Three-dimensional Environmental Measurement of Surroundings Using Camera Pose Estimation Base on Line Features

Shingo Iriyama[1] , Sarthak Pathak[2] and Kazunori Umeda[2]

***Abstract*— This study introduces a cost-effective measurement approach for indoor environments, intended for inspection purposes for autonomous mobile robotics and infrastructure maintenance. The method involves utilizing a spherical camera capable of capturing 360 degrees view of the surroundings, along with a ring laser, to obtain a 3D point cloud representing a cross-section of a room through the structured light method. The camera and laser rotated 360 degrees in the target environment. The camera's orientation is established by comparing the distribution of line features within the room to the three directions in real space. By considering the extent of change in the camera's posture, the method integrates multiple point clouds generated by the structured light method. This results in the creation of a comprehensive 3D point cloud that represents the entire indoor environment.**

I. INTRODUCTION

High-precision 3D measurement technology is extensively utilized for autonomous robot mobility technology and maintenance of large facilities [1]. Self-position estimation for mobile robots in indoor environments involves correlating information acquired from cameras and laser scanners with a preexisting environmental map [2][3]. To achieve precise position estimation, it is important to have a highly accurate three-dimensional map. In maintenance, verifying the correct production of buildings and large products is essential. However, these tasks are typically performed by humans and are both time-consuming and costly. To achieve efficient quality control, acquiring detailed 3D data on the work surface is imperative. Currently, various methods and devices, including omni-directional 3D measurement techniques, are employed. One of these sensor is a LiDAR sensor[4], which measures distance and shape using Time-of-Flight (ToF), by irradiating a light pulse onto an object and measuring the time of a light pulse trip. However, in medium- to short-range environmental sensing, such as indoor environments, the time of a light pulse trip is shorter, and results in lower measurement accuracy. Another disadvantage is the relatively high price of the device itself. Consequently, the demand for accurate and budget-friendly 3D measurement devices is on the rise.

There is a study on 3D measurement in indoor environments using a combination of spherical cameras and structured light method using a ring laser[5]. This method involves fusing the 3D point cloud obtained through the light cutting method with the camera pose information obtained through Structure-from-Motion (SFM) [6] techniques. However, this approach is sensitive to the textures present in the indoor environment, and a lack of textures can adversely affect the accuracy of SFM-based pose estimation and 3D measurement. Additionally, there are studies that utilize a spherical camera for pose estimation, where they match the straight line information in the images with the pre-prepared 3D model of the space [7]. They describe the straight line information on the floor, walls, and other surfaces in the images, as well as the 3D model, as descriptors. By comparing these two sets of descriptors, they calculate the camera's position and orientation when the images were captured. However, it is necessary to prepare an accurate 3D model in advance.

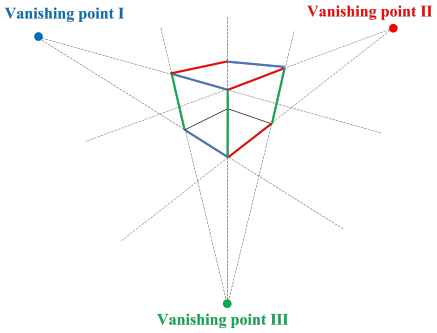


Fig. 1. Three vanishing points [8]

In this study, we propose a 3D measurement method using a spherical camera and a ring laser that is suitable for textureless environments. This method improves upon previous camera pose estimation processes. Specifically, we obtain a point cloud of a cross-sectional indoor area using a camera and a laser with the the structured light method. Simultaneously, we estimate the camera's orientation, and then we integrate multiple laser point clouds based on the amount of rotation. To estimate the camera's pose, we assume that even in an indoor environment with few patterns, the edges of walls and ceilings can be acquired. Therefore, we focus on the line features in the environment for camera pose estimation. Our method is based on the Manhattan world hypothesis [8], assuming that all straight lines lie in one of three principal directions, which are perpendicular to each other. Consequently, all lines are oriented in three directions, corresponding to the three vanishing point directions in real space as shown in Fig.1. By calculating the vanishing point's direction in real space from the line distribution in the

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

[2]Department of Precision Mechanics, Faculty of Science and Engineering, Chuo University, 1-13-27 Kasuga, Bunkyo-ku, Tokyo, Japan. pathak, umeda@mech.chuo-u.ac.jp

image and comparing it with the camera coordinate system's direction, we can estimate the camera's attitude relative to the real space. With this approach, we can accurately obtain 3D information of an entire indoor space using only the prior knowledge that straight lines in the indoor space are perpendicular to each other.

