P3-11

17th International Conference on Machine Vision Applications (MVA)
Fully Online, July 25–27, 2021.

Multiple Fisheye Camera Calibration and Stereo Measurement Methods for Uniform Distance Errors throughout Imaging Ranges

Nobuhiko Wakai
Panasonic Corporation
wakai.nobuhiko@jp.panasonic.com

Takeo Azuma
Omni Vision Technologies
takeo.azuma@ovt.com

Kunio Nobori
Panasonic Corporation
nobori.kunio@jp.panasonic.com

Abstract

This paper proposes calibration and stereo measurement methods that enable accurate distance and uniform distribution of the distance error throughout imaging ranges. In stereo measurement using two fisheye cameras, the distance error varies greatly depending on the measurement direction. To reduce the distance error, the proposed method introduces an effectual baseline weight into the stereo measurement using three or more fisheye cameras and their calibration. Accurate distance is obtained because this effectual baseline weight is the optimum weight in the maximum likelihood estimation. Experimental results show that the proposed methods can obtain an accurate distance with a 94% reduction in error and make the distribution of the distance error uniform.

1 Introduction

Stereo cameras have been widely commercialized as distance sensors for autonomous vehicles, drones, and robots, and fisheye cameras that widen a field of view (FOV) are used for many applications. However, accurate distance measurement throughout imaging ranges with fisheye stereo is a significant challenge because the distance error varies from a small value to infinity depending on the measurement direction with respect to the baseline.

For reducing the variation in the distance error depending on the measurement direction with respect to the baseline, *multi-baseline stereo measurement* is effective when the FOV is relatively narrow. In contrast, when a fisheye camera with a wide FOV is used, as in Fig. 1, the closer the effectual baseline length in the measurement direction is to 0, the closer the distance error is to ∞. Since the conventional multi-baseline stereo technique does not have a mechanism for suppressing such a large error, the measurement error becomes huge.

Considering large errors is also important for calibration. However, in the conventional calibration methods, all points are used equally without considering a large error, so the calibration error becomes huge when a fisheye camera is used. Therefore, to minimize the variation in the distance error, we propose calibration and stereo measurement methods that introduce effectual baseline weights. The proposed methods can obtain an accurate distance and make the error distribution uniform because this effectual baseline weight is the optimum weight in the maximum likelihood estimation.

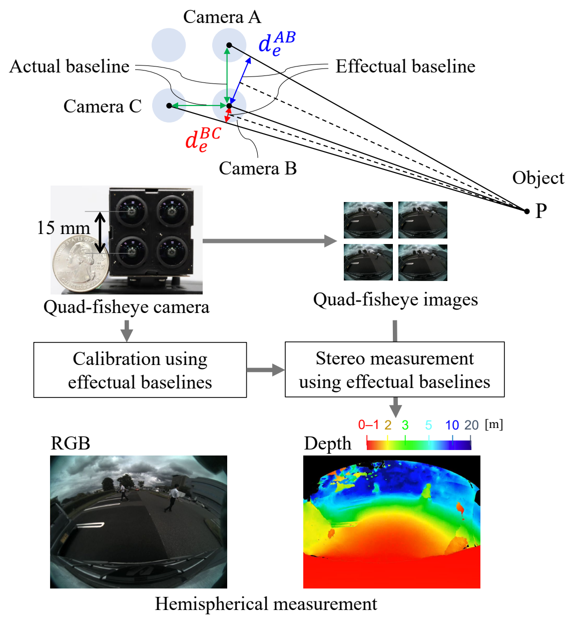

Figure 1: Concept diagram of proposed calibration and measurement methods using effectual baselines.

The major contribution of this paper is that our proposed methods can obtain uniform distance errors over the entire measurement direction for three or more fisheye cameras. Our work is the first to achieve uniform error distribution over the entire measurement direction in practical use case.

2 Related works

Stereo measurement and calibration methods for narrow-view cameras are well-established. In contrast, there has been relatively little work on fisheye stereo measurement, which remains a significant challenge.

Stereo measurement: Stereo measurement is based on triangulation [6]. Theoretically, it is difficult to measure an object with large incident angles, so narrow-view cameras are generally used for stereo measurement. Three or more cameras are typically utilized to improve the measurement accuracy narrow-view cameras. Okutomi *et al.* [9] proposed a pioneering approach to multi-baseline stereo that merges stereo pairs based on each sum of squared errors. Additionally, to select the best stereo pair among cameras, Amat *et al.* [1] presented a method with weights of uncertainty based on radius errors. In contrast to narrow-view cameras, with fisheye cameras it is essential to consider large incident angles causing large distance errors. However, conventional methods do not work well when they look at the large incident angles near the extension of the baseline.

