

Paper:

Estimation of Road Surface Plane and Object Height Focusing on the Division Scale in Disparity Image Using Fisheye Stereo Camera

Tomoyu Sakuda*, Hikaru Chikugo*, Kento Arai*, Sarthak Pathak**, and Kazunori Umeda**

*Precision Engineering Course, Graduate School of Science and Engineering, Chuo University
1-13-27 Kasuga, Bunkyo-ku, Tokyo 112-8551, Japan
E-mail: {sakuda, chikugo, arai}@sensor.mech.chuo-u.ac.jp

**Department of Precision Mechanics, Faculty of Science and Engineering, Chuo University
1-13-27 Kasuga, Bunkyo-ku, Tokyo 112-8551, Japan
E-mail: {pathak, umeda}@mech.chuo-u.ac.jp

[Received March 13, 2023; accepted July 19, 2023]

In this paper, we propose a novel algorithm for estimating road surface shapes and object heights using a fisheye stereo camera. Environmental recognition is an important task for advanced driver-assistance systems. However, previous studies have only achieved narrow measurement ranges owing to sensor restrictions. Moreover, the previous approaches cannot be used in environments where the slope changes because they assume inflexible constraints on the road surfaces. We use a fisheye stereo camera capable of measuring wide and dense 3D information and design a novel algorithm by focusing on the degree of division in a disparity image to overcome these defects. Experiments show that our method can detect an object in various environments, including those with inclined road surfaces.

Keywords: fisheye stereo camera, equirectangular image, drivable region, autonomous vehicle, object detection

1. Introduction

In recent years, active research and development have been conducted on driver-assistance systems. Driver-assistance systems use information obtained from various range sensors (such as sonar and radar sensors) to identify the surrounding conditions and make driving-assistance decisions. In particular, 3D light detection and ranging (LiDAR) and stereo cameras are commonly used to understand the environment around a vehicle in detail, owing to their wide measurement ranges. 3D LiDAR systems have been actively studied owing to their high measurement accuracy, and various environmental measurement methods have been proposed [1–4]. However, 3D LiDAR is expensive. Moreover, it is unable to view color information and its low resolution makes it difficult to detect small objects. In contrast, stereo cameras have been

studied as practical distance sensors because they can also acquire color information, are inexpensive, and have high measurement densities [5–16]. Methods for understanding the surrounding environment can be divided into two categories: ones that estimate the road surface region and determine the locations of obstacles, and ones that detect objects based on color information. One framework for estimating road surface planes and objects is to use UV-disparity [5, 6]. This method compresses information in the horizontal and vertical directions of the image to predict the 3D shape of the environment with a low computational cost. However, it assumes that the roll angle of the camera is 0°; this causes problems when the car is tilted. Seki et al. proposed a method based on a projection matrix [7]. However, this method assumed a single plane, which caused problems when the slope changed. Other studies [8–14] have used deep learning as a framework for object detection. However, deep learning depends on the environment at the time of training. As such, the accuracy is not guaranteed for environments differing from the training data. Meanwhile, stereo cameras generally have a narrow field of view.

In view of the above, in this study, a fisheye stereo camera is used as a distance sensor with a wide viewing angle [17]. Structure from motion (SfM) and epipolar-plane image analyses using a fisheye camera along with a monocular camera have been used as 3D measurement methods [18, 19]. However, these methods face challenges, such as an insufficient distance measurement accuracy. In addition, these analyses are performed with a monocular camera. In contrast, a pseudo-bilateral filter is used in a 3D measurement method herein based on a fisheye stereo camera [20]. The road surface plane is estimated by considering the fact that the estimated plane changes depending on the size of the segmentation of the disparity image. Correspondingly, we propose a method for understanding the environment without relying on deep learning. This method can be applied to a wide range of environments with changes in slopes.



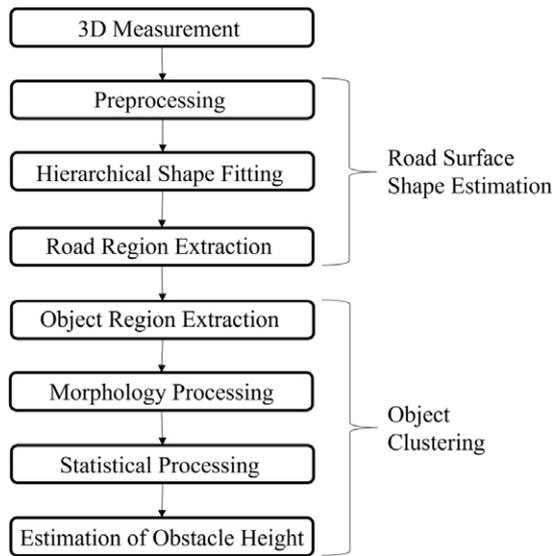


Fig. 1. Flow of proposed method.

2. Outline of the Proposed Method

An overview of the proposed method is shown in Fig. 1. The method consists of three main stages: 3D measurement, road surface plane estimation, and object extraction. For the 3D measurement, we use a pseudo-bilateral filter, and for the road surface plane estimation, we propose a method able to adapt to changes in slope by fitting multiple planes. For object extraction, we perform clustering based on frequency values in the distance after morphological processing.

3. 3D Measurement Using Pseudo-Bilateral Filter

In the 3D measurement stage, the 3D information obtained by an area-based binocular stereo camera is fused with the 3D information obtained by an SfM analysis to obtain a dense and accurate distance image measurement. The weighting is based on the distance between the pixel of interest and its surrounding feature points and that between the disparity of the pixel of interest and that of its surrounding feature points. As this weighting is similar to that of a bilateral filter [21], it is called a pseudo-bilateral filter. The pseudo-bilateral filter performs an SfM analysis on three pairs of images from a total of four images obtained before and after the motion from the left and right cameras to obtain sparse but highly accurate 3D information. The orientation of the baseline length differs for each pair. As the distance accuracy varies greatly depending on the orientation of the baseline length, we weight each SfM pair while considering the orientation of the baseline length in addition to the weights described above. In the 3D measurement method using the pseudo-bilateral filter, the fisheye image is converted to an equirectangular image with reduced distortion as shown in Fig. 2. This helps

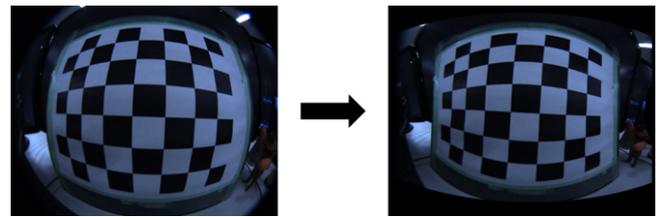


Fig. 2. Equirectangular image.

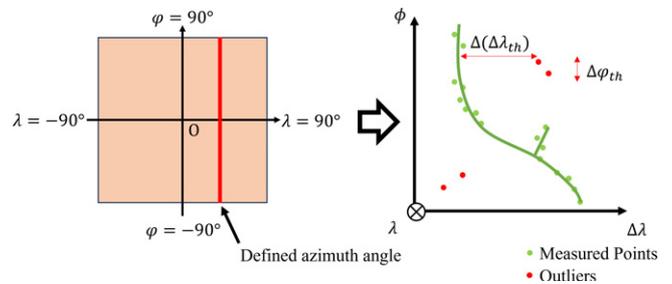


Fig. 3. Disparity-elevation angle space.

avoid adverse effects on the 3D measurements from the distortion inherent in the fisheye image. By processing in the disparity space of the fisheye image, we can ignore the error stretching that occurs when the image is converted to a real space.