Furthermore, to achieve precise orientation estimation, it is crucial to acquire three-way linear information of the indoor area, regardless of the camera's orientation. To address this, we utilized a spherical camera in this study, as it can capture 360 degrees view around the object.

II. PROPOSED METHOD

A. Overview of the Measurement Setup

The overall measurement setup is shown in Fig.2. The setup involves a single fisheye camera and one circular laser attached to a tripod. The measurement process involves rotating the entire setup vertically around the camera's center as the axis while conducting the measurements. The measurement device itself remains fixed at an arbitrary position from the start to the end of the process.

B. Overview of the Proposed Method

The proposed method consists of a process of 3D measurement and a process of pose estimation, as shown in Fig.3. At a single camera position, two images are captured: one showcasing the laser-illuminated scene and the other representing the room itself. Concerning the 3D measurement, the laser-projected image is utilized to extract pixels corresponding to the laser. Employing the light sectioning technique, a 3D point cloud capturing the cross-sectional structure of the room is generated.

Next, we focus on estimating the camera's pose Utilizing deep learning, line features are extracted from the room image. By executing plane fitting on these line features, alterations in the camera's orientation are computed. These two processes are iteratively executed while incrementally rotating the camera within the indoor environment. Using the poses of each camera, the 3D point clouds obtained from all frames are integrated, resulting in the creation of a unified 3D point cloud representation of the entire indoor space.

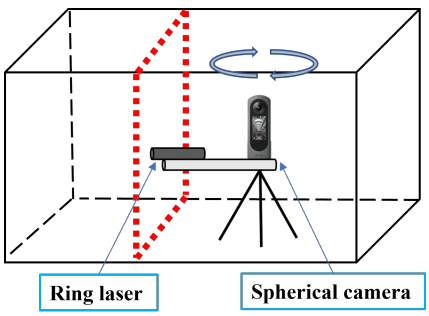


Fig. 2. Measuring setup

C. 3D Measurement

A omnidirectional image showing the laser projection area is captured with a spherical camera. A binary processing is applied to the image, where the red laser area is assigned a pixel value of 255, while the rest is set to a pixel value of 0. Subsequently, the weighted centroid position of the circular laser projection is computed. Taking into account the characteristic circular shape of the laser light in the binary image, a weighted average is calculated in the u and v directions of the image. The combined coordinates of the centroid positions in the u and v directions are regarded as the centroid position of the circular laser projection as shown in Fig.4. Then, utilizing the weighted centroid image and the geometric relationship between the fisheye camera and the circular laser, a 3D point cloud is reconstructed.

D. Concept of Pose Estimation

Fig.5 show the indoor environment and the camera coordinate system. Based on the Manhattan World hypothesis,[3] all the lines indoors face three orthogonal directions. Therefore, these three directions are defined as the world coordinate system. By comparing the direction of the straight lines in the image taken by the rotating camera and the direction of the camera coordinate system, the change in the camera's posture relative to the world coordinate system can be obtained. Therefore, we perform posture estimation using the straight lines in the indoor environment as the feature values.

E. Line Features Extraction

Unified Line Segment Detection (ULSD) [9] based on deep learning detects straight lines in a positively curved a omnidirectional image captured by a spherical camera. ULSD utilizes an end-to-end network to perform line segment detection using the Bezier curve model. This detection approach can accommodate both undistorted perspective projection images and distorted images captured by fisheye

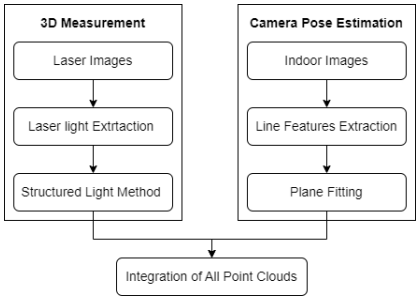


Fig. 3. Flow of the proposed method

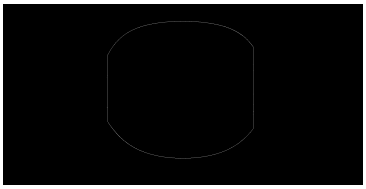


Fig. 4. Extraction of laser light

or spherical cameras. The input image is represented as the indoor scene image shown in Fig. 6(a). The resulting image after line detection is showing in Fig.6(b), where the orange regions represent the detected line areas, and the cyan dots denote the endpoints of the detected lines.

