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

Tomoyu Sakuda[1], Kento Arai[1], Sarthak Pathak[1] and Kazunori Umeda[1]

Abstract— **In this paper, we propose a novel algorithm to estimate road surface shape and object height using a fisheye stereo camera. Environment recognition is an important task for Advanced Driver-Assistance Systems (ADAS). However, previous studies use sensors with a narrow measurement range. They also use the constraint that the road is flat and cannot deal with changes in the road slope. We use a fisheye stereo camera, which can measure wide and dense 3D information, and design a novel algorithm focusing on the scale of division on the disparity image to overcome these defects. Experiments show that our method can detect an object. Furthermore it is shown that measurement accuracy of obstacle height and distance to obstacles strongly depends on the 3D measurement accuracy.**

I. INTRODUCTION

In recent years, there have been many research and development on driver assistance systems. Driver assistance systems support drivers based on information received from range sensors such as LiDAR and stereo cameras, but it is necessary to understand the surrounding environment of the car in order to avoid obstacles. Among the used range sensors, stereo cameras in particular are used to understand the surrounding environment of a car in detail because of their high measurement density and the ability to acquire color images, and various methods for understanding the environment have been proposed [1]−[6].

A method using UV-disparity has been proposed to estimate the road surface area and extract obstacles in order to understand the environment [1][2]. 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. Seki et al. proposed a method that uses the property that the road is a planar area projected on two different image planes [3]. However, this method 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.

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 stereo camera, and the narrow field of view of the range image sensor results in a narrow measurement range. Therefore, in this study, we use a fisheye stereo camera [7] with a wide viewing angle as a range image sensor, and estimate the road surface plane by using the fact that the estimated plane varies depending on the size of the segmentation of the disparity image. In this way, we propose a method for understanding the environment without relying on deep learning, which can be applied to a wide range of environments with varying slopes.

[1]The Course of Precision Engineering, School of Science and Engineering, Chuo University, 1-13-27 Kasuga, Bunkyo-ku, Tokyo, Japan. {sakuda, arai}@sensor.mech.chuo-u.ac.jp

II. OUTLINE OF THE PROPOSED METHOD

An overview of the proposed method is shown in Fig. 1. The method consists of 3D measurement, road surface plane estimation, and obstacle extraction. For the 3D measurement, we use a pseudo-bilateral filter [8], and for the road surface plane estimation, we propose a method that adapts to changes in slope by fitting multiple planes. For obstacle extraction, after morphological processing, clustering is performed using the frequency value in distance.

III. 3D MEASUREMENT USING PSEUDO-BILATERAL FILTERS

In the 3D measurement method using pseudo-bilateral filters, a dense distance image obtained by a region-based binocular stereo camera is fused with a high-precision distance image obtained by Structure from Motion (SfM) to achieve a dense and high-precision 3D measurement [8]. 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 to avoid adverse effects of the distortion inherent in the fisheye on

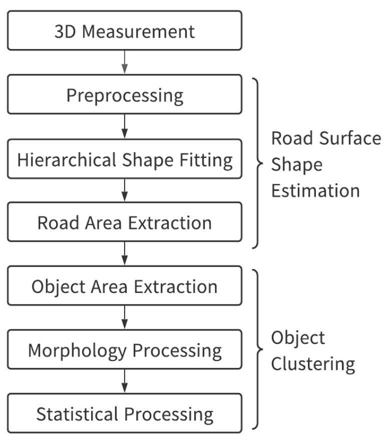


Fig. 1. The flow of the proposed method

2023 IEEE/SICE International Symposium on System Integration (SII) | 979-8-3503-9868-7/23/$31.00 ©2023 IEEE | DOI: 10.1109/SII55687.2023.10039474

the 3D measurement. By processing in the disparity space of the equirectangular image, we do not need to take into account the error amplification that occurs when the image is converted to real space.