4. Estimation of Road Surface Plane

4.1. Preprocessing

In this study, outlier removal is performed as a preprocessing step. The data obtained from the fisheye stereo camera contains outliers which adversely affect the plane estimation. As the difference between the outliers and surrounding data is large, the outliers can be easily discriminated based on the density of the point cloud. However, the calculation of the density of a dense 3D point cloud measured with a fisheye stereo camera is time-consuming. Therefore, to reduce the computational cost, we reduce the number of dimensions. A single azimuth angle is defined and a two-dimensional space represented by the elevation angle ϕ and disparity $\Delta\lambda$ is considered, as shown in Fig. 3. The line indicated by the arrow in Fig. 3 represents the azimuth angle, the non-curvilinear dots represent outliers, and the curved dots represent normal measurement points. Fig. 3 shows that the outliers are mainly in the region of a narrow elevation angle within the region where the difference between the disparity and surrounding points is large. A threshold value $\Delta(\Delta\lambda)_{th}$ is set for the disparity in the space of the elevation angle and disparity at the defined azimuth angle. We search for elevation angles from -90° to 90° . When the disparity changes by more than the threshold value, we examine the width in the elevation direction of the region where the change occurs. If it is less than the threshold $\Delta\phi_{th}$, it is removed as an outlier.

4.2. Plane Estimation

When estimating the plane, the presence of objects or changes in the slope of the road surface can cause the plane to be incorrectly fitted to the point cloud. Therefore, we consider dividing the region for the plane estimation. If we divide the disparity image into smaller regions, the possibility that both objects and the road surface exist in the same area is reduced, and it becomes easier to approximate the road surface shape as a plane. However, if the region to be estimated is excessively small, it will be affected by errors in the measurement of the point cloud. Therefore, to estimate the plane with an appropriate division size, we change the size of the division in a step-by-step manner and perform the plane estimation. We assume that the plane matching the estimated plane parameters before and after the change is the plane that well-fits the point cloud. In addition, there is no limitation on the shape of an object to be detected. A model of the road surface plane in the disparity space is computed to estimate the road surface plane in the disparity space. The equation for the road surface plane in the 3D information (X, Y, Z) in the real space is as follows:

$$a'X + b'Y + c'Z = 1. \quad \dots \quad (1)$$

This equation is converted to a model in the disparity space as follows:

$$\begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = D \begin{pmatrix} \cos \varphi \sin \lambda \\ \sin \varphi \\ \cos \varphi \cos \lambda \end{pmatrix}, \quad \dots \quad (2)$$

$$D = \frac{B \times \cos(\lambda + \Delta\lambda)}{\sin(\Delta\lambda) \times \cos \varphi}. \quad \dots \quad (3)$$

Here, a' , b' , and c' are the road plane parameters in real space, φ is the elevation angle, λ is the azimuth angle, D is the Euclidean distance from the camera, B is the baseline length of the fisheye stereo camera, and $\Delta\lambda$ is the disparity. D is obtained from triangulation in the equirectangular images [17]. After the transformation, $\Delta\lambda$ is sufficiently small to be estimated as follows:

$$\cos(\lambda + \Delta\lambda) \approx \cos \lambda. \quad \dots \quad (4)$$

Therefore, the equation for the road surface plane is as follows:

$$\Delta\lambda = a \cos \varphi \sin \lambda + b \tan \varphi \cos \lambda + c \cos^2 \lambda. \quad \dots \quad (5)$$

In addition, a , b , and c are the road plane parameters in the disparity space, given as follows:

$$\begin{pmatrix} a' \\ b' \\ c' \end{pmatrix} = \frac{1}{B} \begin{pmatrix} a \\ b \\ c \end{pmatrix}. \quad \dots \quad (6)$$

This model of the road surface plane in the disparity space is used to estimate the road surface plane. The procedure is shown as follows:

- i) Divide the disparity image into blocks of square regions. The length of one side of each square is 2^n pixels ($n \in \mathbb{N}$).

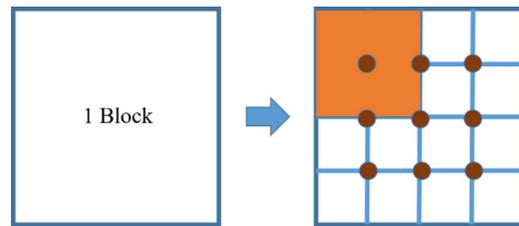


Fig. 4. Split of a block in disparity image.

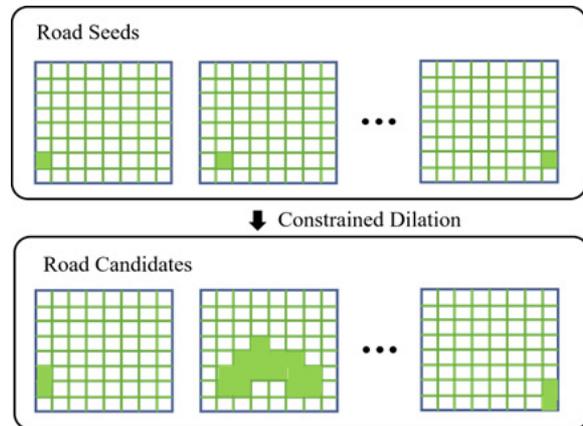


Fig. 5. Road region extraction in disparity image.

- ii) Apply a median filter of size $2^{n-1} \times 2^{n-1}$ (shown in the area centered on the upper left dot) to each block in nine locations, resulting in sections based on three horizontal lines and three vertical lines drawn at equal intervals in the center, as shown in Fig. 4.
- iii) Calculate the plane parameters a , b , and c using the least-squares method based on the values obtained in step ii).
- iv) Halve the length of each block and perform steps ii)–iii), and then step v).
- v) Calculate the similarity of the plane parameters before and after the division; perform step vi) if the similarity is high, and step vii) if the similarity is low.
- vi) Assume that the obtained plane parameters are estimated in the appropriate segmentation and determine the plane parameters.
- vii) Repeat step iv). If the number of attempts exceeds a certain number, the planar parameters cannot be estimated at that location (e.g., the boundary between an object and road surface).

In step iii), Eq. (3) is used to calculate the plane parameters. In this way, the obtained set of planes contains planes fitted to objects. Therefore, the plane corresponding to an object can be removed by focusing on the change in slope in the obtained plane group. As shown in Fig. 5, we prepare several seed points; then, we expand the domain to the region where the change in slope is small at each seed

point. The seed point with the largest region is selected from the obtained regions; the region obtained by extending the seed point is defined as the road surface region.

5. Object Extraction and Height Estimation

5.1. Morphology Processing

The road surface plane obtained as described in Section 4 is used to extract object candidates from the point cloud. To define the road surface in the region where the plane corresponding to the object is estimated, we take the average of the plane parameters in the nearby road surface region and interpolate the plane parameters in the object region. A threshold is set for the absolute value of the difference in disparity with the obtained plane. The points above this threshold are considered as objects. Next, a binary image is generated, with 1 representing the location where the object exists and 0 representing the location where the object does not exist. The image is then subjected to an opening and closing process to remove small regions and holes. Then, clustering is performed considering the connectivity.

5.2. Statistical Processing

As described in Section 5.1, clustering is performed on the binary images. However, because the clustering is performed in a compressed binary space, different objects may be recognized as the same class when occlusion occurs. In addition, erroneous measurement points between a plane and objects may be extracted as objects. Therefore, a frequency distribution of distances is created for each class. False measurement points between objects and the road surface are removed by setting a threshold value for the frequency (as they have lower frequencies). In addition, the same object can be assumed to be continuously distributed at similar distances; thus, if intermittent regions exist, each continuous region can be reassigned to a different class as a different object.

5.3. Estimation of Object Height

The height of each class as a single object is calculated as follows:

$$\text{Height} = \frac{|aX + bZ + cY - B|}{\sqrt{a^2 + b^2 + c^2}} \dots \dots \dots (7)$$

The above is obtained from a calculation as follows:

$$\text{Height} = \frac{|a'X + b'Y + c'Z - 1|}{\sqrt{a'^2 + b'^2 + c'^2}} \dots \dots \dots (8)$$

The above calculations are combined with Eq. (6) based on the relationship between the plane ($aX + bY + cZ = 0$) and distance of the point (X, Y, Z) in three dimensions. For each class, the point farthest from the planar area is selected and used in Eqs. (3) and (8). In this way, the distance to the object and height of the object are calculated.

