Mapping in Indoor Environments Including Transparent Objects Using Stereo Polarization Camera and Projector

Yusuke Ogihara[1], Hiroshi Higuchi[2], Takuya Igaue[2], Qi An[1] and Atsushi Yamashita[1]

***Abstract*— This paper proposes a method for generating maps in indoor environments that include transparent objects by using a stereo polarization camera and projector. Conventional sensors like LiDAR and stereo cameras struggle with glass, as they rely on diffuse reflection, while glass allows light to pass through. In contrast, polarization cameras can measure light polarization and estimate surface normals, enabling depth estimation by combining polarization and RGB information. However, when measuring transparent objects, reflected and transmitted light cancel each other out, reducing polarization contrast, and the RGB information causes the depth estimation to output the depth of objects behind the glass. To address this issue, this paper proposes a novel method that (1) improves the S/N ration in polarization measument via diffuse reflection on non-glass regions and (2) masks out the RGB color from polarimetric depth estimation to not compute depth map of objects behind the glass to obtain depth images that include glass surfaces. Additionally, (3) in the mapping part, depth estimation is repeated at multiple locations, and the results are integrated using self-localization to generate a complete environmental map. Experiments in an indoor environment confirmed the effectiveness of the proposed method, enabling glass-inclusive depth estimation and successful map generation on a mobile robot.**

I. INTRODUCTION

In recent years, the demand for robot operations in indoor environments, such as office buildings, has increased, for security and guidance applications. When these robots use onboard sensors for self-localization and map generation, obtaining depth images is crucial to represent the distances between the camera and surrounding objects. However, indoor environments often contain many glass structures, such as fences, windows, and doors. Detecting and mapping these glass surfaces is essential for the safe navigation of robots [1].

Light Detection and Ranging (LiDAR) and stereo cameras are commonly-used distance sensors for map generation. LiDAR measures distances using the Time of Flight (ToF) method, determining the time it takes for the laser pulse to travel from the robot to surrounding objects and back. Meanwhile, stereo cameras use multiple cameras with a known baseline distance between cameras. By applying the principles of triangulation, stereo cameras measure the distance from the robot to surrounding objects.

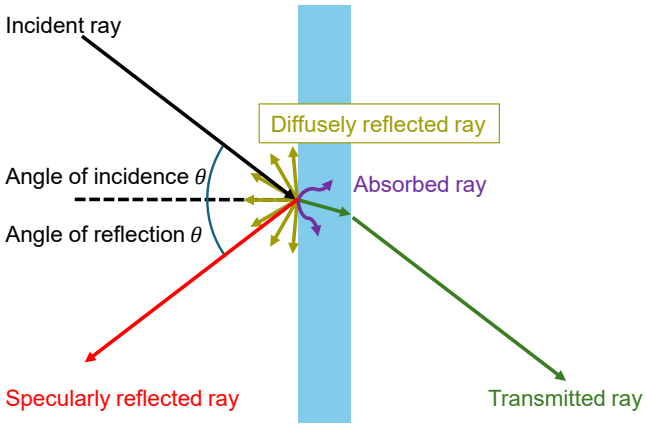


Fig. 1: Optical path of a light ray upon entering a dielectric material. When the incidence angle θ is very small, the specularly reflected light returns in nearly the same direction as the incident light. Consequently, at such angles, LiDAR can estimate the depth of the glass.

As shown in Fig. 1, a light ray that incidents dielectric objects can follow one of four paths: diffuse reflection, absorption, transmission, or specular reflection. Generally, when measuring the coordinates of opaque objects, LiDAR detects diffusely reflected light, which is independent of the incidence angle, allowing for robust distance measurement from various positions. Meanwhile, when measuring the corrdinates of glass, the transmitted and specularly reflected light does not directly return to the LiDAR receiver except when the angle of incidence is close to zero. As a result, the depth output corresponds to the object out of or behind the glass rather than the glass itself. When the angle of incidence is small, specularly reflected light directly returns to the receiver, allowing LiDAR to output the distance of glass. However, because this occurs only within a narrow angular range, only a few three-dimensional (3D) coordinates are obtained, making them susceptible to removal as outliers during map generation. Consequently, problems remain in measuring the distance to glass using LiDAR.

Stereo cameras estimate depth through triangulation by establishing correspondences between points captured by the left and right cameras, using stereo matching or feature point extraction and matching. Stereo cameras can be categorized into passive stereo, which relies on textures in the environ-

*This work was in part supported by JSPS KAKENHI Grant Number 22H03666.

