FISHEYE STEREO CAMERA USING FISHEYE VERTICAL STEREO METHOD

Hikaru Chikugo, Kento Arai, Sarthak Pathak and Kazunori Umeda

Precision Engineering Course, School of Science and Engineering, Chuo University

2024 IEEE International Conference on Image Processing (ICIP) | 979-8-3503-4939-9/24/$31.00 ©2024 IEEE | DOI: 10.1109/ICIP51287.2024.10647828

ABSTRACT

In this paper, we propose a wide-range and high-accuracy fisheye stereo camera using the fisheye vertical stereo method. In stereo measurement with two fisheye cameras placed horizontally, increased mismatching occurs due to template matching along curved epipolar lines. Additionally, because the baseline is horizontal, there is a decrease in the distance measurement accuracy in the left and right areas of the image. Therefore, by placing fisheye cameras vertically and straightening the epipolar lines, we reduce mismatching during template matching in stereo measurement, achieving high-accuracy stereo measurement. Moreover, by making the baseline vertical, we improve the distance measurement accuracy in the left and right areas. Experiments demonstrate that the distance measurement accuracy of our proposed method is higher than that of conventional methods.

***Index Terms*—** Fisheye Stereo Camera, Fisheye Vertical Stereo Method, Three-dimensional Measurement

1. INTRODUCTION

1.1. Background

In recent years, driver assistance systems have become widely adopted [1, 2]. These systems require the understanding of the surrounding environment for purposes such as obstacle detection, primarily utilizing distance information. Stereo cameras, LiDAR, and SONAR are typically used to obtain distance information. However, they often have narrow measurement range and/or low angular resolution of distance measurement, making it difficult to detect thin obstacles. While measurement range can be increased by combining multiple sensors and mechanisms, problems such as restrictions on mounting positions, deterioration of maintainability, and increased costs arise.

Therefore, this study focuses on fisheye stereo cameras. Fisheye cameras, with their approximately $180°$ field of view, enable wide-area measurement with a single unit. Furthermore, their large depth of field results in subjects being less blurred due to distance, which is advantageous for image recognition. These benefits have led to the application of fisheye cameras in automotive vision assistance systems and surveillance camera systems. However, in practice, it is difficult to use fisheye stereo cameras because of the curvature of the image and epipolar lines. This research focuses on the important practical problem improving the accuracy of fisheye stereo by proposing an alternative camera arrangement that leads to easier and more accurate stereo matching.

Corresponding author: chikugo@sensor.mech.chuo-u.ac.jp

1.2. Related Research

Previous research using fisheye cameras includes studies on lane detection, vehicle posture estimation, and obstacle detection [3, 4, 5]. There is also research on generating overhead images using multiple fisheye cameras [6, 7]. Regarding fisheye stereo cameras, previous studies include the following: Hane et al. achieved real-time 3D measurement using the Plane-Sweep method [8]. Abraham et al. simplified stereo matching by applying stereo rectification to fisheye stereo cameras [9]. Moreau et al. proposed a method for 3D environment reconstruction using a fisheye stereo camera with an equisolid angle projection model [10]. However, this method requires significant computational cost. Additionally, there are examples of applying fisheye cameras to practical platforms such as Unmanned Aerial Vehicles (UAVs) and vehicles [11, 12]. These studies convert fisheye images to perspective projection images to simplify the search for corresponding points. However, this conversion causes stretching at the image periphery, making stereo matching in peripheral regions difficult and narrowing the detection range compared to the original fisheye image's angle of view [8, 9, 10, 11, 12]. Moreover, research like Roxas et al.'s uses the Variational Fisheye Stereo (VFS) method without correcting distortion for fisheye cameras [13]. However, the VFS method iteratively optimizes minor variations globally, making it difficult to measure distance and fine structures at close range where large variations occur.

In contrast, Ohashi et al. have proposed a method for 3D measurement that reduces distortion in fisheye images by converting them into equirectangular images [14]. By transforming fisheye images into equirectangular images, which are represented in a Cartesian coordinate system with the azimuth angle λ and elevation angle φ as the horizontal and vertical axes, respectively, this method achieves distortion reduction in fisheye images without stretching the peripheral parts and simplifies the search for corresponding points.