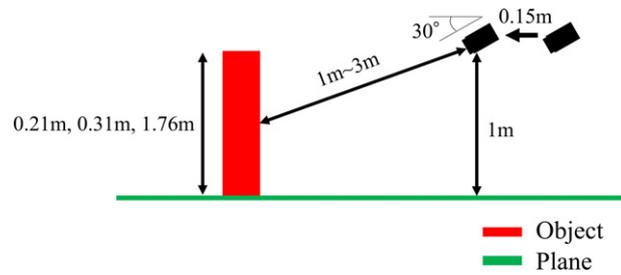


Fig. 6. Experimental conditions.



Fig. 7. Fisheye stereo camera.

6. Experiments for Accuracy Evaluation

6.1. Experiments in Real Environment

6.1.1. Experiments with Objects

In this experiment, we first verified whether objects were correctly extracted on a road surface with no changes in slope. Moreover, the accuracy of the height and distance to the objects were evaluated. The objects were a 0.31-m-high piece of cardboard, a 1.76-m-high human, and a 0.21-m-high plastic bottle. A schematic diagram of the environment is shown in Fig. 6. In Fig. 6, the area on the left side represents the object, the line represents the plane, and the area on the right side represents the camera. The camera was moved from back to front by 0.15 m for the SfM analysis. Fig. 7 shows the fisheye stereo cameras used for our experiments. The cameras were two FLIR Flea3 cameras, each equipped with a SPACE TV1634M fisheye lens. The resolution of both cameras was $1,328 \times 1,048$ pixels and the stereo baseline was 52 mm. The angle of view was 165° in the horizontal direction and 132° in the vertical direction. The objects were measured at azimuth angles of -60° , 0° , and 60° , and 10 shots were taken at distances of 1 m, 2 m, and 3 m from the camera, respectively. The fisheye stereo camera was set at a height of 1 m and tilted at a pitch angle of 30° . The camera was set to move by 0.15 m. The preprocessing threshold was set to 0.004 rad in the disparity direction and 15 pixels in the elevation direction. The similarity value used for the road plane was the inner product of the normalized plane parameters with a threshold of 0.98. For restricted expansion, the inner product of the normalized planar parameters needed to be greater

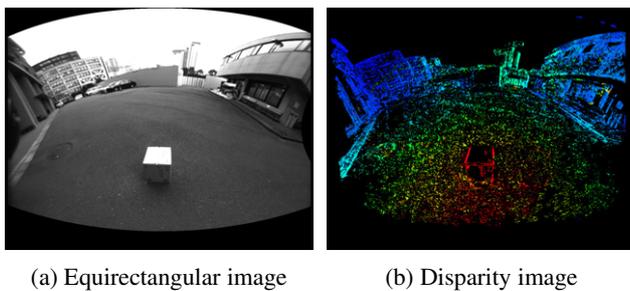
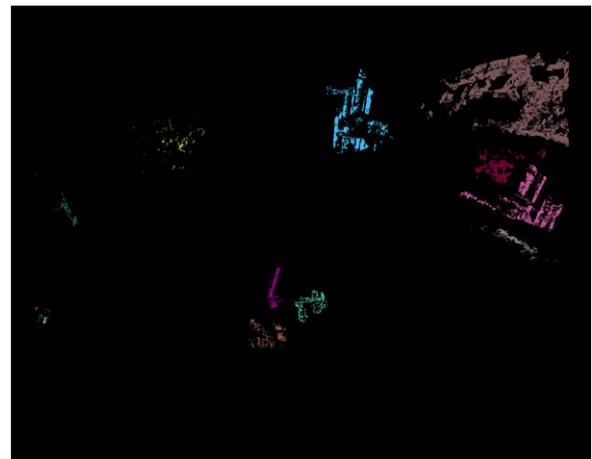


Fig. 8. Images with the cardboard at 1 m.

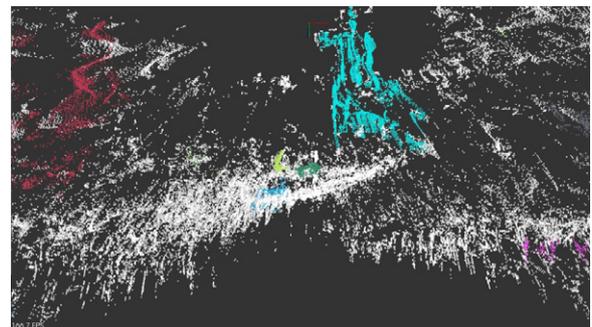
than or equal to the threshold of 0.95 and the average distance between adjacent lines in the image needed to be less than 0.35 m. For clustering, the object region was defined as the region with points with frequency values of 5 or more. The object region was divided into separate objects if the distance between them was greater than 0.3 m. These thresholds were determined experimentally.

6.1.2. Results of Experiments with Objects

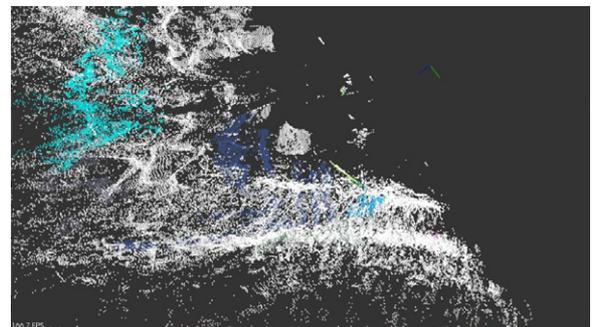
The experimental results are shown in **Figs. 8–17**. **Figs. 8–10** show the results for the cardboard, **Figs. 11–13** for the human, and **Figs. 14–17** for plastic bottles. **Figs. 8, 11, 14, and 15** show the equirectangular and disparity images. **Figs. 9, 12, and 16** show the colored points of the clustered objects. **Figs. 10, 13, and 17** show the measurement errors of the object heights and distances to the object. **Figs. 9(a), 12(a), and 16(a)** show that the target object is extracted. However, objects are incorrectly extracted in other locations. On an actual road, the 3D point cloud of the road surface should be a continuous plane. However, **Figs. 9(c), 12(c), and 16(c)** show that the extracted point cloud of the road surface is not flat. It is thought that the road surface plane could not be estimated correctly owing to errors in the 3D measurements. This is thought to be the reason for the incorrect extractions at locations other than objects. **Figs. 10, 13, and 17** show that the absolute values of the measurement errors of the distance and height at azimuth angles of -60° and 60° are similar to those at the azimuth angle of 0° in many situations. This indicates that the distance and height accuracy does not deteriorate at the edge of the image. However, the maximum error in the distance is approximately 1.0 m and that in the height is approximately 0.7 m; these are not small. This is thought to be owing to errors in the 3D measurement as mentioned above. The error in the height measurement at 1 m when a human is considered as an object is higher when the azimuth angle is 0° than when the azimuth angle is -60° or 60° . This is thought to be owing to the inability to measure the human head, as shown in **Fig. 12(a)**. Furthermore, a plastic bottle could not be detected at 3 m. The disparity image taken as shown in **Fig. 15** shows that the plastic bottle can be measured. This is thought that the plastic bottle was removed as noise by the statistical processing. A distance accuracy of approximately 1.5 m at 50 m and processing speed of approximately 30 fps are required for practical



(a) Clustering image with the cardboard at 1 m



(b) Front view of point clouds with the cardboard at 1 m



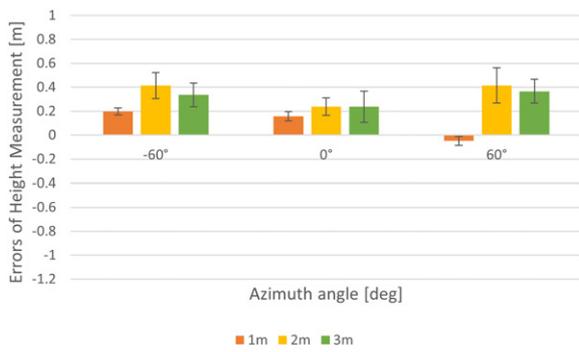
(c) Side view of point clouds with the cardboard at 1 m

Fig. 9. Clustering result with the cardboard.