IV. ESTIMATING ROAD SURFACE PLANE USING HIERARCHICAL STRUCTURE

A. *Preprocessing*

The 3D information obtained from the fisheye stereo camera contains outliers, which have a negative impact on the road surface estimation and should be removed. Since outliers have a large difference in value from the surrounding data, we can remove them by looking for areas with large changes. Fig. 3 shows the measured points and outliers in elevation-disparity space. We search in the elevation direction for each azimuth angle, and the area where the disparity change is greater than the threshold value $\Delta\lambda_{th}$ and less than the elevation width threshold value ϕ_{th} can be considered as an outlier, so we remove the points in that area.

B. *Plane Estimation*

When estimating the plane, the presence of obstacles 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 area for plane estimation. If we divide the disparity image into smaller areas, the possibility that both obstacles and road surface exist simultaneously is reduced, and it becomes easier to approximate the road surface shape to a plane. However, if the area to be estimated is made too small, it is affected by random errors in the measurement of the point cloud. Therefore, in order to estimate the plane with an appropriate division size, we gradually change the size of the division and perform plane estimation in each area. We assume that the plane that matches the estimated plane parameters before and after the change is the plane that fits the point cloud well. The procedure of the method is shown below.

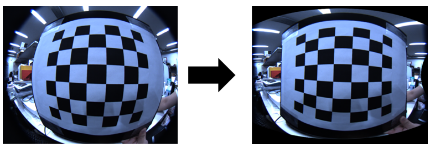


Fig. 2. Equirectangular image

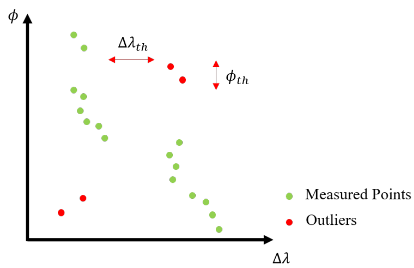


Fig. 3. Disparity-elevation angle space

- Step 1) Divide the disparity image into blocks of square area. The length of one side of the square is $2^n (n \in N)$.
- Step 2) Apply a median filter of size $2^{(n-1)} \times 2^{(n-1)}$ to each block in nine locations, so that the intersection of the six lines drawn at equal intervals is the center, as shown in Fig. 4, according to Eq. (1).
- Step 3) Calculate the plane parameters (a, b, c) by the least-squares method using the values obtained in Step 2.
- Step 4) Divide the length of each block into halves and perform Steps 2 to 3, and then Step 5.
- Step 5) Calculate the similarity of the plane parameters before and after the division, and perform Step 6 if the similarity is high, and Step 7 if the similarity is low.
- Step 6) The obtained planar parameters are assumed to have been estimated at the appropriate partition and the planar parameters are determined.
- Step 7) Repeat Step 4. If the number of attempts exceeds a certain number, the planar parameters cannot be estimated at that location (e.g., the boundary between an obstacle and a road surface).

In Step 3),

$$\Delta\lambda = a\cos\phi\sin\lambda + b\tan\phi\cos\lambda + c\cos^2\lambda \qquad (1)$$

is used to calculate the plane parameters.This equation is obtained by substituting the 3D position calculated from λ and ϕ into the plane equation. Here, (a, b, c) are the plane parameters, λ is the azimuth angle, ϕ is the elevation angle, and $\Delta\lambda$ is the disparity. In this way, the obtained set of planes contains planes that are fitted to obstacles. Therefore, the plane corresponding to the obstacle is removed by focusing on the change in slope of the obtained plane group. As shown in Fig. 5, we prepare several seed points, and expand the domain to the region where the change in slope is small at each seed point. The seed point with the largest area is selected from the final obtained area, and the area obtained by extending the seed point is defined as the road surface area.

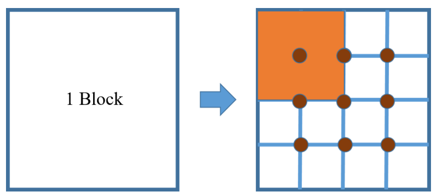


Fig. 4. How to split a block in disparity image

V. OBSTACLE EXTRACTION AND HEIGHT ESTIMATION

A. Morphology Processing

The road surface plane obtained in section IV is used to extract obstacle candidates from the point cloud. In order to define the road surface in the region where the plane corresponding to the obstacle is estimated, we take the average of the plane parameters in the nearby road surface region and interpolate the plane parameters in the obstacle region. A threshold is set for the absolute value of the difference in disparity with the obtained plane, and the points above the threshold are considered obstacles. Next, a binary image is generated, with 1 representing the location where the obstacle exists and 0 otherwise. The image is then subjected to an opening and closing process to remove small regions and holes. Then, clustering is performed via connected components analysis.

