

Obstacle Detection and Height Estimation using Fisheye Stereo Camera Considering Intensity Information

Hikaru Chikugo¹, Tomoyu Sakuda¹, Sarthak Pathak¹ and Kazunori Umeda¹

Abstract— In this paper, we propose a novel method for obstacle detection and height estimation based on disparity and intensity information using a fisheye stereo camera. The method using only disparity information may incorrectly detect road surfaces as obstacles. Therefore, the proposed method detects obstacles by comparing the intensity of the obstacle edges with that of the disparity information. Experimental results show that the proposed method using disparity and intensity information can detect only obstacles without incorrectly detecting road surfaces. It is also shown that the accuracy of the height estimation does not change even though the road surface is not detected incorrectly.

I. INTRODUCTION

In recent years, there have been remarkable advances in automated driving technologies, and many of these technologies have already been put to practical use. Among them, many driver assistance systems are in widespread use. These systems support driving based on three-dimensional (3D) information of the surroundings measured by range sensors such as LiDAR and stereo cameras. To determine whether assistance is required, it is necessary to understand the environment surrounding the vehicle. Among the range sensors used to understand the environment surrounding a vehicle, stereo cameras are used to understand the surrounding environment in detail due to their ability to acquire color information and high measurement density, and various methods for understanding the environment have been proposed [1]-[6].

The methods for recognizing the environment using stereo cameras include estimating the road surface area and extracting obstacle regions using geometric conditions and 3D point clouds, and detecting the location of obstacles based on the color information from a monocular camera. Methods for estimating the road surface area and extracting obstacles include the projection matrix method [1] and the UV-disparity method [2][3]. A method using projection matrix cannot estimate the road surface correctly when the road surface cannot be approximated by a single plane, such as in an environment where the slope changes in the middle of the road. A method using UV-disparity has been proposed to estimate the road surface area and extract obstacles to understand the environment. However, this method assumes that the roll angle of the camera is 0° , so it cannot estimate the road surface correctly when the camera tilts. There have been many studies on using deep learning [4]-[6] as a method to understand the environment by detecting obstacles based only on color

information in images. However, deep learning suffers from the difficulty of adapting to environments that differ from the data used for training, and from low explainability, i.e., the inability to explain the cause of failures when they occur.

All of the above methods use regular stereo camera, and the narrow field of view of the range image sensor results in a narrow measurement range. Sakuda et al. proposed a method for obstacle detection and height estimation corresponding to changes in slopes using a fisheye stereo camera with a wide range of measurement [7, 8]. However, this method sometimes incorrectly detects road surfaces as obstacles because it uses only distance information when there is a measurement error. Therefore, in this study, we propose a method to detect only obstacles without false detection of road surfaces by considering intensity information. The obstacle region extraction is performed again by focusing on the area where the edge of the obstacle overlaps with the obstacle region obtained by disparity information and comparing the intensity information in the surrounding area.

II. PROPOSED METHOD

A. Outline of Proposed Method

Fig. 1 shows an overview of the proposed method. The method is divided into three main areas: 3D measurement, road surface plane estimation, and obstacle classification. 3D measurement is performed using a pseudo-bilateral filter [9]. Road surface plane estimation involves preprocessing, fitting of multiple planes, and extraction of the road surface area. Obstacle classification involves extraction of obstacle regions, statistical processing, and estimation of the height of each obstacle. In this study, we propose an algorithm that considers intensity information in addition to disparity for extracting obstacle regions.

B. 3D Measurement using Pseudo-Bilateral Filters

In this 3D measurement, the 3D information obtained by the area-based binocular stereo camera is fused with the feature point-based 3D information obtained by Structure from Motion (SfM) to achieve a dense and accurate range image measurement. The weighting is based on the distance between the pixel of interest and its surrounding feature points, and the difference between the disparity of the pixel of interest and the disparity of its surrounding feature points. The pseudo-bilateral filter performs SfM on three pairs of images from a total of four images obtained before and after the motion from the left and

¹ Precision Engineering Course, School of Science and Engineering, Chuo University, 1-13-27 Kasuga, Bunkyo-ku, Tokyo, Japan
(Corresponding author: chikugo@sensor.mech.chuo-u.ac.jp)