[1]Y. Ogihara, Q. An and A. Yamashita are with the Department of Human and Engineered Environmental Studies, Graduate School of Frontier Sciences, The University of Tokyo, 5-1-5 Kashiwanoha, Kashiwa, Chiba 277–8563, Japan {ogihara, anqi, yamashita}@robot.t.u-tokyo.ac.jp

[2]H. Higuchi and T. Igaue are with the Department of Precision Engineering, Graduate School of Engineering, The University of Tokyo, Tokyo 113-8656, Japan {higuchi, igaue}@robot.t.u-tokyo.ac.jp

ment, and active stereo, which projects light patterns using a projector or laser. In passive stereo, glass is transparent, meaning that the matching process associates points using textures from objects behind the glass, leading to depth outputs corresponding to the background objects rather than the glass itself. In active stereo, the projected light follows a path similar to LiDAR. Therefore, except when the incidence angle is close to zero, the projected light does not return to the receiver, making depth measurement challenging. Thus, regardless of whether passive or active stereo methods are used, measuring the depth of glass remains highly challenging.

As discussed above, both LiDAR and stereo cameras, which are widely used for robot map generation, face significant difficulties in accurately measuring the coordinates of glass surfaces.

Meanwhile, in the field of computer vision, methods using a polarization camera have been proposed for measuring the shape of transparent objects [2]. Non-polarized light that enters a transparent object linearly polarizes when it reflects and transmits. By measuring the polarization of transparent objects under a non-polarized light source, it is possible to estimate surface normals. Polarization cameras can acquire both the angle of polarization corresponding to the phase of polarized light and degree of polarization corresponding to amplitude. Shao et al. proposed a method for estimating the normals of transparent objects using a single polarization camera [3]. This method models the polarization of light entering glass as a combination of transmission and reflection to estimate surface normals. However, it requires prior knowledge of the light source position and background shape, making it challenging to apply in real-world robot navigation.

A machine learning-based stereo depth estimation method, DPSNet [4], incorporates polarization and RGB images to output disparity maps. This approach enables depth estimation in textureless regions by leveraging both RGB and polarization information. However, because RGB information is used for depth estimation across all regions, areas with textures beyond transparent objects can influence the estimated depth. Also, when the specularly reflected light is weak and the transmitted light from objects behind glass is strong, polarization cannot be accurately measured, degrading the depth estimation accuracy.

To address these limitations, this paper proposes a method for generating environmental maps that include the 3D coordinates of glass under indoor lighting conditions suitable for robot navigation.

II. RELATED WORKS

A. Measurement of Transparent Object Using LiDAR

Jiang et al. developed a neural network that utilizes two characteristics of glass: (1) it is only detectable within a narrow range, and (2) the specular reflection from glass surfaces exhibits high reflection intensity [5]. Yamaguchi et al. combined the neural network-based classification method with glass detection using a polarization camera [6]. While their method detects glass using image-based techniques, it must guess the glass's location from LiDAR measurements of the surrounding frame.

Some LiDAR devices feature a dual-return function, which can capture two distance points along a single laser direction. Some studies have utilized the dual-return function of LiDAR for glass detection [7], [8]. Although these methods enable the generation of maps that include glass coordinates, they have a limitation in that they depends on the transparency of the glass, making it difficult to detect highly transparent glass.

Foster et al. [9] proposed a method for glass classification by analyzing measurement angles for each voxel. While opaque objects are measured from various directions, glass and moving objects are detected from a narrow range of angles. Their method identifies such points as potential glass or moving objects and distinguishes glass using a unique H-shaped structure in the 3D point cloud, formed when LiDAR scans glass perpendicularly. However, this limits them to upright glass that the robot has driven directly in front of.

As discussed above, methods using LiDAR face the limitation that glass detection and range measurement are only possible at the height of the LiDAR sensor.

B. Measurement of Transparent Object Shapes Using Polarization Cameras

Polarization cameras, which can capture polarization information, are used for measuring the shape of transparent objects. When unpolarized light enters a transparent object, it undergoes linear polarization upon reflection and transmission at the object's surface. By measuring the polarization of transparent objects under an unpolarized light source, it is possible to estimate the surface normals of the object. A polarization camera can acquire both the polarization angle, which corresponds to the phase of polarization, and degree of polarization, which corresponds to the amplitude.

Miyazaki et al. developed a method to measure the 3D shape of transparent objects using polarization cameras, utilizing polarization angles [10]. However, this method assumes that both the light source intensity and its polarization state are known in advance.