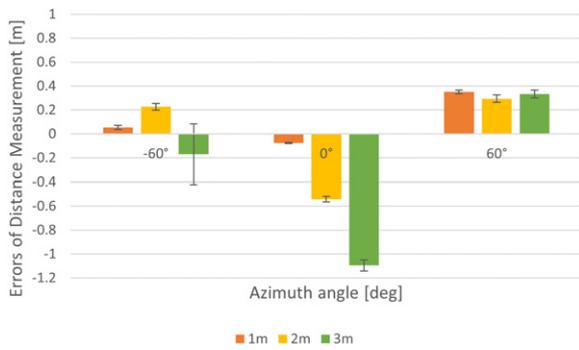
use [22–25]. However, in certain situations, the measurement error reaches up to 3 m and the processing speed is only approximately 0.05 fps. Therefore, a faster processing speed and improved distance measurement accuracy are needed for practical use. This will be considered in future work.

6.1.3. Experiments with Slope

In this experiment, we verified whether the system could recognize a slope as road surface without mistakenly detecting it as an object in an environment where the slope changes, as shown in **Fig. 18**. The position and orientation of the fisheye stereo camera and threshold values were the same as in Section 6.1.1.

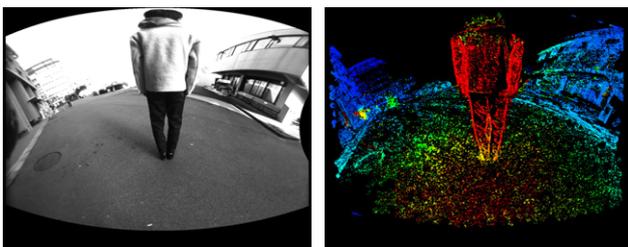


(a) Height of the cardboard



(b) Distance of the cardboard

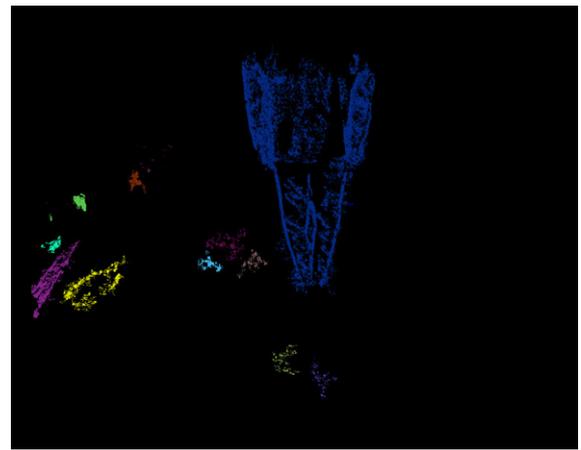
Fig. 10. Measurement error with the cardboard.



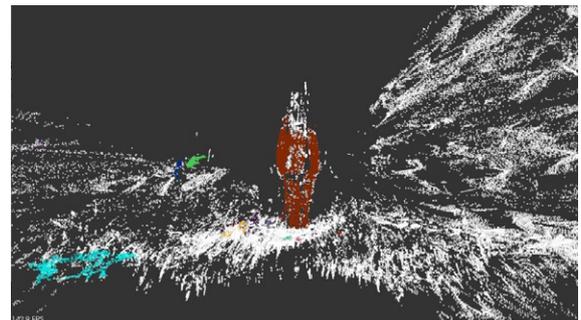
(a) Equirectangular image

(b) Disparity image

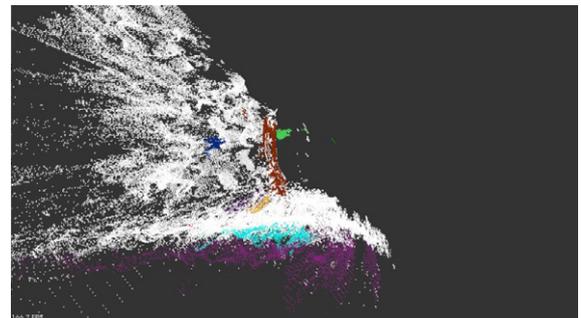
Fig. 11. Images with the human at 1 m.



(a) Clustering image with the human at 1 m



(b) Front view of point clouds with the human at 1 m



(c) Side view of point clouds with the human at 1 m

Fig. 12. Clustering result with the human.

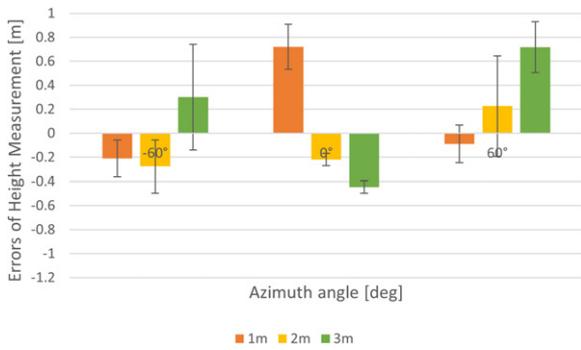
6.1.4. Results of Experiments with Slope

The experimental results are shown in **Figs. 19** and **20**. **Fig. 19** shows the equirectangular and disparity images, and **Fig. 20** shows the clustered objects as colored points. **Fig. 20** shows that the road surface is not detected incorrectly as an object; moreover, the tree on the right side of **Fig. 19(a)** and fence on the left side of the image are correctly detected as objects. This demonstrated the usefulness of the system in managing an environment where the slope changes.

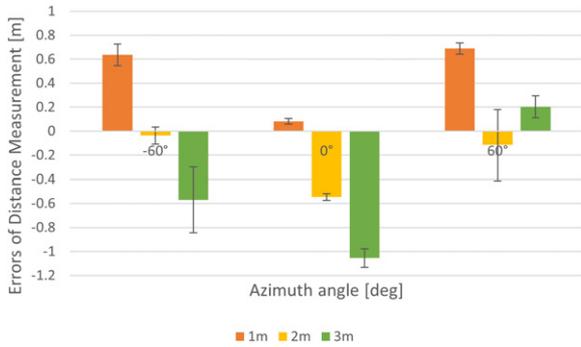
6.2. Simulation Experiments

In this section, simulation experiments were conducted for various environments that were difficult to prepare in a real environment. As shown in **Fig. 21**, the experimental environments included an environment with a variable slope, one with a hole, one with occlusion, and one with

an assumed parking space. The parking space was 3 m wide and 6 m deep and was surrounded by a 1 m wall. In this simulation experiment, we used the 3DCG software Blender [a]. In Blender, the render engine was Cycles, and the lens was a panoramic equirectangular cylindrical view. The horizontal and vertical resolution were set to 1,328 and 1,048 pixels, respectively. The horizontal and vertical angles of view were set to 180° and 142°, respectively. The size of the image sensor in the horizontal direction was set to 36 mm and that in the vertical direction was set automatically. The camera movement was set to 0.15 m as in the experiment described in the previous section, and the object was set at 0.5 m in height. The simulation results are shown in **Figs. 22–25** and **Table 1**. **Figs. 22–25** show an equirectangular image, the