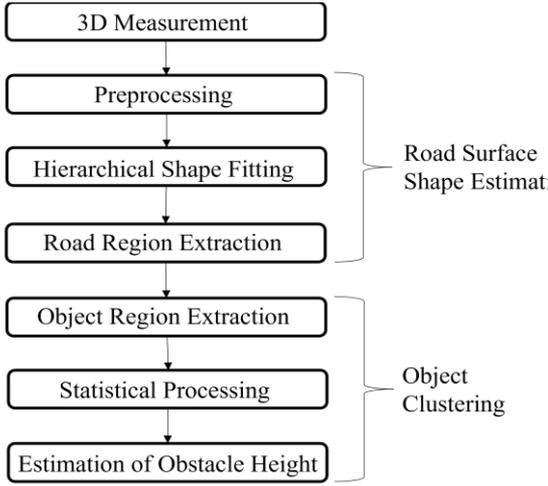


Figure 3. Flow of proposed method

right cameras to obtain sparse but highly accurate 3D information. The orientation of the baseline length differs for each pair. Since the distance accuracy varies greatly depending on the orientation of the baseline length, we weight each SfM pair considering the orientation of the baseline length in addition to the weights described above [9].

Fig. 2 shows the conversion from a fisheye image to an equirectangular image. 3D measurement using a fisheye camera is adversely affected by the distortion inherent in fisheye images. To reduce this effect, the image is converted to an equirectangular image before measurement [10].

C. Estimation of Road Surface Plane

In road surface plane estimation, the road surface plane parameters are obtained by segmenting the disparity images measured in Section II (B). As shown in Fig. 3, limited expansion is performed multiple times with large values of vertical coordinates, and the areas are compared. The area with the largest area because of the comparison is the road surface area [7, 8].

D. Extraction of Obstacle Regions using Disparity

The obstacle region is extracted from the road surface planes obtained in Section II (C). The road surface may not be estimated at the points corresponding to the obstacle region. Therefore, the plane parameters in such areas are estimated using the average of the plane parameters in the road surface area near the obstacle region. Points where the difference in disparity from the estimated road surface is greater than a threshold value are considered obstacles. A binary image is generated, where 1 is the location where the obstacle exists and 0 is the location where the obstacle does not exist. Hereafter, this image is called a candidate point image.

E. Extraction of Obstacle Regions using Intensity

As described in Section I, the conventional method uses only distance information, which may result in incorrect detection of road surfaces when there is a measurement error. Therefore, an algorithm for obstacle region extraction using intensity information is added to the algorithm for obstacle

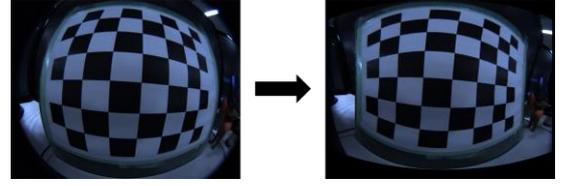


Figure 1. Equirectangular image

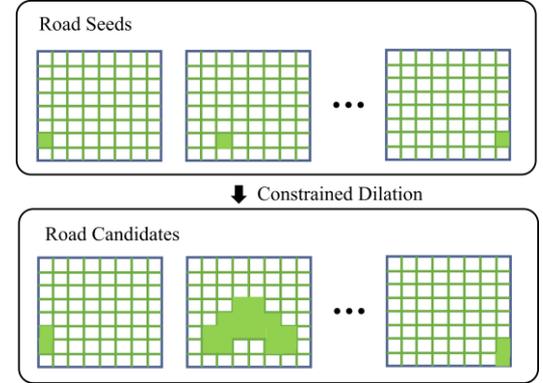


Figure 2. Road area extraction in disparity image

region extraction using disparity information described in Section II (D). A flowchart of this algorithm is shown in Fig. 4. The additional algorithm focuses on the overlapped region between the edge of the obstacle and the obstacle region obtained from the disparity information. If the overlapped area is small, it is removed as noise by the statistical process in Section II (F). Therefore, the overlapped region is surrounded by a bounding box, and the intensity is compared within the bounding box. Then, the obstacle region is extracted again. The procedure is shown below.