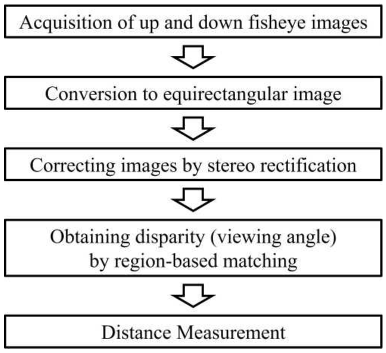


Fig. 1: The flow of proposed method

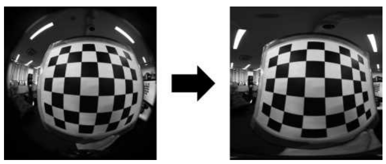


Fig. 2: Conversion to equirectangular images

However, the method still suffers from increased mismatching due to template matching along curved epipolar lines, leading to inadequate distance measurement accuracy. On the other hand, Iida et al. have proposed the pseudo-bilateral filter as a method that fuses region-based stereo measurement with feature-based Structure from Motion (SfM) [15]. Although the introduction of time-series images significantly improved distance measurement accuracy, the processing load is too high for real-time performanc. Additionally, time-series processing has the drawback of being inapplicable when the vehicle is stationary, such as during parking. Moreover, the methods by Ohashi et al. and Iida et al., which involve placing cameras horizontally for stereo measurement, result in insufficient distance measurement accuracy in the image's left and right areas compared to near the center due to the direction of the baseline length [14, 15]. When considering automotive applications, measurement accuracy in the left and right camera areas becomes crucial. Also, placing cameras on the left and right sides makes it easier to extract vertical edges in the environment but more challenging to extract horizontal edges. When envisioning camera sensors for automotive or mobile robot applications, the environment often contains horizontal edges, such as cars, curbs, and steps.

Therefore, when considering deployment on vehicles or mobile robots, it is necessary to accurately extract horizontal edges in the environment. Our study proposes the new concept of placing fisheye cameras vertically and processing them in the equirectangular projection. This leads to straight epipolar lines, simplifying stereo matching and making it more accurate. By having the baseline length vertical, distance estimation accuracy in the left and right areas also improves. The effectiveness of our proposed method is demonstrated through simulation experiments and real-world experiments to evaluate distance accuracy compared to conventional methods [14].

2. METHOD

2.1. Outline of the Proposed Method

The flow of the proposed method is shown in Fig.1. Initially, fisheye images are captured from the upper and lower cameras. These captured fisheye images are then transformed into equirectangular images with reduced distortion. Next, the transformed equirectangular images are corrected using stereo rectification with a checkerboard. Then, using the stereo-rectified equirectangular images, the disparity is obtained by performing template matching along the image's vertical epipolar lines. Finally, distance is calculated using the obtained disparity.

2.2. Conversion to Equirectangular Images

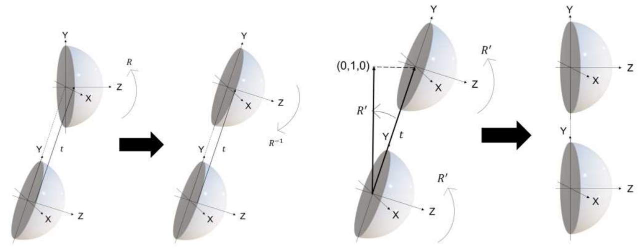
In this study, stereo measurement with reduced distortion effects is achieved by converting fisheye images into equirectangular images. However, since the output images are prone to errors, image correction is employed to mitigate the impact of these errors. Therefore, image correction that considers the camera's intrinsic parameters is performed to reduce the impact of errors. The OcamCalib Toolbox proposed by Scaramuzza et al. is used for estimating the intrinsic parameters [16]. The conversion of fisheye images to equirectangular images employs the method proposed by Ohashi et al. [14]. The image conversion results are shown in Fig. 2. From Fig. 2, the vertical distortion of the checkerboard is reduced.

2.3. Stereo Rectification

When conducting distance measurement with two cameras, it is common practice, for simplification, to align each camera's optical axis perpendicular to the baseline. In reality, the optical axes and the baseline of the cameras may not be perfectly perpendicular due to discrepancies during camera installation. Therefore, correction is performed using stereo rectification with checkerboards. The steps for stereo rectification are shown below.