Straight lines within the omnidirectional image can be projected onto the circumference of a 3D spherical coordinate system, as shown in Fig.7. Furthermore, the normal vectors of the straight lines, projected onto the spherical coordinate system and pointing in the same direction, possess the property of lying on the same plane. In other words, by analyzing the normal vectors, the orientation of the world coordinate system in space can be determined. Therefore, unit normal vectors are defined as line features. The relationship between the 3D spherical coordinate system's lines and the normal vectors as line features is illustrated in Fig.6. The specific procedure for pose estimation is outlined below.

- The line in the image is converted to 3D coordinates on the unit sphere.
- The unit normal vector, $\mathbf{n_{12}}$ of a line is obtained by randomly selecting two points, $\mathbf{p_1}$ and $\mathbf{p_2}$, from the point set of the line and computing the outer product.

F. Plane Fitting for Estimation Camera Pose

In this study, we assume an environment where all lines in space are oriented in three orthogonal directions. As a result, the normal vectors of the lines can be classified into three planes. As shown in Fig.8, as the camera's pose changes, both the distribution of lines in the omnidirectional image and the distribution of unit normal vectors change accordingly. Thus, by fitting the three planes of the camera's coordinate system and the three planes of the unit normal vectors, the camera's pose with respect to the world coordinate system can be determined. Specifically, we employ the Levenberg-Marquardt method [10] to obtain a rotation matrix that minimizes the distance errors between these planes.

Let $n_k(k = 1, 2, ..., n)$ represent the unit normal vectors, and $W_l(l = 1, 2, ..., n)$ denote the three axes of the world coordinate system. The camera attitude change is expressed by a rotation matrix R sing Euler angles (α, β, γ) The distance d_{kl} between n_k is defined as the absolute value of the inner product in (1).

$$d_{kl} = |n_k \cdot R \cdot W_l| \tag{1}$$

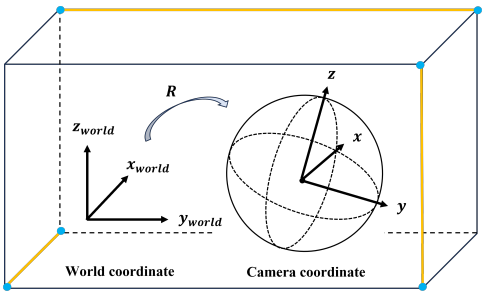
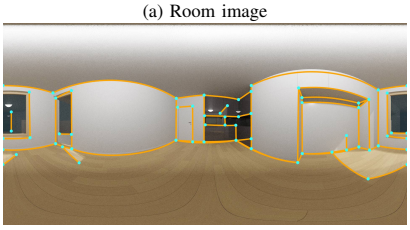


Fig. 5. Coordinate system

(a) Room image



(b) Line image

Fig. 6. Line detection using ULSD

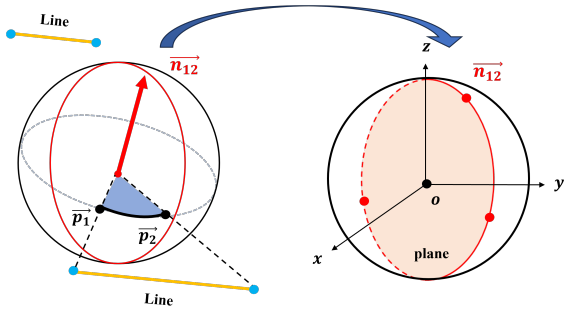


Fig. 7. Definition of line features

For each normal vector, the smallest of the three inner product values is represented as d_k in (2).

$$d_k = min(d_{k1}, d_{k2}, d_{k3}) \tag{2}$$

The parameters (α, β, γ) that minimize the inner product values of all normal vectors are calculated using the Levenberg-Marquardt method, as shown in (3).

$$(\alpha, \beta, \gamma) = \arg\min_{\alpha,\beta,\gamma} \sum_{k=1}^{n} d_k^2$$

The amount of rotation refers to the extent of attitude change relative to the world coordinate system of the indoor environment, as depicted in Fig.3. Using the calculated rotation matrix, the 3D point clouds from multiple laser beams can be fused together, enabling the measurement of the entire indoor environment's shape.