B. Statistical Processing

In section V(A), we performed clustering. However, since the process is based on binary images, obstacles with occlusions are combined and extracted as the same obstacle. Therefore, for each obstacle extracted by clustering, we focus on the frequency value of distance. Since the distance value measured from the same obstacle is continuous, if there is an intermittent point, it is extracted as a different obstacle. In addition, points with low frequency values are removed as mismeasured points.

VI. EXPERIMENTS FOR ACCURACY EVALUATION

A. Experimental Conditions

We conducted an experiment to verify that the obstacles could be extracted correctly. We also evaluated the accuracy of the distance to the obstacle and the height of the obstacle. We used a 0.5 m high cardboard as the object. A schematic of the environment is shown in Fig. 6. The obstacles were placed so that the center of the obstacle was at an azimuth angle of 0°. 10 shots were taken at distance of 1 m, 2 m, and 3 m between the obstacle and the camera. Fig. 7 shows the appearance of the fisheye stereo camera used in this experiment. The fisheye stereo camera was placed at a height of 1 m with a pitch angle of 30°. The camera movement also was set to 0.15 m in the horizontal direction. 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 plane estimation is the inner product of the normalized plane parameters, with a threshold of 0.98. For the restricted expansion, the limits were that the inner product of the normalized plane parameters should be greater than the threshold value of 0.95 and that the average distance of the lines in the adjacent lines in the image should be less than 0.35 m. For the clustering of obstacles, areas with points with a frequency value of 40 or more were considered to be obstacle areas, and within an obstacle area, if the distance was discontinuous by more than 0.3 m, it was divided into separate obstacles.

B. Experimental Results

The experimental results are shown in Fig. 8 - 10. Fig. 8 shows the images captured during the experiment. Fig. 9 is a 3D point cloud display of the obtained 3D information, and each obstacle is distinguished by color. The obstacles in the center are detected correctly, but non-existent obstacles are detected on the left and right. This is caused by a discontinuous systematic errors in the 3D measurement. The maximum error in the obstacle height and distance to the obstacle is 0.3 m, but as can be seen in Fig. 9, the shape of the detected objects looks correct, making it unlikely that this is due to false obstacle detection. Therefore, the errors in the obstacle height and distance to the obstacle are considered to be largely caused by 3D measurement errors.

C. Evaluation Experiment for 3D Point Cloud

In order to remove the influence of the error of 3D measurement, we manually calculated the height of the obstacle

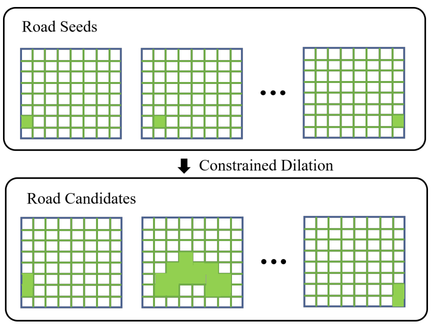


Fig. 5. Road area extraction in disparity image

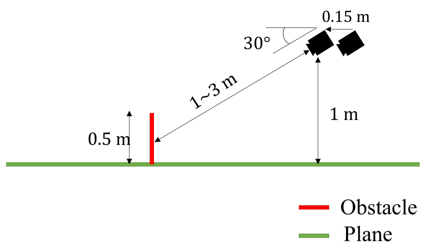


Fig. 6. Experimental condition

Fig. 7. Appearance of the Fisheye Stereo Camera

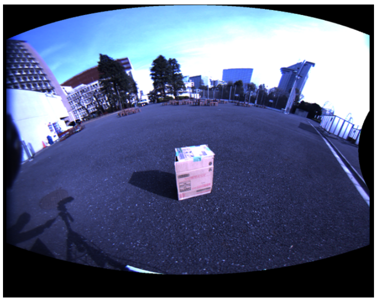
and the distance to the obstacle in the 3D point cloud, and evaluated the accuracy of the method. We manually fitted a plane to the target obstacle as shown in Figure 11, and calculated the distance and height to the obstacle.