First, estimate the corresponding points between images. To estimate the corresponding points between images, checkerboards are used. Fisheye images of the checkerboard are converted into equirectangular images. Then, find each corresponding point of the checkerboard between the images. The corresponding points obtained through this process are in equirectangular coordinates (λ, φ), and are converted into three-dimensional coordinates (x, y, z).This process is performed on images of the checkerboard taken from various angles. Then, using all the obtained corresponding points, estimate the external parameters.

Next, using all the corresponding points obtained with the checkerboard, estimate the external parameters between the cameras. Initially, calculate the fundamental matrix using the five-point algorithm from the obtained corresponding points



(a) Pose transformation using a pose matrix (b) Transformation of coordinate axis using position vector

Fig. 3: Rectification by transformation of the 3D coordinates

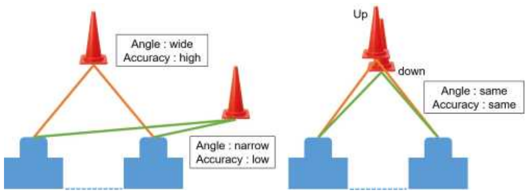


Fig. 4: Variation of distance measurement accuracy

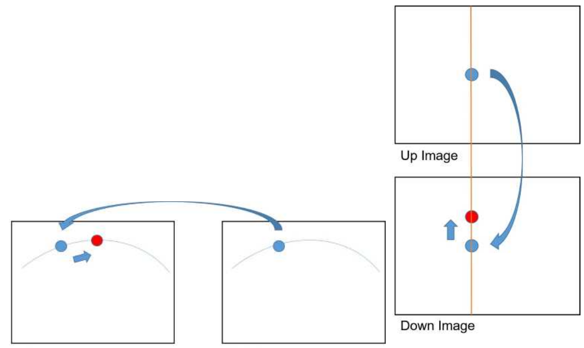
[17]. Then, by determining the external parameters from the fundamental matrix, find the translation vector t and the rotation matrix R, which represent the position and orientation of the lower camera as seen from the upper camera.

From the obtained external parameters, stereo images are rectified by rotating the 3D coordinates transformed from equirectangular coordinates [18]. The process overview is shown in Fig.3. Initially, 3D coordinates projected onto a hemispherical surface are obtained from the equirectangular images based on the azimuth angle λ and elevation angle φ. From the 3D coordinates projected onto the hemispherical surface, as shown in Fig.3(a), a posture transformation of the 3D coordinates is performed using the rotation matrix R. Subsequently, as shown in Fig.3(b), a transformation of the coordinate axis using the translation vector t is performed, resulting in the 3D coordinates being in a parallel state. Finally, by converting the 3D coordinates back into equirectangular coordinates, the images are rectified.

2.4. Fisheye Vertical Stereo Method

First, the advantages of the fisheye vertical stereo method are discussed. As shown in Fig.4, the distance measurement accuracy varies significantly across image locations depending on the orientation of the baseline. Therefore, by arranging the cameras vertically, an improvement in accuracy for the left and right areas is expected due to the vertical baseline.

The appearance of epipolar lines is shown in Fig.5. In traditional fisheye stereo methods, template matching along distorted epipolar lines led to increased mismatches and a decrease in accuracy [14]. In this study, leveraging the characteristic of equirectangular images where distortion is reduced in the vertical direction, epipolar lines can be easily straightened. Although it is possible to straighten epipolar lines when cameras are placed horizontally, creating equirectangular images in such a case requires rotating the image by 90°, necessitating additional processing steps. Furthermore, straightening the epipolar lines is expected to reduce mismatches during template matching. Discussing 3D measurement using the fisheye vertical stereo method, it is noted that in stereo matching, incorrect correspondences can occur due to estimation errors of internal and external parameters, lighting conditions, and other factors. This phenomenon, known as mismatches, negatively impacts obstacle recognition, necessitating the exclusion of such errors. Hence, this study employs a texture filter as a mismatch removal method.



