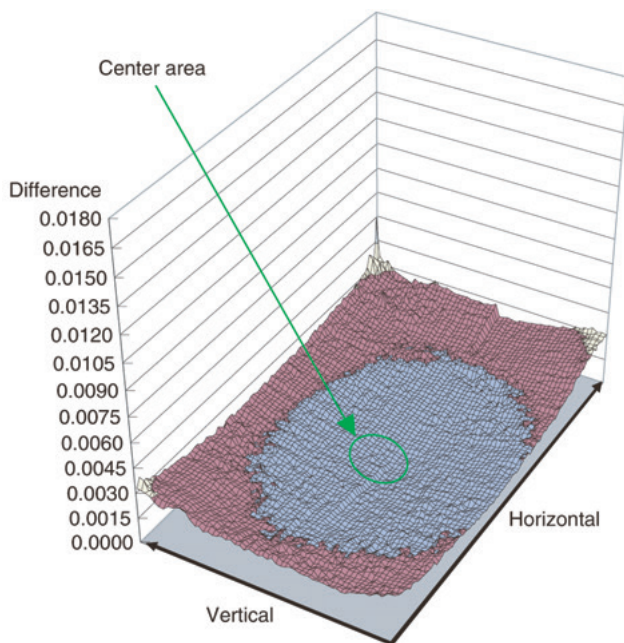


Fig. 19 Error for polynomial approximation with no shield

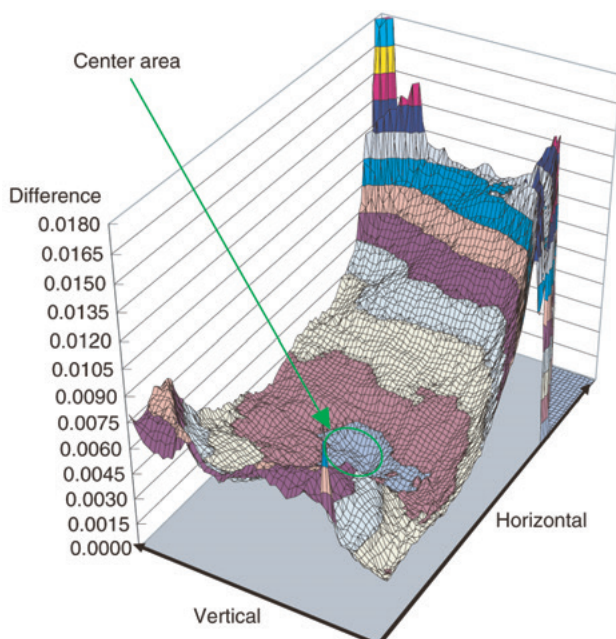


Fig. 20 Error for polynomial approximation with shield

4.4. Irregular Distortion Caused by a Shield

Figures 19 and 20 compare an approximation of distortion by the conventionally used 5th-order polynomial, and measured values obtained by the newly developed system, in order to examine the image distortion irregularity. The lower left portion of the image corresponds to the bottom of the graphs. Tangents of the measured values and approximation values were obtained for the angles α and θ shown in Fig. 2, and the square root of the sum of the squares of the difference between $\tan \alpha$ and $\tan \theta$ is displayed on the height axis as the error. Error of one pixel equals approximately 0.0015, and converted to distance error at 1 m, this corresponds to approximately 9 mm in the x direction.

Figure 19 shows the approximation error with no shield, and shows that the error is regular. This error is caused by a spiral offset, so it cannot be eliminated merely by polynomial correction that expands and contracts the image radially from the center of the image distortion.

Figure 20 shows the approximation error with a shield. The undulations in the graph show that the shield causes extremely complex distortion. The range with error of one pixel or less is limited to the purple range around the green circle in the center of the screen. The peripheral areas with error of 0 are areas where calibrations were not performed. This is because the peak positions in these areas could not be estimated due to poor focus, to the extent that there were multiple peaks, and thus the ray path could not be estimated.

5. Conclusion

This paper presented a camera calibration method that takes deviation from the pinhole camera model into account by tracing the path of the rays incident on a camera, and the accuracy of a stereo camera system that uses these measurement results.

This study confirmed the following items.

- (1) Camera calibration can be performed accurately, even when there is irregular image distortion.
- (2) The incident rays do not converge, and instead spread around the principal point.

A stereo camera system with camera calibration performed using this newly developed method obtained more accurate, relative 3D position estimation results than conventional methods, and made it possible for a robot to perform grasping motion.

References

- (1) Sakagami, S., Watanabe, R., Aoyama, C., Matsunaga, S., Higaki, N., Fujimura, K.: The intelligent ASIMO: system overview and integration, Proceedings of the 2002 IEEE/RSJ International Conference on Intelligent Robots and Systems(IROS), Vol. 3, p. 2478-2483 (2002)
- (2) Higaki, N., Yokoyama, T., Fujimura, K.: Gesture Recognition Algorithm for Understanding Man-Machine Interaction and Its Adaptation for New ASIMO, HONDA R&D Technical Review, Vol. 15, No.1, p. 95-102 (2003)
- (3) Goldberg, S., Maimone, M., Matthies, L.: Stereo Vision and Rover Navigation Software for Planetary Exploration, 2002 IEEE Aerospace conference Proceedings, Vol. 5, p. 2025-2036 (2002)

- (4) Sabe, K., Fukuchi, M., Gutmann, J., Ohashi, T., Kawamoto, K., Yoshigahara, T.: Obstacle Avoidance and Path Planning for Humanoid Robots using Stereo Vision, Proceedings of the IEEE International Conference on Robotics and Automation (ICRA), Vol. 1, p. 592-597 (2004)
- (5) Faugeras, O., Hotz, B., Mathieu, H., Vieville, T., Zhang, Z., Fua, P., Theron, E., Moll, L., Berry, G., Vuillemin, J., Bertin, P., Proy, C.: Real time correlation-based stereo: algorithm, implementations and applications, INRIA Research Report 2168 (1994)
- (6) Kanade, T., Okutomi, M.: A stereo matching algorithm with an adaptive window, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 16, No. 9, p. 920-932 (1994)
- (7) Zhang, Z., Deriche, R., Faugeras, O., Luong, Q.: A Robust Technique for Matching Two Uncalibrated Images Through the Recovery of the Unknown Epipolar Geometry, Artificial Intelligence Journal, Vol. 78, p. 87-119 (1996)
- (8) TSAI, R.: A Versatile Camera Calibration Technique for High-Accuracy 3D Machine Vision Metrology Using Off-theShelf TV Cameras and Lenses, IEEE J. Robotics and Automation, Vol. 3, No. 4, p. 323-344 (1987)
- (9) Shih, S.: Accuracy analysis on the estimation of camera parameters for active vision systems, Proceedings of the 13th International Conference on Pattern Recognition, p. 930-5, Vol. 1, (1996)
- (10) Shih, S.: Kinematic calibration of a binocular head using stereo vision with the complete and parametrically continuous model, Proceedings of the SPIE - The International Society for optical Engineering, Vol. 1825, p. 643-57 (1992)

■ Author ■



Chiaki AOYAMA