Shao et al. proposed a method for estimating the surface normals of transparent objects using a single polarization camera [3]. Their approach formulates the polarization of light upon entering glass as a combination of transmitted and reflected components. Because the polarization angles of reflected and transmitted light differ by 90 degrees, they cancel each other out, reducing the degree of polarization. This corresponds to a decrease in amplitude of the polarization signal, leading to a lower signal-to-noise ratio (S/N) for the polarization data.

As a result, the polarization information of the light reflected from the glass surface becomes difficult to obtain due to the influence of transmitted light. To mitigate the effects of transmission, this method imposes constraints on the measurement environment, such as using a black, flat background and uniform light source. However, in indoor

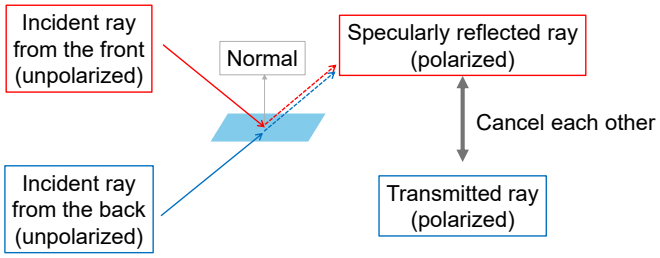


Fig. 2: Optical path when measuring a transparent object. The specularly reflected light is polarized perpendicular to the plane of incidence, while the transmitted light is polarized parallel to the plane of incidence, causing them to cancel each other out.

environments, fluorescent lights and other light sources often exist behind glass, making it challenging to control both the background and lighting conditions.

All of these methods impose strong constraints on light sources and backgrounds, making them unsuitable for environments where robots operate, where the positions of light sources are unknown and the background conditions are highly variable.

C. Application of Polarization Cameras to Mobile Robots

As an application of polarization cameras in mobile robots, methods utilizing stereo polarization cameras for depth estimation have been proposed.

Berger et al. developed a robust stereo matching method for specular reflective objects using a polarization camera [11]. This method employs polarization angles for matching regions with high degrees of linear polarization. However, if the object behind a transparent object has a texture, there is a risk that the distance to the background object will be obtained instead of the transparent object.

A machine learning-based depth estimation method using stereo polarization cameras, known as DPSNet [4], has also been proposed. DPSNet is a neural network that takes polarization and RGB images as input and outputs a disparity map. By combining RGB and polarization information, it enables depth estimation in textureless regions. The network was trained using a synthetic indoor dataset as well as real datasets collected in indoor environments, such as auditoriums and rooms, making it applicable to diverse environments with varying lighting conditions and backgrounds where robots operate.

However, in environments containing glass, these methods are designed to output the depth of objects behind the glass rather than estimating the depth of the glass itself.

III. PROPOSED METHOD

A. Overview of the Proposed Method

Challenges in measuring transparent objects with a polarization camera include the following: (1) Low S/N in Polarization Images: As shown in Fig. 2, the polarizations of transmitted and reflected light cancel each other out, reducing the S/N. (2) Depth Estimation: In RGB images, the textures of objects behind the transparent object are captured, causing the estimated depth to correspond to the background rather than the glass itself. To address these challenges, the proposed method (1) imporoves S/N ration in polarization measument via diffuse reflection on non-glass regions and (2) masks out the RGB color from polarimetric depth estimation to not compute the depth map of objects behind the glass. Additionally, to improve interpretability and usability for the downstream process, (3) in the mapping part, depth estimation is repeated at multiple locations and the generated environmental map includes glass.

Figure 3 depicts the flow of the proposed single-location depth estimation system. In the proposed method, selective projection of non-glass regions is performed to enhance the S/N of polarization information. Because this projection requires glass segmentation to obtain the non-glass regions, the proposed method captures images twice. Unlike DPSNet, which uses only images captured under ambient lighting for depth estimation, the proposed method combines images taken under ambient lighting with those captured under selective projection for depth estimation. In the first capture, images are taken without projection to segment glass regions. Based on the segmentation results, selective projection is performed onto non-glass areas during the second capture. Because direct light does not reach objects behind the glass, the amount of light transmitted through the glass and reaching the camera is minimized. Additionally, light projected onto the surrounding walls and window frames reflects off the glass surface, enabling the capture of polarization information from the reflected light. Therefore, the second capture is performed with selective illumination, and the obtained image is combined with the first capture's image to serve as input for the depth estimation algorithm. In the mapping part of the proposed method, the depth maps estimated at multiple locations are integrated to generate an environmental map.