- Step 1) Edges are extracted from the equirectangular image measured in Section II (B). First, a bilateral filter is applied to the equirectangular image to smooth it. Next, a Laplacian filter is applied to extract the edges of obstacles while suppressing the extraction of road surface edges by binarization and expansion.
- Step 2) The area where the obstacle region obtained in Section II (D) and the edge obtained in Step 1 overlap is considered likely to be an obstacle. Therefore, the area where they overlap is the obstacle region.
- Step 3) Considering connectivity, the obstacle region obtained in Step 2 is enclosed by a bounding box. However, when there is a wall and a person, as shown on the left in Fig. 5, the obstacle region obtained in Step 2 is surrounded by a bounding box, as shown on the right in Fig. 5, because the processing is performed on a binary image. Then, if the intensity of an obstacle such as a wall or a person is similar to that of the road surface, the road surface becomes an obstacle region. Therefore, the image is divided into 64 segments and the obstacle region is surrounded by the bounding box. The number of segments was determined experimentally.

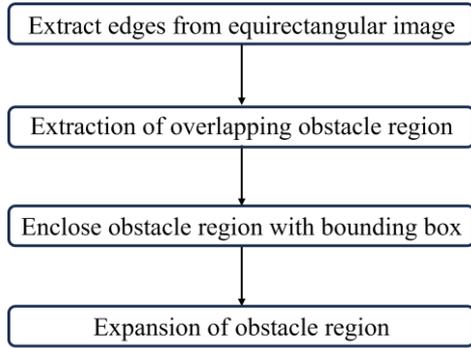


Figure 6. Flowchart of obstacle detection by intensity

- Step 4) The intensity is compared within each bounding box. The intensity of the obstacle region obtained in Step 2 is compared with the intensity in the surrounding area. If they are similar, the area is defined as the obstacle region. The surrounding area is an obstacle region if the difference in intensity between the obstacle region and its surrounding region obtained in Step 2 is within ± 5 . This intensity threshold was determined experimentally.

F. Statistical Processing

Clustering is performed on the images obtained in Section II (E), where the connected regions are considered as one class. However, since the distance from the camera is not considered, they are clustered as the same obstacle due to occlusion and so on. Therefore, the classes are divided again by considering the distance from the camera. For each class, if there are more than a threshold number of points with similar distance values, the region is a class. If the similar distance points are less than the threshold, the class is considered as noise and removed.

G. Obstacle Height Estimation

Calculate the height of each class as a single obstacle.

$$H = \frac{|aX + bZ + cY - B|}{\sqrt{a^2 + b^2 + c^2}} \quad (1)$$

is calculated at the points included in the class, and the maximum value is the obstacle height for that class. Here, (a, b, c) are the planar parameters, H is the height of the obstacle, (X, Y, Z) are the 3D coordinates, and B is the baseline length of the fisheye stereo camera.

III. EXPERIMENTS FOR ACCURACY EVALUATION

A. Experimental Conditions

In this experiment, a 0.31m high cardboard was used as an obstacle. We verified that the proposed method could detect only the obstacle without incorrectly detecting the road surface. We also evaluated the percentage of the cardboard as an obstacle was detected, the accuracy of the distance to the cardboard, and the accuracy of the height of the cardboard. The experimental environment was a flat environment where the angle of slope did not change. Fig. 6 shows a conceptual diagram of the experiment. The red region represents the

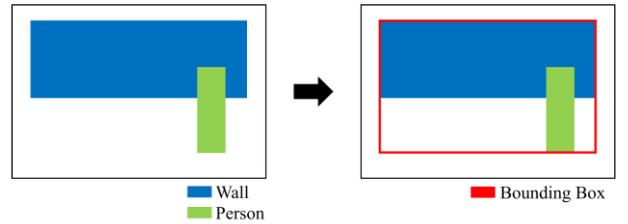


Figure 4. Processing by bounding box

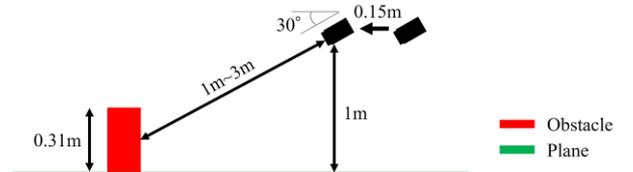


Figure 5. Experimental condition

obstacle and the green region represents the road surface. The black region represents the fisheye stereo camera, which translates 0.15 m in 1 frame from back to front. The camera is tilted at a 30° angle toward the road surface. Fig. 7 shows the appearance of the fisheye stereo camera used in the experiment. The fisheye camera was a Flea3 from FLIR and the fisheye lens was a TVI634M from SPACE. The resolution of the camera is 1328×1048 pixels, the baseline length is 52 mm, and the angle of view is 165° horizontally and 132° vertically. The obstacle was placed so that the center of the obstacle was at azimuth angles of -60° , 0° , and 60° . The camera captured 10 images each at distances of 1m, 2m, and 3m from the obstacle. The distance to the obstacle and the number of captures were the same as for the conventional method [7, 8].