Camera calibration: Stereo measurement methods require the specific camera parameters. The numerous camera calibration methods that have been proposed so far are classified into methods for monocular cameras [12, 15], binocular cameras [2, 3, 8, 10, 11, 13, 14, 16, 17], and three or more cameras [4, 5].

As a typical example of a calibration method that uses a narrow-view camera, Tsai [12] developed a method that minimizes the reprojection errors in image coordinates. This is importance because calibration methods for monocular cameras lack the relation of cameras in stereo measurement. To optimize fisheye stereo cameras, calibration methods using two steps, *i.e.*, separated optimization for intrinsic and extrinsic parameters, have been proposed [3, 11]. However, large errors near the extension of the baseline have not been taken into account. To reduce errors and/or improve the ranges of view, three or more cameras are typically used for the measurement. Findeisen *et al.* [4] presented an approach for surveying an indoor scene that uses three omnidirectional cameras to estimate the full hemispherical depth. However, similar to the stereo camera case, objects near the extension of the baseline causes large errors in this method.

As discussed above, it has been difficult for previous works to perform precise measurements over the entire measurement direction using fisheye cameras.

3 Proposed Method

This section describes our proposed methods of stereo measurement and calibration using baseline weights to improve the measurement accuracy using multiple cameras.

3.1 Stereo measurement using baseline weights

Our stereo measurement method uses baseline weights to reduce distance errors including large incident angles by merging the points of multiple stereo pairs. The weights minimize the variance of distance errors depending not on actual baselines but on effectual ones.

3.1.1 Definition of effectual baselines

Let us define the effectual baseline that represents the effectual length in triangulation. As shown in Fig. 2 (*top*), we

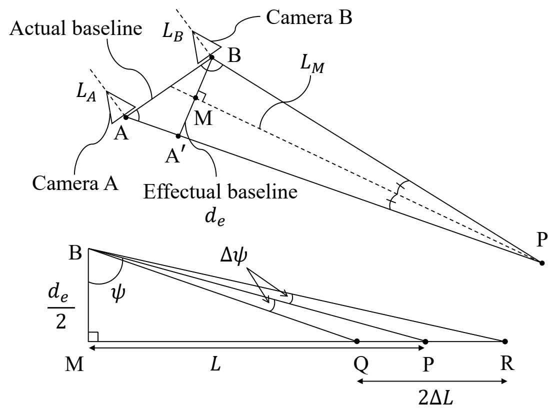

Figure 2: Illustration of the effectual baseline d_e and error estimation. Top: Definition of effectual baseline. Bottom: Error estimation using the effectual baseline.

assume that camera A and camera B are located at point A and point B, respectively, to measure the 3D point P. The actual baseline between camera A and camera B is $\overline{AB}$. Although the baseline is $\overline{AB}$, $\overline{A'B}$ is the effectual baseline for measuring point P under the condition that L_M is the bisector of the angle at P, and $\overline{A'B}$ is orthogonal to $\overline{MP}$. Note that M is located at the middle of $\overline{A'B}$, and the optical axes of camera A and camera B are L_A and L_B, respectively. We call $\overline{A'B}$ the effectual baseline d_e in this paper.

3.1.2 Error estimation of stereo measurement

The purpose of the error estimation is to describe the relation among object distance, angle errors, and effectual baselines under practical conditions. Using effectual baselines, we estimate the stereo measurement errors in Fig. 2 (*bottom*). We assume here that there are angle errors in $\pm\Delta\psi$ causing a position shift from P to Q or R. In this case, $\overline{MP}$ and $\overline{QR}$ are L and $2\Delta L$, respectively. The distance error ΔL against L for measuring P is estimated as

$$\Delta L \approx (\overline{PQ} + \overline{PR})/2. \tag{1}$$

Actually, the error distribution is a complex 3D shape depending on the object coordinates. However, we regard the errors as two representative points, *i.e.*, near point Q and far point R, because the errors are small enough to be simplified. Under this assumption, we derive ΔL using L, $\Delta\psi$, and d_e based on Eq. (1) as follows

$$\begin{aligned} \Delta L &= d_e/4 \cdot (\tan(\psi + \Delta\psi) - \tan(\psi - \Delta\psi)) \\ &= \frac{d_e(\tan^2\psi + 1)\tan\Delta\psi}{2(1 - \tan^2\psi \tan^2\Delta\psi)}. \end{aligned} \tag{2}$$