(a) Epipolar lines in horizontal stereo (b) Epipolar lines in vertical stereo

Fig. 5: Template matching in an equirectangular image

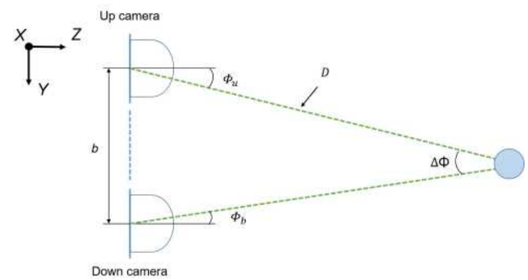


Fig. 6: Relation between measurement object and camera

Disparity is determined from the upper and lower images using the fisheye vertical stereo method. The Sum of Absolute Difference (SAD) is used to evaluate similarity during template matching. As shown in Fig.6, the distance to the measurement target is calculated using the formula (1):

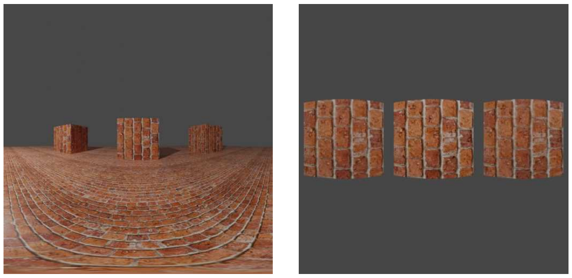
$$D = b\frac{\cos\varphi_d}{\sin\Delta\varphi} \tag{1}$$

Here, φ_u and φ_d represent the elevation angles of each camera, b is the baseline length, $\Delta\varphi$ is the disparity elevation angle, and D represents the Euclidean distance from the upper camera to the measurement target.

3. SIMULATION EXPERIMENTS

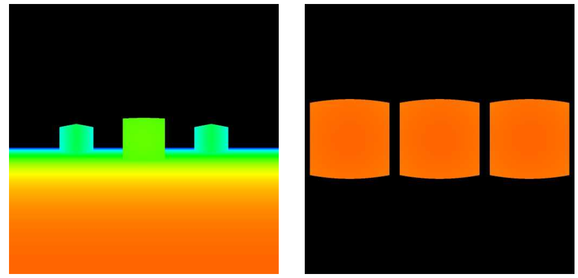
3.1. Experimental Conditions

In this experiment, two accuracy evaluation experiments were conducted using the 3D CG software Blender. The comparison was made against fisheye stereo measurement with



(a) First environment (b) Second environment

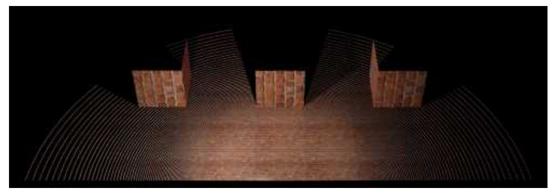
Fig. 7: Experimental environments



(a) First environment (b) Second environment

Fig. 8: True value

(a) front view



(b) top view

Fig. 9: 3D point cloud of true value

cameras placed on the left and right sides [14]. In Blender, the Cycles render engine was used, and the lens was set to produce panoramic equirectangular images. The camera resolution was set to 1048×1048 pixels, with a field of view of $180°$ both horizontally and vertically, and a baseline length between the cameras of $0.072\mathrm{m}$. In the first accuracy evaluation experiment, the true values obtained from Blender were compared with the measurement results of each method. The experimental environment is shown in Fig.7(a). In the second accuracy evaluation experiment, the Root Mean Squared Error (RMSE) between the true values obtained from Blender and the distance values obtained from each measurement was compared. The experimental environment is shown in Fig.7(b). The elevation angle in the image was kept constant at $0°$, and square blocks were placed at measurement positions with azimuth angles of $-60°$, $0°$, and $60°$. The side of the block was set to $1\mathrm{m}$, and the distance from the camera to the block was measured at five locations from $1\mathrm{m}$ to $9\mathrm{m}$.

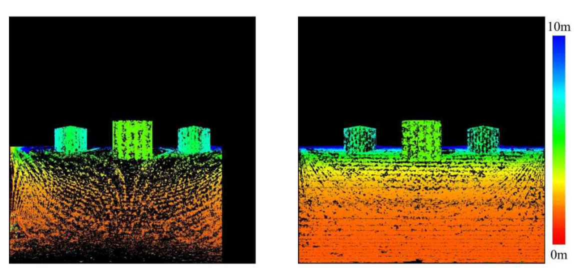
3.2. Experimental Results

The experimental results are shown in Figs.8-14. Fig.8 presents the true values obtained from Blender. Fig.9 displays these true values as a 3D point cloud. Fig.10 shows the measurement results of each method in the first accuracy evaluation experiment. Figs.11 and 12 represent the measurement results of each method in the first accuracy evaluation experiment as 3D point clouds. Fig.13 displays the measurement results of each method in the second accuracy evaluation experiment, and Fig.14 illustrates the RMSE for each distance in the second accuracy evaluation experiment.