In the case of obstacle height,

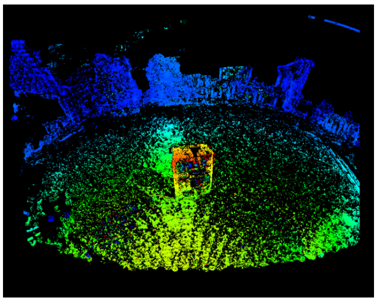
$$y \cos \frac{\pi}{6} + z \sin \frac{\pi}{6} + d = 0 \tag{2}$$

is fitted to the top and bottom of the obstacle, and calculate

$$h = d_{top} - d_{bottom} \tag{3}$$

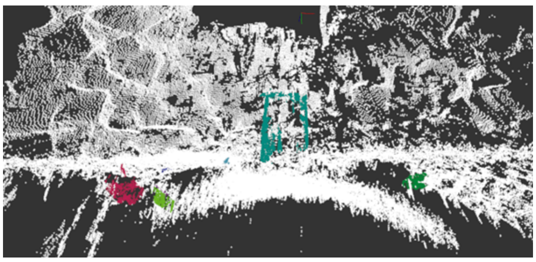


(a) Equirectangular Image

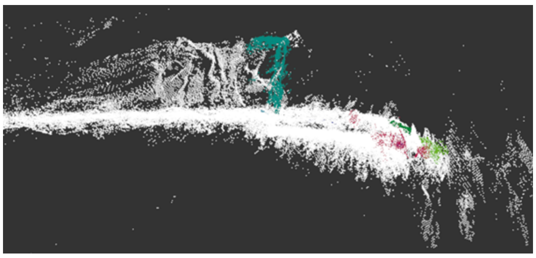


(b) Disparity Image

Fig. 8. Images with the target obstacle at 1 m



(a) The front view



(b) The side view

Fig. 9. Point clouds with the target at 1 m. The colored points are the detected objects.

to obtain the true value. In the case of the distance to the obstacle,

$$y \cos \frac{2\pi}{3} + z \sin \frac{2\pi}{3} + d = 0 \tag{4}$$

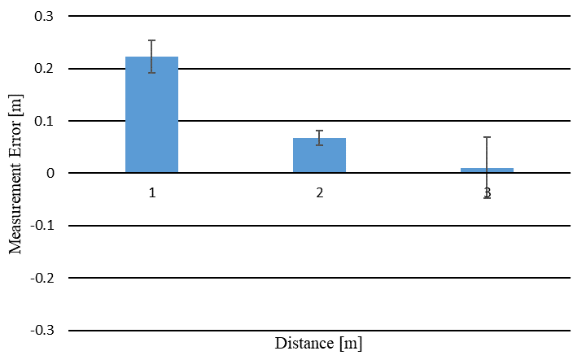
is fitted to the bottom of the obstacle, and