III. SIMULATION EXPERIMENT

The simulation experiment environment was set to a low-texture indoor scene [11] in the 3D computer graphics integrated development environment Blender, as shown in Fig.9. In the simulation environment, the true values of camera rotation and distance to the measurement target can be obtained, enabling quantitative evaluation.

The input images had a resolution of 1920×960. The camera was placed indoors in a way that the image center was

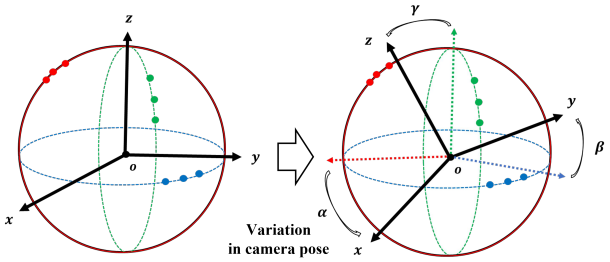


Fig. 8. Change in line features after camera pose change

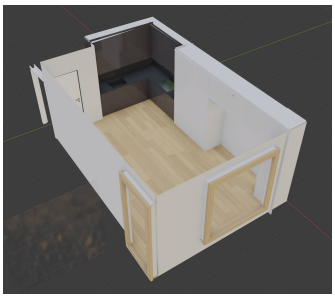


Fig. 9. Indoor room

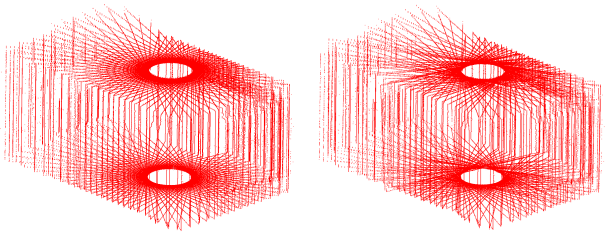
aligned with one direction in the world coordinate system. The rotation axis was set to be only in the vertical direction, and the rotation per cycle was set to 5 degrees. The camera's pose was rotated 72 times to cover a complete 360 degrees view around the environment. The results of the simulation experiment were evaluated in two aspects: the variation in camera pose and the accuracy of 3D measurement.

A. Rotation Results

Quaternion rotations were calculated to assess the camera pose estimation based on linear features. The estimation results are presented in Table 1. The error angle was obtained by subtracting the estimated angle from the true value of the quaternion angle, and the standard deviation of the error angle for the 72 measurements was 1.31 degrees. This error is primarily caused by the process of extracting straight lines in the omnidirectional image of the room using deep learning. The results of the straight-line detection in Fig.6(b) indicate that the output image includes a detected straight line at a position that deviates from the actual room edge. This can be interpreted as an incorrect camera posture estimation, representing a failure in the process.

B. 3D measurement results

Fig.10(a) displays the point cloud of the entire room fused based on the known camera posture, while Fig. 10(b) shows the result using the proposed method. To evaluate the impact of the posture estimation result on the fusion of the 3D point cloud, we utilized the 3D point cloud processing software Cloud Compare. The direction vector of the wall plane at 1.93 meters from the camera, which is sensitive to the amount of vertical rotation, and the plane error are calculated and shown in Table 2. The true value of the direction vector is $(\mathbf{1}, \mathbf{0}, \mathbf{0})$, and the plane error represents



(a) True rotation (b) Estimated rotation

Fig. 10. The point cloud of the entire blender room

the standard deviation of the distance error of each point group from the estimated plane. The results are depicted in Fig.11, which shows the point clouds for plane estimation, and Fig.12, which displays the resulting planar area.

Fig.10(b) indicates that the 3D point cloud in the room matches the Blender model, demonstrating the effectiveness of the proposed method. However, even with the known camera pose estimation, the direction vectors are not precisely $(\mathbf{1}, \mathbf{0}, \mathbf{0})$ as expected. Instead, they are measured as $(1.00, 6.20 \times 10^{-3}, -0.54 \times 10^{-3})$, leading to a plane error value of 23.4 mm. The reason for this error in the known amount of rotation case is attributed to the distortion of the omnidirectional image. This because of characteristics of a spherical camera. Consequently, the upper and lower portions of the image, such as the floor and ceiling in Fig.6(a), are projected with stretching. This stretching effect also occurs when the laser-irradiated area is projected onto the omnidirectional image, potentially affecting the calculation of the laser's weighted center of gravity.