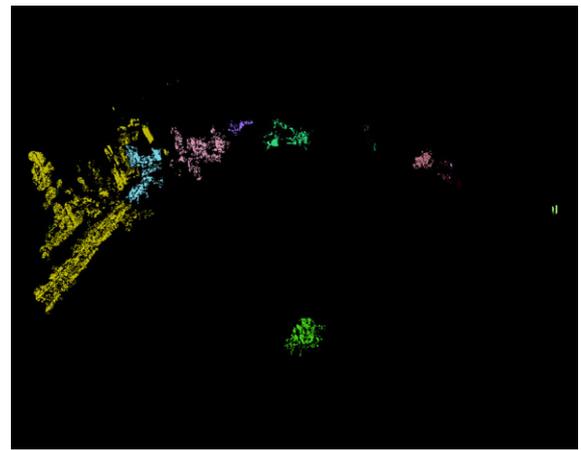


(a) Height of the human

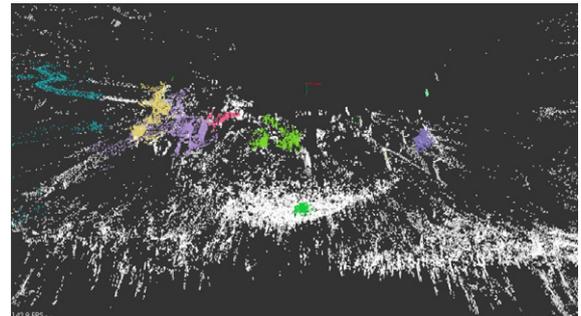


(b) Distance of the human

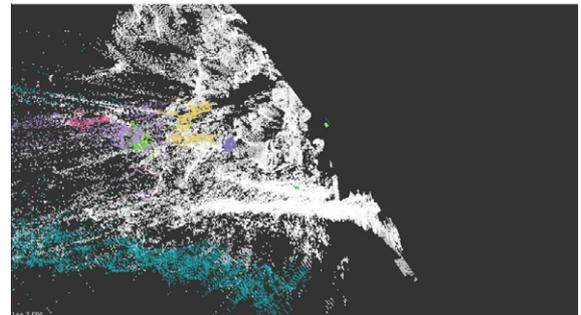
Fig. 13. Measurement error with the human.



(a) Clustering image with the plastic bottle at 1 m

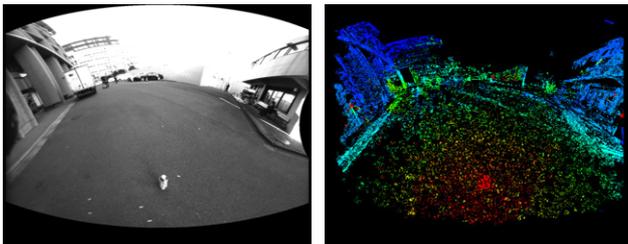


(b) Front view of point clouds with the plastic bottle at 1 m



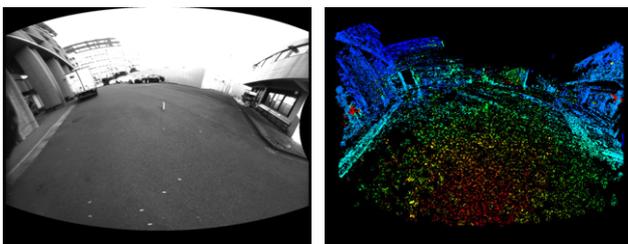
(c) Side view of point clouds with the plastic bottle at 1 m

Fig. 16. Clustering result with the plastic bottle.



(a) Equirectangular image (b) Disparity image

Fig. 14. Images with the plastic bottle at 1 m.

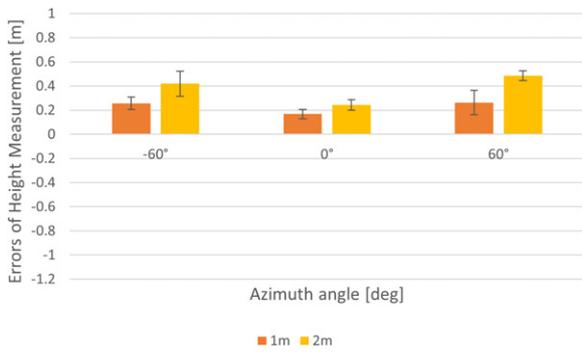


(a) Equirectangular image (b) Disparity image

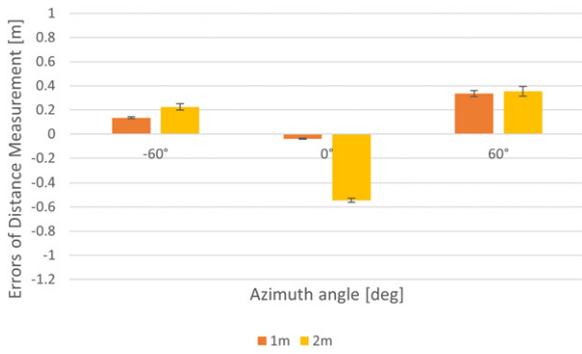
Fig. 15. Images with the plastic bottle at 3 m.

disparity images, and the 3D point clouds colored with the clustered objects in each environment, respectively. The objects are correctly extracted in each environment. Moreover, the objects are not incorrectly extracted when the angle changes. It is thought that the road surface plane estimation described in Section 4 did not extract the loca-

tions with changes in inclination (as shown in **Figs. 22** and **25**) as objects. It is also thought that the statistical process described in Section 5.2 can extract the objects as separate objects even in an environment with occlusions. **Figs. 22(d), 23(d), 24(d), and 25(d)** show that the extracted road surface points are planar, unlike those in the experimental results in Section 6.1. It is considered that the reason why the point cloud of the road surface extracted as described in Section 6.1 was not planar was owing to the 3D measurement error. **Table 1** shows the error in the distance between the object and camera and that for height of the object. These results are more accurate than those from the experiment described in Section 6.1. The 3D point clouds shown in **Figs. 22–25** are correctly extracted. It is thought that the accuracy of the previous



(a) Height of the plastic bottle



(b) Distance of the plastic bottle

Fig. 17. Measurement error with the plastic bottle.

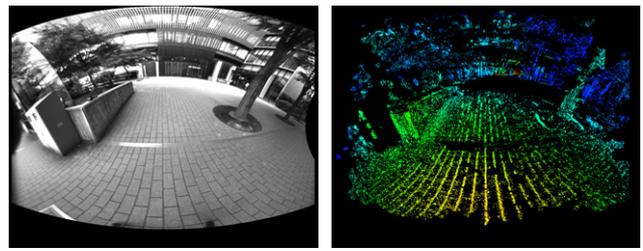


(a) View from nearby



(b) View from afar

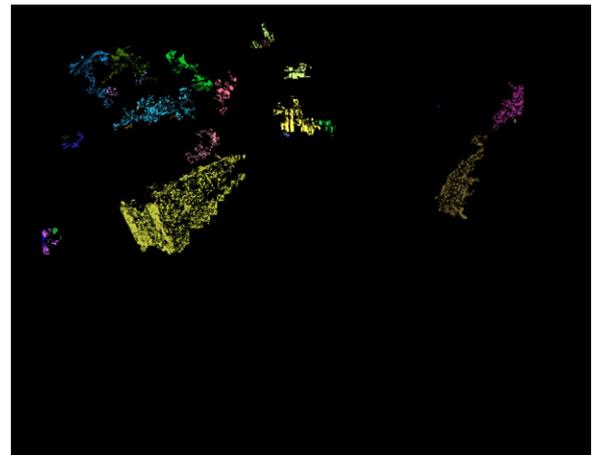
Fig. 18. Environment for an experiment with slope.



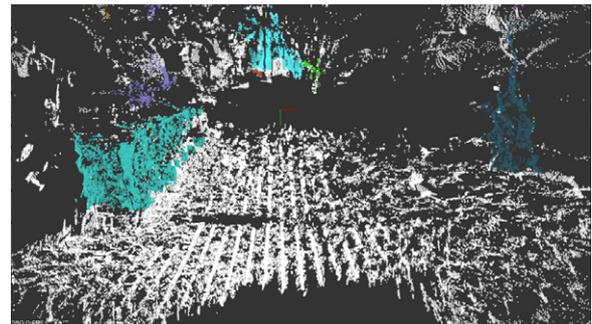
(a) Equirectangular image

(b) Disparity image

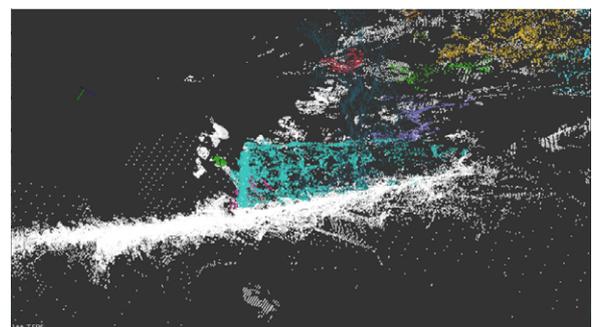
Fig. 19. Images with the slope.