$$D = \sqrt{d^2 + 0.75^2} \tag{5}$$

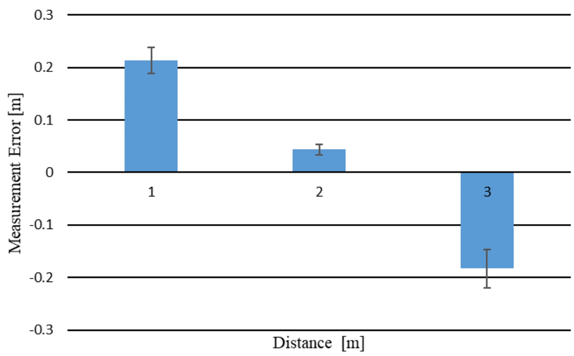
is calculated to find the distance to the obstacle D, where (x, y, z) are the coordinates of the camera coordinate system of the right camera after the movement, and d is a parameter to be fitted manually. The planes that are applied to the top of the obstacle and the boundary between the obstacle and the road surface are denoted by the subscripts top and $bottom$. 0.75 m is the camera height of 1 m minus half of the obstacle height of 0.5 m. The results of the experiment are shown in Fig. 12, which shows that the error is much smaller than the experimental results in Section VI(B). It can be seen that the accuracy of the method is greatly affected by the error of the 3D measurement, but it is able to fit the 3D point cloud with good accuracy.

VII. CONCLUSION

In this study, we proposed a method for estimating the height of road surface planes and obstacles using 3D point clouds obtained with a fisheye stereo camera, focusing on the segmentation scale of the disparity image. Experiments showed that the method was able to detect obstacles in an



(a) Height of the object



(b) Distance to the object

Fig. 10. Measurement error

environment with little change in slope. We also showed that the error in the measurement of the height of the obstacle and the distance to the obstacle is mainly caused by the 3D measurement error. As a future prospect, it is necessary to verify whether it is possible to estimate the road surface plane and extract obstacles in the environment where occlusion occurs.

REFERENCES

[1] 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.

[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] 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.

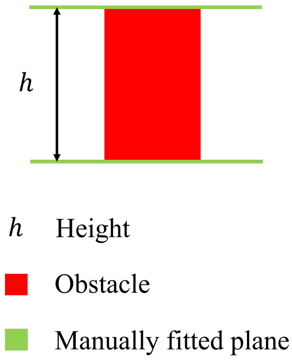
[4] P. Li, X. Chen and S. Shen,"Stereo R-CNN Based 3D Object Detection for Autonomous Driving", 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 7636-7644, 2019.

[5] 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), 2019.

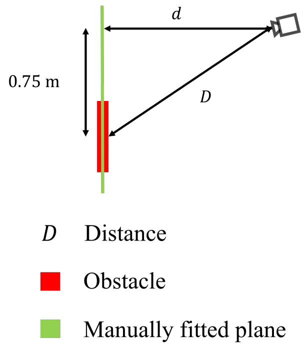
[6] X. Chen, K. Kundu, Y. Zhu, H. Ma, S. Fidler and R. Urtasun, "3D object proposals using stereo imagery for accurate object class detection", Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017.

[7] 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), 2017.12.

[8] 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, 2020.1.

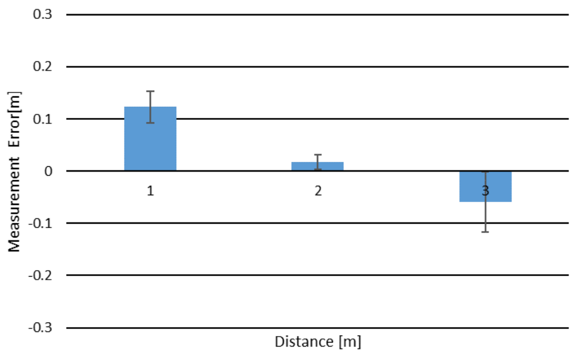


(a) Height

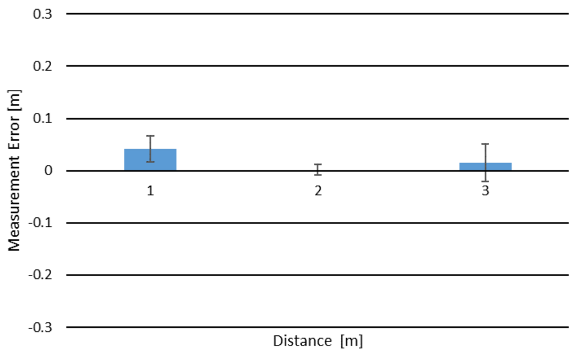


(b) Distance

Fig. 11. Manual calculation of height and distance of target obstacle



(a) Height of the object



(b) Distance to the object

Fig. 12. Measurement error