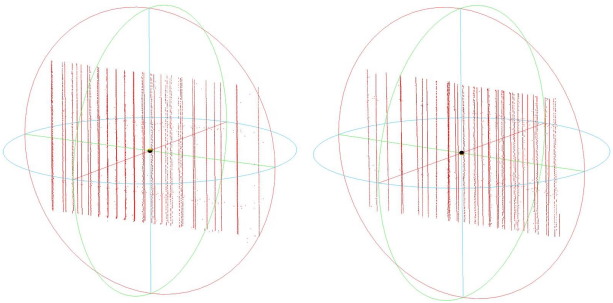
Comparing the results between the case where the amount of rotation is already known and the case where the amount of rotation is estimated, the error value of the direction vector and the plane error are larger in the latter case. This can be attributed to errors in the camera's orientation estimation, in addition to the distortion of the omnidirectional image.

TABLE I Quaternion rotations

Frame	True angle [°]	Estimated angle [°]	Error angle [°]
1	0.00	0.812	-0.812
2	5.00	5.48	-0.483
3	10.0	9.98	-0.0674
4	15.0	15.5	-0.512
5	20.0	17.8	2.20
6	25.0	24.5	0.471
7	30.0	29.3	0.729
8	35.0	34.4	0.611
9	40.0	43.3	-3.34
10	45.0	45.2	-0.210

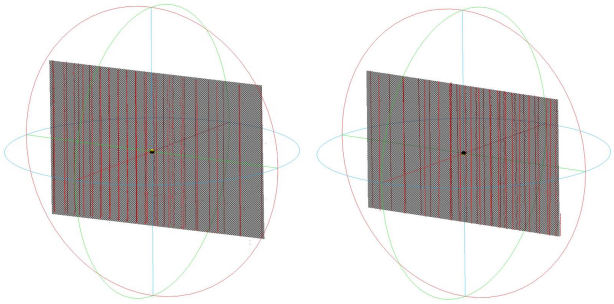
TABLE II Quaternion rotations

Rotation	Normal vector	Planar deviation
True	$(\mathbf{1.00}, \mathbf{6.20 \times 10^{-3}}, \mathbf{-0.54 \times 10^{-3}})$	23.4 [mm]
Estimated	$(\mathbf{1.00}, \mathbf{8.13 \times 10^{-3}}, \mathbf{-1.51 \times 10^{-3}})$	39.9 [mm]



(a) True rotation (b) Esttimated rotation

Fig. 11. Point cloud of a plane



(a) True rotation (b) Esttimated rotation

Fig. 12. Estimated plane area

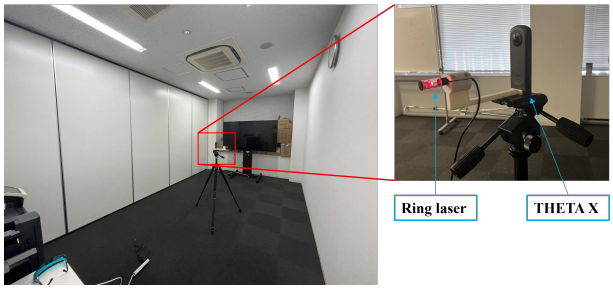
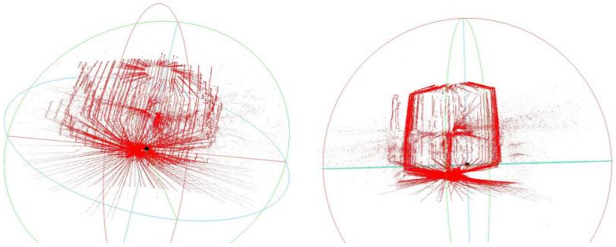


Fig. 13. Room 2629 and measurement device



(a) Oblique viewpoint (b) Side viewpoint

Fig. 14. The point cloud of the entire room 2629

IV. Experiments in a Real-World Environment

The experimental environment and measurement devices are shown in Fig. 13. The experiment was conducted in Room 2629 at the Korakuen Campus of Chuo University. The experimental apparatus is a RICOH THETA X and a 19.6mW ring laser. The measurement method is the same as in the simulation experiment. In this chapter, we only describe the results of the 3D measurements.