(a) Clustering image with the slope



(b) Front view of point clouds with the slope



(c) Side view of point clouds with the slope

Fig. 20. Clustering result with the slope.

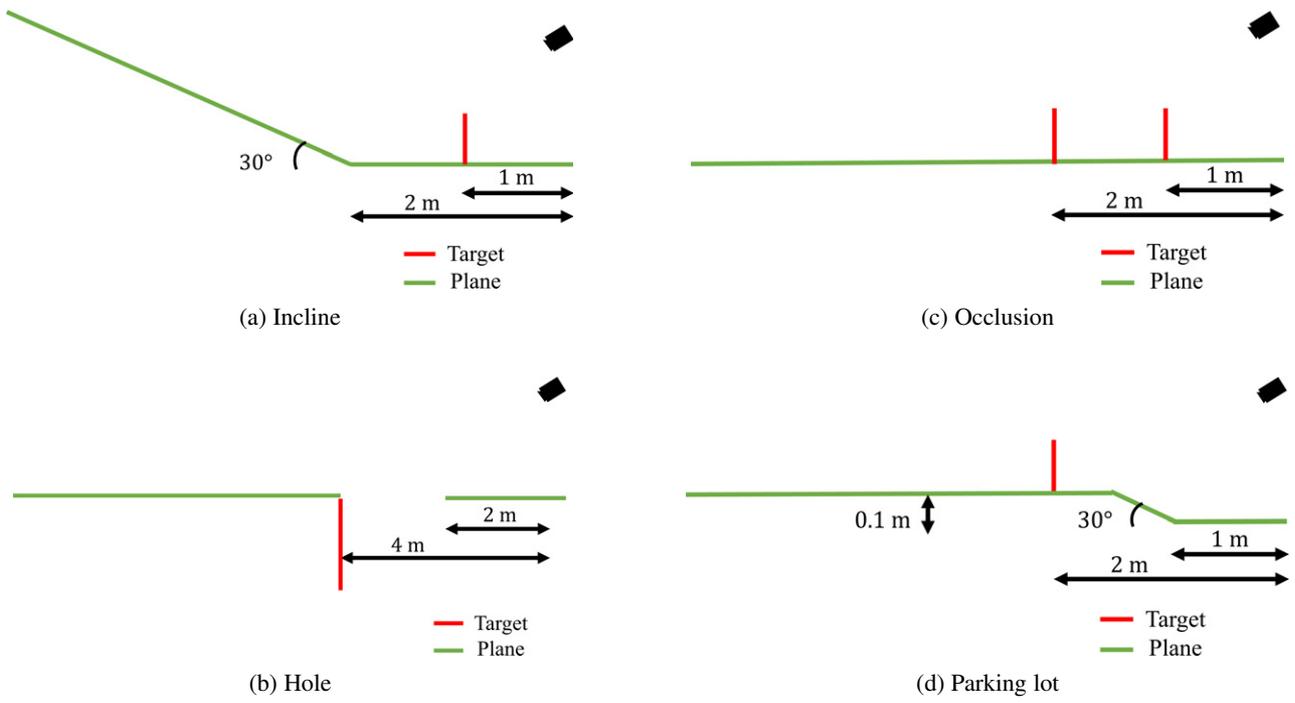
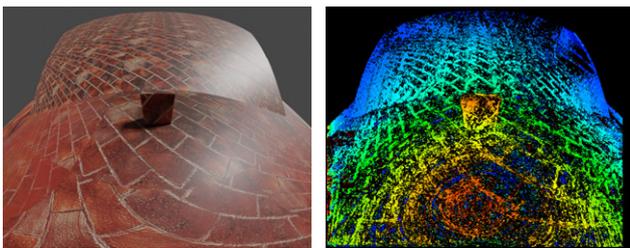
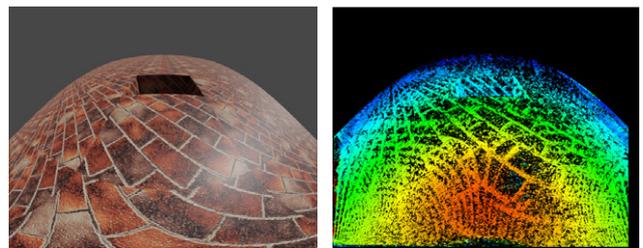


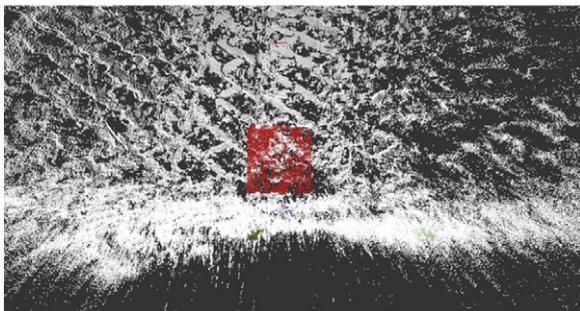
Fig. 21. Simulation conditions.



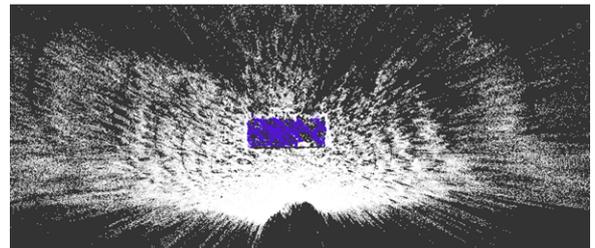
(a) Equirectangular image (b) Disparity image



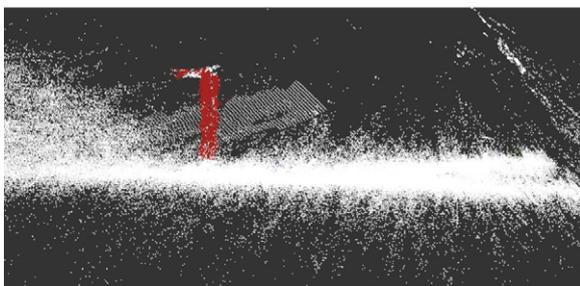
(a) Equirectangular image (b) Disparity image



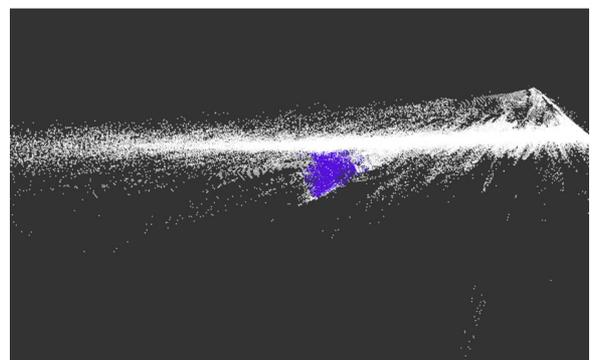
(c) Point cloud from front view



(c) Point cloud from front view



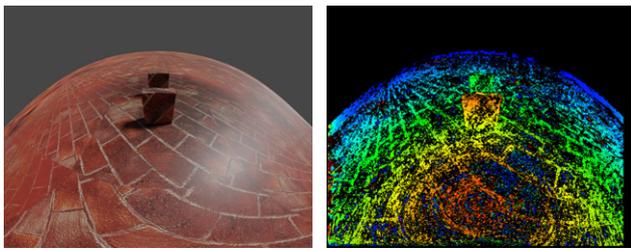
(d) Point cloud from side view



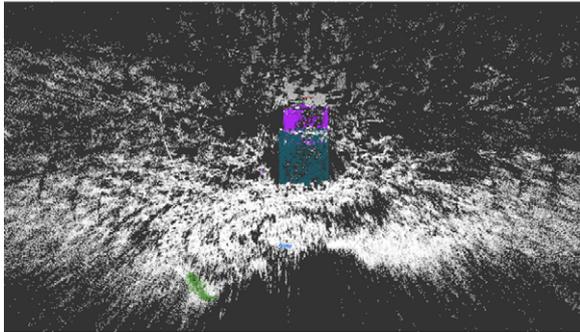
(d) Point cloud from side view

Fig. 22. Results of inclined environment.