B. Selective Projection on the Opaque Regions

When measuring polarization in indoor environments, the angles of polarization of transmitted and reflected light differ, and in the case of transparent objects, transmitted and reflected light cancel each other's polarization and reduces the degree of polarization, i.e., the S/N of the polarization information. The proposed method reduces the contribution of transmitted light by selectively projecting light onto non-glass regions, allowing the acqisition of polarization information from reflected light.

In the first capture, images are taken under ambient lighting without projection. A stereo polarization camera can capture both RGB and polarization images for the left and right cameras. The RGB and polarization images obtained from the left camera are denoted as ${}^{L}\mathbf{I}_c \in \mathbb{R}^{h\times w\times 3}$ and ${}^{L}\mathbf{I}_p \in \mathbb{R}^{h\times w\times 2}$, respectively. Similarly, the RGB and polarization images obtained from the right camera are denoted as ${}^{R}\mathbf{I}_c \in \mathbb{R}^{h\times w\times 3}$ and ${}^{R}\mathbf{I}_p \in \mathbb{R}^{h\times w\times 2}$, respectively. The polarization images consists of two chaccels, each representing the degree of polarization and the angle of polarization. Here, $h, w \in \mathbb{N}$

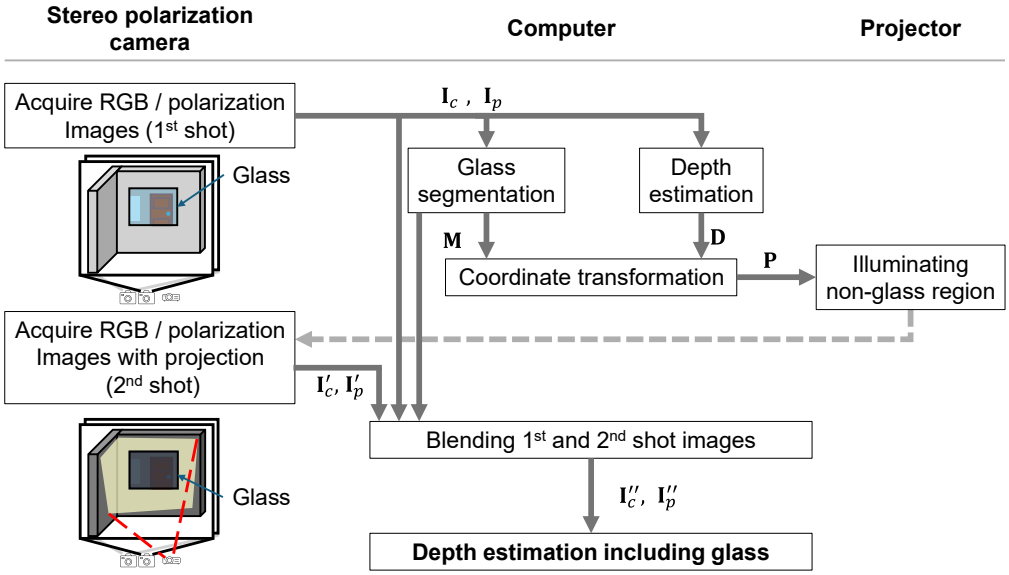


Fig. 3: Flow of the proposed depth estimation system.

represent the height and width of the acquired images in pixels, respectively. The subscripts L and R indicate images captured by the left and right cameras, respectively. However, in the following, these subscripts are omitted when the same processing is applied to both cameras.

Glass segmentation is required to selectively project light onto non-glass regions. Here, $\mathbf{I}_c$ and $\mathbf{I}_p$ are the inputs, and binary images $\mathbf{M} \in \mathbb{R}^{h \times w}$ that represents glass regions, as shown in Fig. 4, are obtained from the segmentation results:

$$\mathbf{M}(i,j) = \begin{cases} 1 & \text{if glass is included in the pixel} \\ 0 & \text{else} \end{cases}. \quad (1)$$

To detect glass regions, we employ PGSNet [12], a glass segmentation algorithm that performs segmentation using polarization and RGB information. Although PGSNet enables glass segmentation, the unstable contour shapes and potential false detections make it difficult to directly estimate depth from the segmentation results.

Based on the segmentation results, images for selective projection onto non-glass regions are created. Because the segmentation results $\mathbf{M}$ represent glass regions in the image coordinate system, coordinate transformation to the projector's coordinate system is necessary for projection. The following describes the process of transforming the segmentation result from the left camera into the projector coordinate system. The same procedure is applied to the right camera segmentation result.