A. 3D measurement results

Fig.14(a),(b) shows the point cloud of the entire room from two viewpoints. Considering the shape of the room to be measured, it is clear that the point cloud is not correctly reconstruceted. The reason for this is that the positional relationship between the camera and the laser is not accurately acquired. Currently, the point cloud is calculated using an approximate positional relationship based on the dimensions of the measurement device. It is necessary to calibrate the distance and posture of the camera and laser.

V. Conclusion

We proposed a method for 3D measurement of indoor environments using a fisheye camera and circular laser projection. By focusing on lines in the images, it becomes possible to estimate camera poses even in textureless spaces, and multiple laser point clouds can be fused. Utilizing the characteristics of the fisheye camera, the normal vectors of lines in spherical coordinates were calculated, defining them as line features. The distribution of these line features changes as the camera's pose varies. The amount of this change was computed using the Levenberg-Marquardt method to estimate the camera's pose. Simulation experiments were conducted using Blender to evaluate the change in camera pose and the plane estimation of 3D point clouds. This confirmed the effectiveness of the proposed method.

While this study utilized deep learning for line extraction, it was observed that the accuracy of this detection affected the accuracy of camera pose estimation. Therefore, for further improvement in accuracy, combining methods like the Canny edge detection [12] alongside deep learning could enhance the line extraction process. In addition, good results have not been obtained in real-world conditions. A calibration method needs to be devised to determine the exact positional relationship between the camera and the laser.

References

[1] M. Golparvar-Fard, J. Bohn, J. Teizer, S. Savarese, and F. Pena-Mora, Evaluation of image-based modeling and laser scanning accuracy fore-merging automated performance monitoring techniques, Automation in Construction, vol. 20, no. 8, pp. 1143–1155, 2011.

[2] P. Besl and N. McKay, A method for registration of 3-dshapes, IEEE Transactions on Pattern Analysis and Machine Intelligence, 14-2, 239/256 (1992)

[3] F. Dellaert, D.Fox, W. Burgard and S. Thrun, Monte Carlo Localization for MobileRobots, In Proc. of IEEE International Conference on Robotics and Automation, 1322/1328 (1999)

[4] ME Warren, [Automotive LIDAR Technology], 2019 Symposium on VLSI Circuits, Kyoto,Japan,2019,ppC254-C255.

[5] M. Kawata, H. Higuchi, S. Pathak, A. Yamashita, and H.Asama, Scale Optimization of Structure from Motion for Structured Light-based All-round 3D Measurement, 2021IEEE International Conference on Systems, Man, and Cybernetics(SMC), Melbourne, Australia, 2021, pp. 442-449.

[6] J. L. Schonberger and J.-M. Frahm, Structure-from-motion revisited," in Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR2016), Las Vegas, Nevada, USA, 2016, pp. 4104-4113.

[7] T. Goto, S. Pathak, Y. Ji, H. Fujii, A. Yamashita and H. Asama, Line-based Global Localization of a Spherical Camera in Manhattan Worlds, Proceedings of the 2018 IEEE International Conference on Robotics and Automation (ICRA2018), pp. 2296-2303, Brisbane(Australia), May 2018.C. J. Kaufman, Rocky Mountain Research Lab., Boulder, CO, private communication, May 1995.

[8] J. Coughlan, and A. Yuille, The Manhattan World Assumption: Regularities in Scene Statistics which Enable Bayesian Inference, NIPS (2000).

[9] H. Li, H. Yu, W. Yang, L. Yu and S. Scherer, ULSD: Unified Line Segment Detection across Pinhole,Fisheye, and Spherical Cameras, ISPRS Journal of Photogrammetryand Remote SensinVolume 178, August 2021, pp. 187-20.

[10] D.W. Marquardt, An Algorithm for Least-Squares Estimation of Nonlinear Parameters, Journal of the Society for Industrial and Applied Mathematics, vol.11, no.2, pp. 431-441, 1963.

[11] `https://www.blender.org/download/demo-files/cycles`

[12] J. Canny, A Computational Approach to Edge Detection", IEEE Trans. Pattern Anal. Machine Intell, Vol. 8, No. 6, pp. 679-698 (1986).