From Figs.8-12, it is observed that the proposed method yields denser results compared to the traditional method, indicating that the proposed method produces results closer to the true values. The results represented as 3D point clouds show that the proposed method has fewer measurement errors than the traditional method. Additionally, the number of measurement points is $218,991$ for the traditional method and $449,357$ for the proposed method, demonstrating that the proposed method can perform denser distance measurements. This is attributed to the fact that correspondences are easier to find during stereo matching as the epipolar lines are straightened from curves. Moreover, measurement results in the left and right areas of the images are improved, likely due to the reduction in disparity angle errors at the image edges because of the baseline orientation.

3.3. Discussion

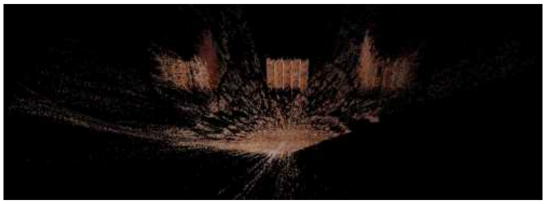
From Fig.14, it is evident that the RMSE is reduced at each distance with the proposed method, especially showing significant accuracy improvement in the left and right areas of the images at azimuth angles of $-60°$ and $60°$. This confirms the accuracy improvement in the left and right areas of the images by changing the baseline orientation. Accuracy improvements are also observed near the center of the images at an azimuth angle of $0°$, likely due to the reduction in mismatches as the epipolar lines become straight. The proposed method shows a larger error at $1\mathrm{m}$ distance compared to $2\mathrm{m}$ and $3\mathrm{m}$. Generally, the closer the measurement target, the larger the disparity and the wider the disparity search range. At $1\mathrm{m}$ distance, some areas exceeded the set disparity search range, leading to mismatches. Hence, a larger error at $1\mathrm{m}$ distance is observed.



(a) Conventional method (b) Proposed method

Fig. 10: Measurement results in first environment

(a) front view



(b) top view

Fig. 11: 3D point cloud of conventional method

(a) front view

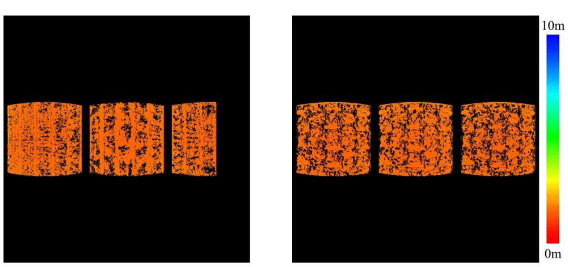
(b) top view

Fig. 12: 3D point cloud of proposed method

4. EXPERIMENTS IN REAL ENVIRONMENT

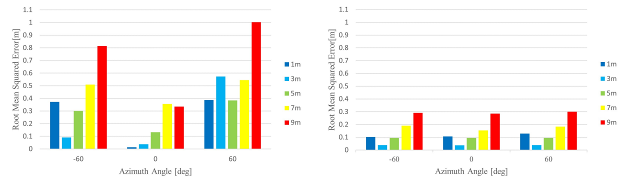
4.1. Experimental Conditions

A qualitative and quantitative evaluation of the proposed method was conducted, with the comparison being made against fisheye stereo measurement using cameras placed on the left and right sides [14]. The cameras used were from FLIR's Blackfly series, and the fisheye lenses were provided by Edmund Optics. The camera resolution was set at 1048×1048 pixels, with a baseline length of $0.072\mathrm{m}$ and a field of view of $180°$ both horizontally and vertically. The fisheye stereo camera setup used in the experiments is shown in Fig.15. Range images measured in indoor and outdoor environments were assessed. The experimental environment and measurement locations for the quantitative evaluation are depicted in Fig.16, with different colors indicating different measurement locations. The measurement distances were set at $1\mathrm{m}$, $3\mathrm{m}$, $5\mathrm{m}$, $7\mathrm{m}$, and $9\mathrm{m}$, using black and white paper as the measurement targets. The targets were moved left and right to be captured at each of the five locations shown in Fig.16. The true values were measured with a laser rangefinder. Measurements were taken by focusing on a point and measuring the points above and below it, repeating this process five times to obtain a total of 15 points. The error was evaluated based on the average and standard error of these 15 points.