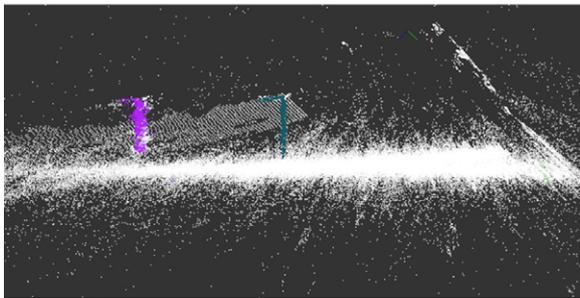
Fig. 23. Results for environment with a hole.



(a) Equirectangular image (b) Disparity image



(c) Point cloud from front view



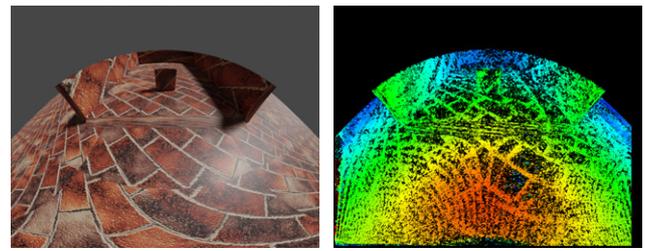
(d) Point cloud from side view

Fig. 24. Results for environment with occlusion.

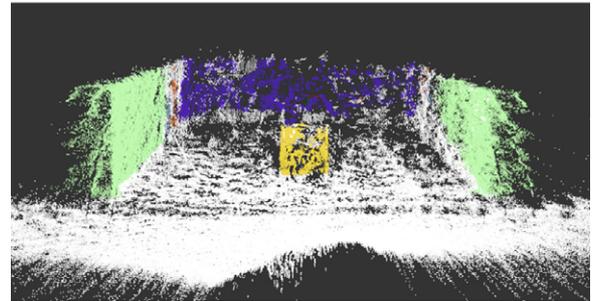
distances and heights was reduced by errors in the estimation of the planar parameters owing to errors in the 3D measurements.

7. Conclusion

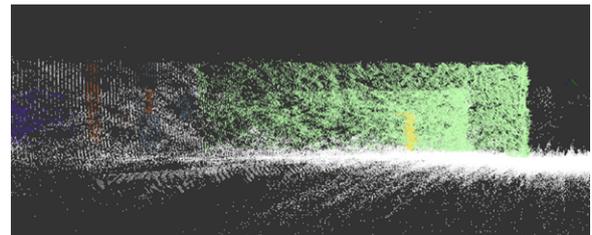
In this paper, we proposed a method for road surface plane estimation and object extraction using a hierarchical structure based on 3D information obtained from a fisheye stereo camera. Experimental results show that objects can be extracted in a flat environment. In addition, experiments with changes in road slope show that the system can estimate the road surface without mistaking the change of slope as an object. Simulation results show that the method is also applicable to environments with holes and changes in slopes. However, the system is currently limited in its ability to recognize the environment because it only processes a disparity image. Including color images can provide vital cues to cover for situations where the disparity is inaccurate. Future works will aim at in-



(a) Equirectangular image (b) Disparity image



(c) Point cloud from front view



(d) Point cloud from side view

Fig. 25. Results for environment with parking lot.

Table 1. Errors of distance and height.

	Distance [m]	Height [m]
Inclined plane	0.19	0.01
Occlusion (front)	0.19	0.09
Occlusion (back)	0.04	-0.04
Parking lot	0.02	0.01

creasing the accuracy and speed by considering the consistency of the time series. In addition, it is important to improve the accuracy of 3D measurement using the fish-eye stereo camera, to verify images that include motion blur, and to verify whether the occlusion can be identified as another obstacle in an actual environment.

References:

- [1] Y. Wei, C. Gong, and S. Chen, "Harnessing U-disparity in point clouds for obstacle detection," 2017 4th IAPR Asian Conf. on Pattern Recognition (ACPR), pp. 262-267, 2017. <https://doi.org/10.1109/ACPR.2017.106>
- [2] X. Chen, H. Ma, J. Wan, B. Li, and T. Xia, "Multi-view 3D object detection network for autonomous driving," 2017 IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), pp. 6526-6534, 2017. <https://doi.org/10.1109/CVPR.2017.691>
- [3] M. Engelcke, D. Rao, D. Z. Wang, C. H. Tong, and I. Posner, "Vote3Deep: Fast object detection in 3D point clouds using efficient convolutional neural networks," 2017 IEEE Int. Conf. on