We regard $\tan^2\Delta\psi$ as 0 since $\Delta\psi$ is a minute value. Equation (2) is rewritten as

$$\begin{aligned} \Delta L &\approx d_e/2 \cdot (\tan^2\psi + 1)\tan\Delta\psi \\ &= 2L^2/d_e \cdot (1 + (d_e/L)^2/4)\tan\Delta\psi. \end{aligned} \tag{3}$$

Practically speaking, d_e/L is a minute value since L is significantly longer than d_e. Under the assumption of $(d_e/L)^2 = 0$, we thus obtain the simple relation shown in

$$\Delta L \approx 2L^2/d_e \cdot \tan\Delta\psi. \tag{4}$$

Finally, we assume that $\tan\Delta\psi = \Delta\psi$ and $\Delta\psi$ is constant in Eq. (4). We simply describe the relation among ΔL, L, and d_e as

$$\Delta L \approx 2L^2\Delta\psi/d_e \propto L^2/d_e. \tag{5}$$

It is important for fisheye cameras to take the effectual baseline d_e with L^2 into account because the range of d_e in fisheye cameras is wider than that in narrow-view cameras. In addition, the relation between the distance error and the effectual baseline length shows that short-baseline cameras tend to have the large errors described in Eq. (5).

3.1.3 Effectual baseline weights

To improve the measurement accuracy, we use effectual baselines for weights of merging ${}_n\mathrm{C}_2$ points (n is the number of cameras). The effectual baseline weight w satisfies the maximum likelihood that minimizes the variance of distance errors to obtain the accurate distance shown in

$$w_{i,j} = d_{e,i,j}^2, \tag{6}$$

where i and j are the indices of 3D points and cameras, respectively. If the effectual baseline is approximately 0, we remove the stereo pair to avoid adding infinity errors. Further, we define the normalized weight to merge the 3D point written in

$$w'_{i,j} = \frac{w_{i,j}}{\sum_{j=1}^{{}_n\mathrm{C}_2} w_{i,j}}. \tag{7}$$

Each 3D point is thus measured using the normalized weight, as

$$P_i = \sum_{j=1}^{{}_n\mathrm{C}_2} w'_{i,j} U_{i,j}, \tag{8}$$

where P_i is the 3D point of index i using ${}_n\mathrm{C}_2$ pairs of cameras, and $U_{i,j}$ is a measured 3D point of index i using a stereo pair index j by standard stereo measurement. Therefore, we use P_i as a measured point taking effectual baselines into consideration in both real-time stereo measurement and the calibration described next.

3.2 Calibration using effectual baselines

Here we describe our efficient calibration method for multiple cameras over the entire measurement direction. Our method appropriately optimizes camera parameters owing to effectual baseline weights.

3.2.1 Objective function for multiple cameras

For multiple stereo measurement using three or more cameras, we define the objective function using 3D measurement errors directly to optimize camera parameters based on Eq. (5), as

$$J_{multi} = \sum_{i=1}^{N} \frac{||P_i - \hat{P}_i||_2}{L_i^2}, \tag{9}$$

where N is the number of calibration data, P_i is the merged point in Eq. (8), and $\hat{P}_i$ is the ground-truth world coordinate of the calibration data. Each calibration error is normalized using the square distance.

3.2.2 Objective function for stereo cameras

We define the objective function J_{stereo} using effectual baselines for stereo cameras since there is only one measured point in each calibration data. Effectual baselines are used for weighting calibration errors instead of merging measured points, as

$$J_{stereo} = \sum_{i=1}^{N} \frac{d_{e,i}^2 \, ||P_{stereo,i} - \hat{P}_i||_2}{L_i^2}, \tag{10}$$

where P_{stereo} is the standard stereo measurement point.

Thanks to looking at effectual baselines, our weighting method enables uniform error distribution in fisheye stereo cameras. Note that our method can be applied for any cameras owing to its objective function independent of camera models.

4 Experiments

This section presents that the effectiveness of our weighting method using effectual baselines.