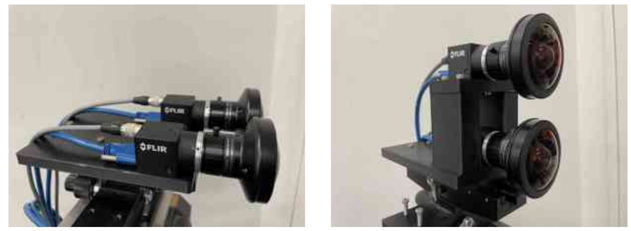
(a) Conventional method (b) Proposed method

Fig. 13: Measurement results in quantitative evaluation at 1m



(a) Conventional method (b) Proposed method

Fig. 14: Root mean squared error at each distance



(a) Conventional method (b) Proposed method

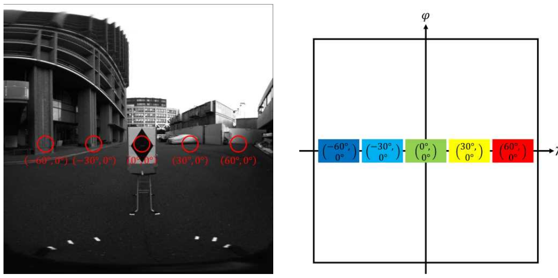
Fig. 15: Fisheye stereo camera

4.2. Experimental Results

The experimental results are shown in Figs.17-19. Fig.17 displays the equirectangular images and range images captured in indoor and outdoor environments. Fig.18 presents the range images for each method at a measurement distance of $1\mathrm{m}$ and an azimuth angle of $0°$. Fig.19 show the average errors and standard deviations at each measuring distance. The colors of the bar graphs correspond to the measurement locations indicated in Fig.16(b). From Fig.17, it can be seen that it was possible to measure distances to cars, curbs, people, and other obstacles, especially in the left and right areas of the images. However, texture-less surfaces like walls could not be measured, likely due to the use of the texture filter and the reliance on template matching for finding correspondences between upper and lower images.

4.3. Discussion

From Fig.18, the proposed method shows fewer mismatches and improved distance accuracy at farther distances. This improvement is attributed to the straightening of epipolar lines during template matching and the presence of many horizontal edges in the environment. Additionally, Fig.18(b) indicates that while vertical edges in the target at the image center were not measured accurately, horizontal edges were successfully measured, likely due to the vertical orientation of



(a) Outdoor environment (b) Measurement points

Fig. 16: Experimental conditions

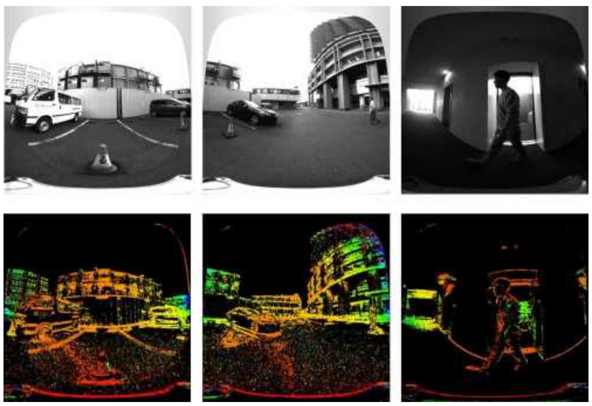
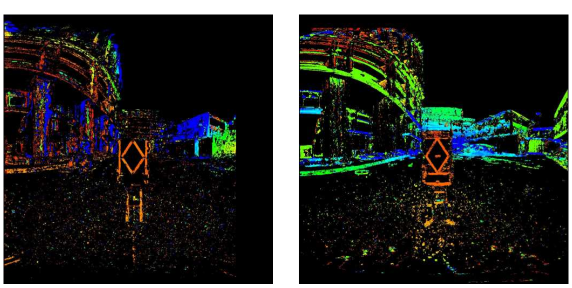


Fig. 17: Results of distance measurement (Same color representation as Fig. 10 and 13.)

the baseline. Fig.19 shows that, similar to the simulation experiments conducted in Section Ⅲ, distance accuracy has improved. At $1\mathrm{m}$, the distance accuracy of each method is comparable. However, from $3\mathrm{m}$ to $9\mathrm{m}$, the overall distance accuracy of the proposed method is enhanced, especially in the left and right areas of the images at azimuth angles of $-60°$ and $60°$, indicating a significant improvement. This improvement is likely due to the reduction of errors in the left and right areas of the images due to the vertical baseline. Furthermore, the reduction in field-of-view angle errors in the left and right areas results in comparable distance accuracy between the image center and the left and right areas.