- Robotics and Automation (ICRA), pp. 1355-1361, 2017. <https://doi.org/10.1109/ICRA.2017.7989161>
- [4] C. R. Qi, W. Liu, C. Wu, H. Su, and L. J. Guibas, "Frustum PointNets for 3D object detection from RGB-D data," 2018 IEEE/CVF Conf. on Computer Vision and Pattern Recognition, pp. 918-927, 2018. <https://doi.org/10.1109/CVPR.2018.00102>
- [5] M. Liu, C. Shan, H. Zhang, and Q. Xia, "Stereo vision based road free space detection," 2016 9th Int. Symp. on Computational Intelligence and Design (ISCID), pp. 272-276, 2016. <https://doi.org/10.1109/ISCID.2016.2072>
- [6] 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 Trans. on Intelligent Transportation Systems, Vol.19, No.3, pp. 758-773, 2018. <https://doi.org/10.1109/TITS.2017.2700628>
- [7] A. Seki and M. Okutomi, "Robust obstacle detection in general road environment based on road extraction and pose estimation," 2006 IEEE Intelligent Vehicles Symp., pp. 437-444, 2006. <https://doi.org/10.1109/IVS.2006.1689668>
- [8] P. Li, X. Chen, and S. Shen, "Stereo R-CNN based 3D object detection for autonomous driving," 2019 IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR), pp. 7636-7644, 2019. <https://doi.org/10.1109/CVPR.2019.00783>
- [9] B. Li, W. Ouyang, L. Sheng, X. Zeng, and X. Wang, "GS3D: An efficient 3D object detection framework for autonomous driving," 2019 IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR), pp. 1019-1028, 2019. <https://doi.org/10.1109/CVPR.2019.00111>
- [10] X. Chen, K. Kundu, Y. Zhu, H. Ma, S. Fidler, and R. Urtasun, "3D object proposals using stereo imagery for accurate object class detection," IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol.40, No.5, pp. 1259-1272, 2018. <https://doi.org/10.1109/TPAMI.2017.2706685>
- [11] C. Zhou, Y. Liu, Q. Sun, and P. Lasang, "Vehicle detection and disparity estimation using blended stereo images," IEEE Trans. on Intelligent Vehicles, Vol.6, No.4, pp. 690-698, 2021. <https://doi.org/10.1109/TIV.2020.3049008>
- [12] H. Wu, H. Su, Y. Liu, and H. Gao, "Object detection and localization using stereo cameras," 2020 5th Int. Conf. on Advanced Robotics and Mechatronics (ICARM), pp. 628-633, 2020. <https://doi.org/10.1109/ICARM49381.2020.9195365>
- [13] T. Akita, Y. Yamauchi, and H. Fujiyoshi, "Stereo vision by combination of machine-learning techniques for pedestrian detection at intersections utilizing surround-view cameras," J. Robot. Mechatron., Vol.32, No.3, pp. 494-502, 2020. <https://doi.org/10.20965/jrm.2020.p0494>
- [14] S. Tanaka and Y. Inoue, "Outdoor human detection with stereo omnidirectional cameras," J. Robot. Mechatron., Vol.32, No.6, pp. 1193-1199, 2020. <https://doi.org/10.20965/jrm.2020.p1193>
- [15] T. Senoo, A. Konno, Y. Wang, M. Hirano, N. Kishi, and M. Ishikawa, "Tracking of overlapped vehicles with spatio-temporal shared filter for high-speed stereo vision," J. Robot. Mechatron., Vol.34, No.5, pp. 1033-1042, 2022. <https://doi.org/10.20965/jrm.2022.p1033>
- [16] T. Hayakawa, Y. Moko, K. Morishita, Y. Hiruma, and M. Ishikawa, "Tunnel lining surface monitoring system deployable at maximum vehicle speed of 100 km/h using view angle compensation based on self-localization using white line recognition," J. Robot. Mechatron., Vol.34, No.5, pp. 997-1010, 2022. <https://doi.org/10.20965/jrm.2022.p0997>
- [17] A. Ohashi, F. Yamano, G. Masuyama, K. Umeda, D. Fukuda, K. Irie, S. Kaneko, J. Murayama, and Y. Uchida, "Stereo rectification for equirectangular images," 2017 IEEE/SICE Int. Symp. on System Integration (SII), pp. 535-540, 2017. <https://doi.org/10.1109/SII.2017.8279276>
- [18] K. Terabayashi, H. Mitsumoto, T. Morita, Y. Aragaki, N. Shimomura, and K. Umeda, "Measurement of three-dimensional environment with a fish-eye camera based on structure from motion-error analysis," J. Robot. Mechatron., Vol.21, No.6, pp. 680-688, 2009. <https://doi.org/10.20965/jrm.2009.p0680>
- [19] K. Terabayashi, T. Morita, H. Okamoto, and K. Umeda, "3D measurement using a fish-eye camera based on EPI analysis," J. Robot. Mechatron., Vol.24, No.4, pp. 677-685, 2012. <https://doi.org/10.20965/jrm.2012.p0677>
- [20] 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," 2020 IEEE/SICE Int. Symp. on System Integration (SII), pp. 343-348, 2020. <https://doi.org/10.1109/SII46433.2020.9025910>
- [21] C. Tomasi and R. Manduchi, "Bilateral filtering for gray and color images," 6th Int. Conf. on Computer Vision, pp. 839-846, 1998. <https://doi.org/10.1109/ICCV.1998.710815>
- [22] S.-H. Wang and X.-X. Li, "A real-time monocular vision-based obstacle detection," 2020 6th Int. Conf. on Control, Automation and Robotics (ICCAR), pp. 695-699, 2020. <https://doi.org/10.1109/ICCAR49639.2020.9108018>
- [23] A.-L. Hou, X. Cui, Y. Geng, W.-J. Yuan, and J. Hou, "Measurement of safe driving distance based on stereo vision," 2011 6th Int. Conf. on Image and Graphics, pp. 902-907, 2011. <https://doi.org/10.1109/ICIG.2011.27>
- [24] Y. Liu, H. Huang, J. Guo, R. Yuan, M. Xia, and Y. Chen, "Research on target distance detection technology of vehicle assisted driving based on monocular vision," 2021 11th Int. Conf. on Information Technology in Medicine and Education (ITME), pp. 17-20, 2021. <https://doi.org/10.1109/ITME53901.2021.00014>
- [25] N. Baha and M. Tolba, "Towards real-time obstacle detection using stereo images," 2015 Science and Information Conf. (SAI), pp. 672-679, 2015. <https://doi.org/10.1109/SAI.2015.7237214>

Supporting Online Materials:

- [a] Blender Foundation Official Website. <https://www.blender.org/> [Accessed January 11, 2023]



Name:

Tomoyu Sakuda

Affiliation:

Daihatsu Motor Co., Ltd.

Address:

1-1 Daihatsucho, Ikeda, Osaka 563-8651, Japan

Brief Biographical History:

2022 Received M.Eng. degree in Precision Engineering from Chuo University
2022- Daihatsu Motor Co., Ltd.



Name:

Hikaru Chikugo

Affiliation:

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

Address:

1-13-27 Kasuga, Bunkyo-ku, Tokyo 112-8551, Japan

Brief Biographical History:

2023 Received B.Eng. degree in Precision Mechanics from Chuo University

Membership in Academic Societies:

- The Japan Society of Mechanical Engineers (JSME)



Name:
Kento Arai

Affiliation:
Information Services International-Dentsu, Ltd.

Address:
2-17-1 Konan, Minato-ku, Tokyo 108-0075, Japan

Brief Biographical History:
2023 Received M.Eng. degree in Precision Engineering from Chuo University
2023- Information Services International-Dentsu, Ltd.



Name:
Sarthak Pathak

ORCID:
0000-0002-5271-1782

Affiliation:
Assistant Professor, Department of Precision Mechanics, Faculty of Science and Engineering, Chuo University

Address:
1-13-27 Kasuga, Bunkyo-ku, Tokyo 112-8551, Japan

Brief Biographical History:
2014 Received Bachelor of Technology and Master of Technology degrees from Department of Engineering Design, Indian Institute of Technology Madras
2017 Received Ph.D. degree, Department of Precision Engineering, The University of Tokyo
2017- Postdoctoral Researcher, Department of Precision Engineering, The University of Tokyo
2018- JSPS Postdoctoral Research Fellow, Department of Precision Engineering, The University of Tokyo
2020- Project Assistant Professor, Department of Precision Engineering, The University of Tokyo
2021- Assistant Professor, Department of Precision Mechanics, Faculty of Science and Engineering, Chuo University

Main Works:

- S. Pathak, A. Moro, H. Fujii, A. Yamashita, and H. Asama, "Spherical video stabilization by estimating rotation from dense optical flow fields," J. Robot. Mechatron., Vol.29, No.3, pp. 566-579, 2017.
- D. Kim, S. Pathak, A. Moro, A. Yamashita, and H. Asama, "Self-supervised optical flow derotation network for rotation estimation of a spherical camera," Advanced Robotics, Vol.35, No.2, pp. 118-128, 2021.
- S. Pathak, A. Moro, A. Yamashita, and H. Asama, "A decoupled virtual camera using spherical optical flow," 2016 IEEE Int. Conf. on Image Processing (ICIP), pp. 4488-4492, 2016.

Membership in Academic Societies:

- Institute of Electrical and Electronic Engineers (IEEE)
- The Japan Society for Precision Engineering (JSPE)
- The Robotics Society of Japan (RSJ)



Name:
Kazunori Umeda

ORCID:
0000-0002-4458-4648

Affiliation:
Professor, Department of Precision Mechanics, Faculty of Science and Engineering, Chuo University

Address:
1-13-27 Kasuga, Bunkyo-ku, Tokyo 112-8551, Japan

Brief Biographical History:
1994 Received Ph.D. degree in Precision Machinery Engineering from The University of Tokyo
1994- Lecturer, Department of Precision Mechanics, Faculty of Science and Engineering, Chuo University
2003-2004 Visiting Worker, National Research Council Canada
2006- Professor, Department of Precision Mechanics, Faculty of Science and Engineering, Chuo University

Main Works:

- M. Shinozaki, M. Kusanagi, K. Umeda, G. Godin, and M. Rioux, "Correction of color information of a 3D model using a range intensity image," Computer Vision and Image Understanding, Vol.113, No.11, pp. 1170-1179, 2009.
- T. Kuroki, K. Terabayashi, and K. Umeda, "Construction of a compact range image sensor using multi-slit laser projector and obstacle detection of a humanoid with the sensor," 2010 IEEE/RSJ Int. Conf. on Intelligent Robots and Systems, pp. 5972-5977, 2010.
- S. Yan, S. Pathak, and K. Umeda, "PointpartNet: 3D point-cloud registration via deep part-based feature extraction," Advanced Robotics, Vol.36, No.15, pp. 724-734, 2022.

Membership in Academic Societies:

- The Robotics Society of Japan (RSJ)
- The Japan Society for Precision Engineering (JSPE)
- The Japan Society of Mechanical Engineers (JSME)
- The Society of Instrument and Control Engineers (SICE)
- The Institute of Electronics, Information and Communication Engineers (IEICE)
- Institute of Electrical and Electronics Engineers (IEEE)