4.1 Experimental setup

To focus on camera geometry, we used ground-truth image coordinates of calibration points instead of stereo matching for evaluation. For the real-time depth estimation, semi-global matching [7] was performed for stereo matching.

Quad-fisheye camera: We developed quad-fisheye cameras with short baselines for the semispherical stereo measurement shown in Fig. 1. Quad-fisheye lenses are placed $15\,\mathrm{mm}$ apart on a 25-mm square base and capture synchronized images in 1280×960 pixels. The diameter and focal length of each fisheye lens is $9\,\mathrm{mm}$ and $0.88\,\mathrm{mm}$, respectively. The projection is based on stereographic projection with a large FOV (~190°). For non-directional measurement and long baselines, the vertexes of a regular polygon are appropriate camera positions.

Calibration targets: The ground truth of 1898 points in the world coordinate was measured with a 3D laser scanner

Table 1: Comparison of RMS relative distance errors (%).

Calibration methods	Merging measured points	4 cameras	3 cameras	2 cameras
Tsai's method [12]	center of gravity	18.25	24.40	39.41
our baseline weighting	center of gravity	3.60	4.98	8.19
our baseline weighting	our baseline weighting	1.13	1.41	–

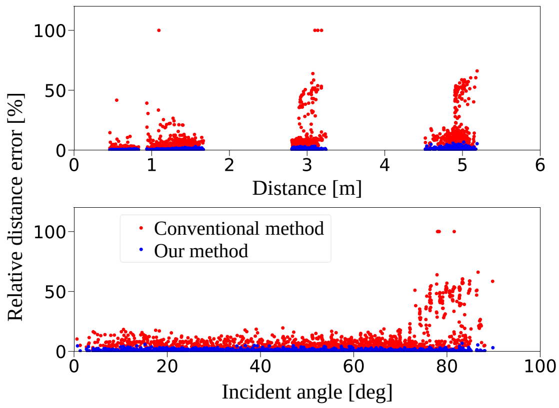

Figure 3: Relative distance errors using four cameras. The conventional method and our method correspond to the first and third rows in Tab. 1, respectively. Top: Errors against the distance between the camera and ground truth of each point. Bottom: Errors against the incident angle.

(FARO Focus^{3D} X 130) and the corresponding image coordinates were detected in sub-pixel precision. These points consisted of a wide range of distances (0.5–5.2 m) and incident angles (0°–90°).

4.2 Experimental results

To simplify errors, we used a relative distance error:

$$\varepsilon_i \quad = \quad ||P_{est,i} - \hat{P}_i||_2 \;/\; L_i, \tag{11}$$

where i is the index of calibration point, $P_{est,i}$ is the measured point, $\hat{P}_i$ is the ground truth of 3D point, L_i is the distance between $\hat{P}_i$ and the center of gravity among quad-fisheye cameras. Note that we clamped errors with 100% for the maximum error. The input of a set of image coordinates was given using the same points for calibration.

Table 1 shows a comparison of the root-mean-square (RMS) relative distance errors, where "3 cameras" indicates that the lower-left, upper-left, and upper-right cameras in Fig. 1 were used, and "2 cameras" indicates the upper-left and upper-right cameras were used. In Tsai's method [12], each camera in the quad-fisheye was calibrated individually. Note that we extended Tsai's method to adapt fisheye images. The extension substituted reprojection errors on image sensor coordinates for image coordinates to address ~90° incident angles. Since it was not necessary to merge measured points in the two-camera case, we ignore the merging measured points. As we can see, the distance error was reduced by 94% when our baseline weighting was used for both calibration and merging measured points. To avoid stereo measurement with short effectual baselines, at least three cameras are required. Thus, our method effectively reduced errors in the cases of three or more cameras.

Additionally, we found that using our method resulted in mean reprojection errors among cameras of 0.19, 0.23, and 0.18 pixels in the four-, three- and two-camera cases, respectively. In contrast, the mean reprojection errors among cameras using Tsai's method [12] were 0.60, 0.70, and 0.57 pixels. Note that the reprojection error of a camera is the RMS errors in a set of points. These results demonstrate that our calibration achieves fewer reprojection errors than when Tsai's method [12] is used.

Figure 3 shows the error distribution against the distance and angles. The conventional method had large errors compared to ours. In particular, points with more than ~70° incident angles tended to have large errors. In contrast, our method had small relative distance errors from 0° to 90° incident angles, *i.e.*, over the entire measurement direction, and achieved uniform error distribution even if the incident angles were large.