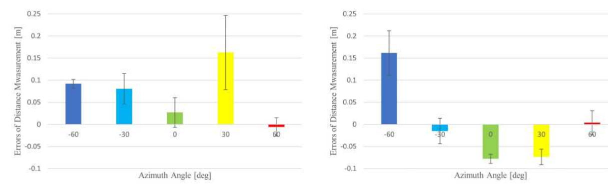


(a) Conventional method (b) Proposed method

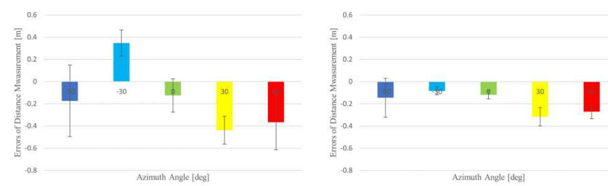
Fig. 18: Range images at 1m (Same color representation as Fig. 10 and 13.)

5. CONCLUSION

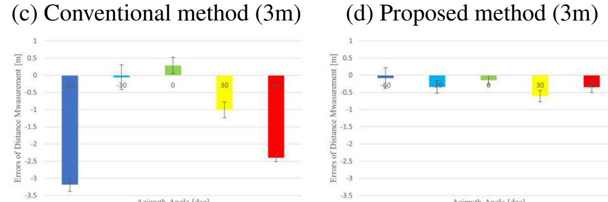
In this study, we proposed the fisheye vertical stereo method, which involves vertically arranging fisheye cameras. Through comparative experiments with conventional methods, we demonstrated the increase in accuracy as compared to the conventional fisheye stereo method. Future works include reducing mismatches and improving distance accuracy through image reprojection optimization. Another avenue is to aim for robust distance measurement in texture-less environments by integrating distance information obtained through deep learning. In addition, straightening the epipolar lines allows processing fisheye images in the same way as regular perspective stereo images, providing access to the vast array of stereo disparity estimation methods. Evaluating these methods and further improving accuracy will also be considered.



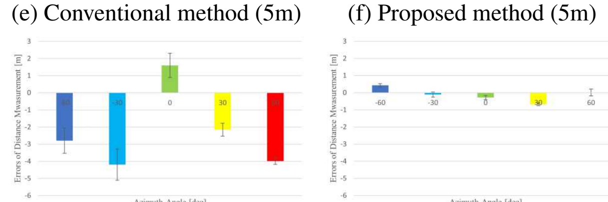
(a) Conventional method (1m) (b) Proposed method (1m)



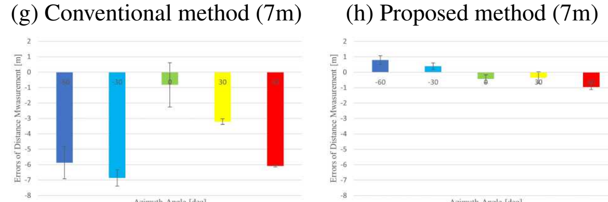
(c) Conventional method (3m) (d) Proposed method (3m)



(e) Conventional method (5m) (f) Proposed method (5m)



(g) Conventional method (7m) (h) Proposed method (7m)



(i) Conventional method (9m) (j) Proposed method (9m)

Fig. 19: Mean and standard deviation of errors from 1m to 9m

6. REFERENCES

[1] Jau-Woei Perng, Pei-Yu Liu, Kai-Quan Zhong, and Ya-Wen Hsu, "Front object recognition system for vehicles based on sensor fusion using stereo vision and laser range finder," in *2017 IEEE International Conference on Consumer Electronics - Taiwan (ICCE-TW)*, 2017, pp. 261–262.

[2] Nobuo Sasaki, Naoyuki Iijima, and Daiki Uchiyama, "Development of ranging method for inter-vehicle distance using visible light communication and image processing," in *2015 15th International Conference on Control, Automation and Systems (ICCAS)*, 2015, pp. 666–670.