B. Experimental Results

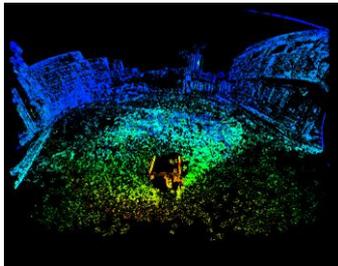
The experimental results are shown in Fig. 8-12 and Table I-II. Fig. 8 shows the equirectangular image and the disparity image, Fig. 9 and Fig. 10 show the clustered images of obstacles and the 3D point cloud display of the obtained 3D information with each obstacle distinguished by color. Fig. 11 shows the average and standard error of the measurement error of the distance between the obstacle and the camera. Fig. 12 shows the average and standard error of the measurement error of the obstacle height. Table I and Table II show the percentage of obstacles detected without false detection of the road surface. The brackets show the percentage of detected obstacles. Fig. 9 and Fig. 10 show that the conventional method can detect obstacles, but it incorrectly detects road surfaces in the lower left corner of the image. The proposed method detects the obstacles correctly without incorrectly including the road area. The obstacle clustered in the upper right corner of the image represents a building. The cardboard as the obstacle is not detected in its entirety, and only the edges of the top surface of the cardboard are detected. However, this is not a problem because it is sufficient to detect only a part of the obstacle. Fig. 11 and Fig. 12 show that the proposed method is as accurate as the conventional method in terms of distance and height. The lower accuracy at azimuth 0° compared to other azimuth angles is due to the pseudo-



Figure 8. Fisheye stereo camera



(a) Equirectangular image



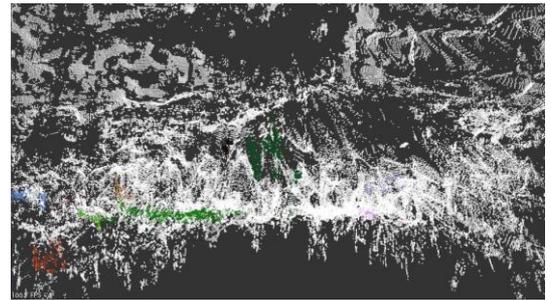
(b) Disparity image

Figure 9. Images with the cardboard at 1 m

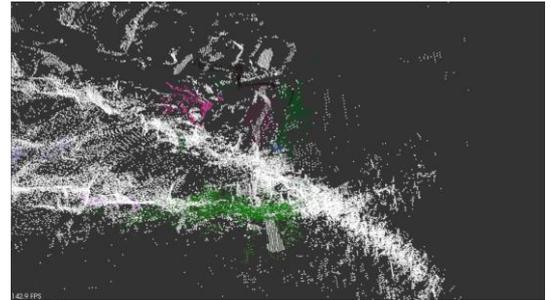
bilateral filter; the epipolar direction of the SfM is along the optical axis, and the measurement accuracy near the center of the image is lower. Table I and Table II show that the conventional method can detect obstacles in almost all situations, but it incorrectly detects the road surface in all situations. The proposed method was able to detect only obstacles in all situations without incorrectly detecting the road surface. However, the proposed method failed to detect obstacles several times in the experiments at 3m and -60° and 0° . Fig. 13 and Fig. 14 show the respective images before clustering. Fig. 13 and Fig. 14 show that the cardboard region was extracted as an obstacle region extraction. However, it is considered that they were removed as noise in the statistical processing described in Section II (F). Also, the processing speed was approximately 0.05 fps. Such a processing speed is insufficient for automatic driving. In future, we aim to improve this processing speed.