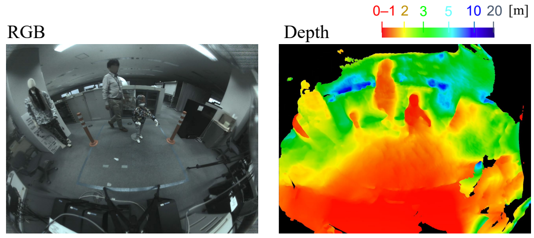

Figure 4: Example of indoor depth estimation using our quad-fisheye camera.

The outdoor and indoor depth maps using our quad-fisheye camera are shown in Figs. 1 and 4. The shapes of depth could accurately represent objects such as cars and persons, though the depth maps had some errors and noise.

5 Conclusion

To obtain an accurate distance and uniform distance error throughout imaging ranges using multiple fisheye cameras, we proposed baseline weighting for calibration and stereo measurement. Experimental results showed that our methods achieved a uniform error distribution and reduced the RMS relative distance errors by 94% using four fisheye cameras.

References

[1] J. Amat, M. Frigola, and A. Casals. Selection of the best stereo pair in a multi-camera configuration. In *Proceedings of IEEE International Conference on Robotics and Automation*, volume 4, pages 3342–3346, 2002.

[2] S. R. E. Datondji, Y. Dupuis, P. Subirats, and P. Vasseur. Rotation and translation estimation for a wide baseline fisheye-stereo at crossroads based on traffic flow analysis. In *Proceedings of IEEE International Conference on Intelligent Transportation Systems*, pages 1534–1539, 2016.

[3] M. Drulea, I. Szakats, A. Vatavu, and S. Nedevschi. Omnidirectional stereo vision using fisheye lenses. In *Proceedings of IEEE International Conference on Intelligent Computer Communication and Processing*, pages 251–258, 2014.

[4] M. Findeisen and G. Hirtz. Trinocular spherical stereo vision for indoor surveillance. In *Proceedings of Canadian Conference on Computer and Robot Vision*, pages 364–370, 2014.

[5] R. I. Hartley. Computation of the quadrifocal tensor. In *Proceedings of European Conference on Computer Vision*, pages 20–35, 1998.

[6] R. I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, second edition, 2004.

[7] H. Hirschmüller. Accurate and efficient stereo processing by semi-global matching and mutual information. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, volume 2, pages 807–814, 2005.

[8] Y. Jiang, Q. Zhu, and Q. Li. A novel practical approach of camera calibration for binocular stereo vision system. In *Proceedings of International Conference on Advanced Computer Theory and Engineering*, volume 2, pages 243–247, 2010.

[9] M. Okutomi and T. Kanade. A multiple-baseline stereo. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(4):353–363, 1993.

[10] J. Schneider, C. Stachniss, and W. Förstner. On the accuracy of dense fisheye stereo. *IEEE Robotics and Automation Letters*, 1(1):227–234, 2016.

[11] Shigang Li. Real-time spherical stereo. In *Proceedings of International Conference on Pattern Recognition*, volume 3, pages 1046–1049, 2006.

[12] 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*, 3(4):323–344, 1987.

[13] L. Wu and B. Zhu. Binocular stereovision camera calibration. In *Proceedings of IEEE International Conference on Mechatronics and Automation"*, pages 2638–2642, 2015.

[14] F. Yamano, H. Iida, K. Umeda, A. Ohashi, D. Fukuda, S. Kaneko, J. Murayama, and Y. Uchida. Improving the accuracy of a fisheye stereo camera with a disparity offset map. In *Proceedings of France-Japan and 10th Europe-Asia Congress on Mechatronics*, pages 94–97, 2018.

[15] Z. Zhang. A flexible new technique for camera calibration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(11):1330–1334, 2000.

[16] F. Zhao and Z. Jiang. A new algorithm for three-dimensional construction based on the robot binocular stereo vision system. In *Proceedings of International Conference on Intelligent Human-Machine Systems and Cybernetics*, volume 2, pages 302–305, 2012.

[17] C. Zhou, D. Tan, and H. Gao. A high-precision calibration and optimization method for stereo vision system. In *Proceedings of International Conference on Control, Automation, Robotics and Vision*, pages 1–5, 2006.