[3] Chang-Hoon Kum, Dong-Chan Cho, Moon-Soo Ra, and Whoi-Yul Kim, "Lane detection system with around view monitoring for intelligent vehicle," in *2013 International SoC Design Conference (ISOCC)*, 2013, pp. 215–218.

[4] Shigang Li, Hideki Oshima, Isao Nakanishi, and Kikuo Fujimura, "Lane departure estimation by side fisheye camera," in *2011 IEEE International Conference on Mechatronics and Automation*, 2011, pp. 1004–1009.

[5] Tomoyu Sakuda, Hikaru Chikugo, Kento Arai, Sarthak Pathak, and Kazunori Umeda, "Estimation of road surface plane and object height focusing on the division scale in disparity image using fisheye stereo camera," *Journal of Robotics and Mechatronics*, vol. 35, no. 5, pp. 1354–1365, 2023.

[6] Takaaki Sato, Alessandro Moro, Atsushi Sugahara, Tsuyoshi Tasaki, Atsushi Yamashita, and Hajime Asama, "Spatio-temporal bird's-eye view images using multiple fish-eye cameras," in *Proceedings of the 2013 IEEE/SICE International Symposium on System Integration*, 2013, pp. 753–758.

[7] Cheng-Ming Huang, Ting-Yi Yang, Kuang-Yow Lian, Shih-Che Chien, and Min-Fong Luo, "Fish-eye cameras calibration for vehicle around view monitoring system," in *2014 IEEE International Conference on Consumer Electronics - Taiwan*, 2014, pp. 225–226.

[8] Christian Häne, Lionel Heng, Gim Hee Lee, Alexey Sizov, and Marc Pollefeys, "Real-time direct dense matching on fisheye images using plane-sweeping stereo," in *2014 2nd International Conference on 3D Vision*, 2014, vol. 1, pp. 57–64.

[9] Steffen Abraham and Wolfgang Förstner, "Fish-eye-stereo calibration and epipolar rectification," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 59, no. 5, pp. 278–288, 2005.

[10] Julien Moreau, Sebastien Ambellouis, and Yassine Ruichek, "Equisolid fisheye stereovision calibration and point cloud computation," 11 2013, vol. XL-7/W2.

[11] Nicola Krombach, David Droeschel, and Sven Behnke, "Evaluation of stereo algorithms for obstacle detection with fisheye lenses," 08 2015, vol. II-1/W1.

[12] Deukhyeon Kim, Jinwook Choi, Hunjae Yoo, Ukil Yang, and Kwanghoon Sohn, "Rear obstacle detection system with fisheye stereo camera using hct," *Expert Systems with Applications*, vol. 42, no. 17, pp. 6295–6305, 2015.

[13] Menandro Roxas and Takeshi Oishi, "Variational fisheye stereo," *IEEE Robotics and Automation Letters*, vol. 5, no. 2, pp. 1303–1310, 2020.

[14] Akira Ohashi, Fumito Yamano, Gakuto Masuyama, Kazunori Umeda, Daisuke Fukuda, Kota Irie, Shuzo Kaneko, Junya Murayama, and Yoshitaka Uchida, "Stereo rectification for equirectangular images," in *2017 IEEE/SICE International Symposium on System Integration (SII)*, 2017, pp. 535–540.

[15] Hirotaka Iida, Yonghoon Ji, Kazunori Umeda, Akira Ohashi, Daisuke Fukuda, Shuzo Kaneko, Junya Murayama, and Yoshitaka Uchida, "High-accuracy range image generation by fusing binocular and motion stereo using fisheye stereo camera," in *2020 IEEE/SICE International Symposium on System Integration (SII)*, 2020, pp. 343–348.

[16] Davide Scaramuzza, Agostino Martinelli, and Roland Siegwart, "A toolbox for easily calibrating omnidirectional cameras," in *2006 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2006, pp. 5695–5701.

[17] D. Nister, "An efficient solution to the five-point relative pose problem," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 6, pp. 756–770, 2004.

[18] Sarthak Pathak, Alessandro Moro, Atsushi Yamashita, and Hajime Asama, "A decoupled virtual camera using spherical optical flow," in *2016 IEEE International Conference on Image Processing (ICIP)*, 2016, pp. 4488–4492.