(a) Clustering image



(b) Front view of point clouds



(c) Side view of point clouds

Figure 7. Clustering results at 1m in conventional method

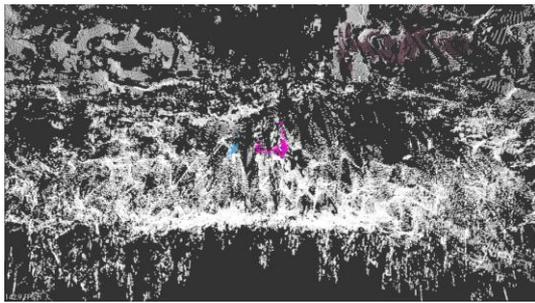
IV. CONCLUSION

In this study, we proposed a method to correctly detect obstacles and estimate their height without incorrectly detecting the road surface using disparity and intensity information obtained from a wide range of dense 3D information. Experiments showed that only obstacles can be detected without incorrectly detecting road surfaces. The proposed method is as accurate as the conventional method in terms of distance and height, even though it only detects obstacles without incorrectly detecting road surfaces.

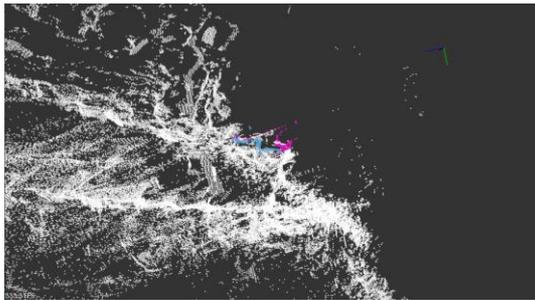
For future works, it is necessary to verify whether the proposed method can detect only obstacles by experiments in complex environments. It is also important to improve the accuracy of 3D measurement using the fisheye stereo camera.



(a) Clustering image



(b) Front view of point clouds

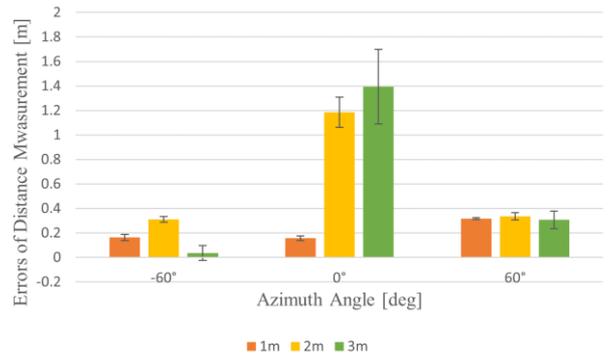


(c) Side view of point clouds

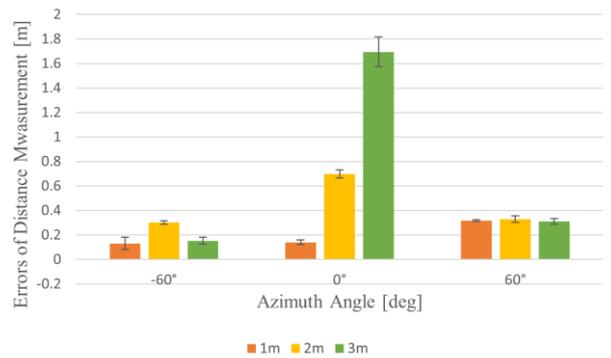
Figure 10. Clustering result at 1m in proposed method

REFERENCES

- [1] A. Seki and M. Okutomi, "Robust Obstacle Detection in General Road Environment Based on Road Extraction and Pose Estimation", 2006 IEEE Intelligent Vehicles Symposium, pp. 437-444, 2006.
- [2] W. Song, Y. Yang, M. Fu, F. Qiu and M. Wang, "Real-Time Obstacles Detection and Status Classification for Collision Warning in a Vehicle Active Safety System", IEEE Transactions on Intelligent Transportation Systems (ITSS), vol. 19, no. 3, pp. 758-773, 2018.
- [3] M. Liu, C. Shan, H. Zhang and Q. Xia, "Stereo Vision Based Road Free Space Detection", 2016 9th International Symposium on Computational Intelligence and Design (ISCID), pp. 272-276, 2016.
- [4] X. Chen, K. Kundu, Y. Zhu, H. Ma, S. Fidler and R. Urtasun, "3D object proposals using stereo imagery for accurate object class detection", Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1259-1272, 2017.
- [5] P. Li, X. Chen and S. Shen, "Stereo R-CNN Based 3D Object Detection for Autonomous Driving", 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 7636-7644, 2019.
- [6] B. Li, W. Ouyang, L. Sheng, X. Zeng and X. Wang, "GS3D: An Efficient 3D Object Detection Framework for Autonomous Driving", 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1019-1028, 2019.

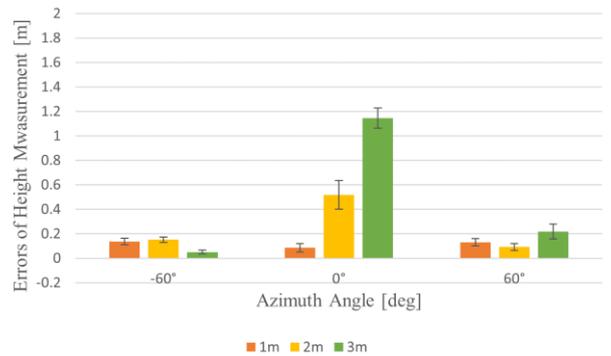


(a) Conventional method

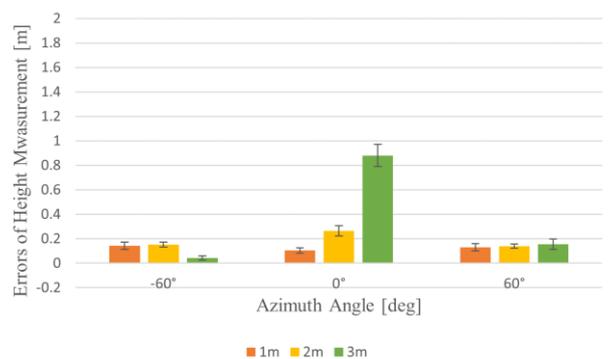


(b) Proposed method

Figure 11. Standard error in distance with the cardboard



(a) Conventional method



(b) Proposed method

Figure 12. Standard error in height with the cardboard

TABLE I. PERCENTAGE OF THE CARDBOARD DETECTED ONLY IN CONVENTIONAL METHOD (PERCENTAGE OF DETECTED OBSTACLES)

Distance [m]	Azimuth angle [deg]		
	-60	0	60
1	0 (100)	0 (100)	0 (100)
2	0 (100)	0 (80)	0 (100)
3	0 (100)	0 (70)	0 (100)

TABLE II. PERCENTAGE OF THE CARDBOARD DETECTED ONLY IN PROPOSED METHOD (PERCENTAGE OF DETECTED OBSTACLES)

Distance [m]	Azimuth angle [deg]		
	-60	0	60
1	100 (100)	100 (100)	100 (100)
2	100 (100)	100 (100)	100 (100)
3	100 (90)	100 (80)	100 (100)

- [7] T. Sakuda, K. Arai, S. Pathak and K. Umeda, "Estimation of Road Surface Shape and Object Height Focusing on the Division Scale in Disparity Image Using Fisheye Stereo Camera", 2023 IEEE/SICE International Symposium on System Integration (SII2023), pp. 1-5, 2023.
- [8] T. Sakuda, H. Chikugo, K. Arai, S. Pathak, and K. Umeda, "Estimation of Road Surface Plane and Object Height Focusing on the Division Scale in Disparity Image Using Fisheye Stereo Camera", J. Robot. Mechatron., Vol.35 No.5, 2023 to be published.
- [9] H. Iida, Y. Ji, K. Umeda, A. Ohashi, D. Fukuda, S. Kaneko, J. Murayama, and Y. Uchida., "High-accuracy Range Image Generation by Fusing Binocular and Motion Stereo Using Fisheye Stereo Camera", Proc. Of 2020 IEEE/SICE International Symposium on System Integration (SII2020), pp.343-348, 2020.
- [10] A. Ohashi, F. Yamano, G. Masuyama, K. Umeda, D. Fukuda, K. Irie, S. Kaneko, J. Murayama, and Y. Uchida, "Stereo Rectification for Equirectangular Images", Proc. Of 2017 IEEE/SICE International Symposium on System Integration (SII2017), pp. 535-540, 2017.



(a) Equirectangular image



(b) Candidate point image

Figure 13. Images with the cardboard at 3 m and 0°



(a) Equirectangular image



(b) Candidate point image

Figure 14. Images with the cardboard at 3 m and -60°