This transformation requires the intrinsic parameters of the polarization cameras and projector $\mathbf{K}_L, \mathbf{K}_P \in \mathbb{R}^{3\times3}$ and the extrinsic parameters between the projector and camera, including the rotation matrix $\mathbf{R} \in \mathbb{R}^{3\times3}$ and translation vector $\mathbf{t} \in \mathbb{R}^3$, as well as depth information $\mathbf{D} \in \mathbb{R}^{h\times w}$. Therefore, depth estimation is performed using the polarization and RGB images from the first capture. At this stage, the estimated depth corresponds to objects behind the glass, and the depth of the glass itself cannot be obtained. First, the 3D coordinates $(x, y, z) \in \mathbb{R}^3$ of pixels identified as glass in the left camera coordinate system are determined. Given a pixel $(i, j) \in \mathbb{R}^2$, where $\mathbf{M}(i, j) = 1$, its 3D coordinates can be computed using the intrinsic matrix of the left camera $\mathbf{K}_L$:

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = z \cdot \mathbf{K}_L^{-1} \begin{bmatrix} i \\ j \\ 1 \end{bmatrix}, \quad (2)$$

where z represents the depth, obtained from the depth image $z = \mathbf{D}(i, j)$.

Next, these 3D points are projected onto the projector image plane. The corresponding projector image coordinates $(u_P, v_P) \in \mathbb{R}^2$ for each (x, y, z) can be obtained using the intrinsic matrix of the projector $\mathbf{K}_P$ and the extrinsic parameters between the projector and the camera, $\mathbf{R}, \mathbf{t}$:

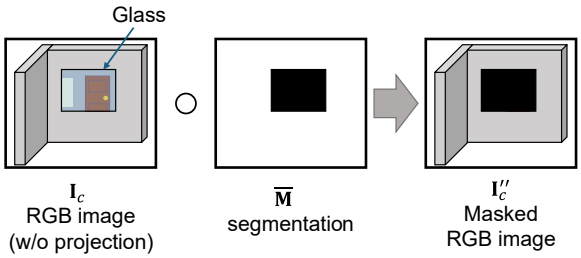
$$s \begin{bmatrix} u_P \\ v_P \\ 1 \end{bmatrix} = \mathbf{K}_P \, [\mathbf{R}|\mathbf{t}] \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix}, \quad (3)$$

where $s \in \mathbb{R}$ is a scale parameter.

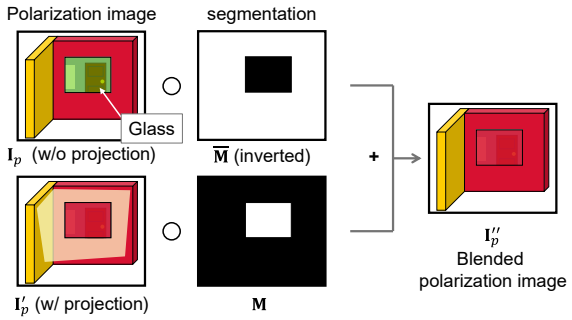
Through this transformation, the projector image coordinates (u_P, v_P) corresponding to pixels identified as glass in the camera image are obtained. Because u_P and v_P are real-valued, they are rounded to the nearest integers: $i_P = \mathrm{round}\,(u_P)$ and $j_P = \mathrm{round}\,(v_P)$, where $\mathrm{round}\,(x)$ represents rounding x to the nearest integer. (i_P, j_P) forms glass segmentation result ${}^P\mathbf{M}$ in the projector coordinate system through make the projected glass pixels as 1 and the other pixels as 0:

$${}^P\mathbf{M}(k,l) = \begin{cases} 1 & \text{if } (k,l) = (i_P, j_P) \\ 0 & \text{else} \end{cases}. \quad (4)$$

To illuminate non-glass regions, the projector input image $\mathbf{P}$ is derived from ${}^P\mathbf{M}$. The derivation involves a bit-



(a) Blending RGB images.



(b) Blending polarization images.

Fig. 4: Blending images captured by the polarization camera.

inversion of ${}^{P}\mathbf{M}$, followed by the application of morphological closing operations, which serve as a filtering step for image refinement.

C. Masking RGB Information for Depth Estimation

In the second capture, the stereo polarization camera captures images while the projector illuminates non-glass regions: $\mathbf{I}_c' \in \mathbb{R}^{h\times w\times 3}$ and $\mathbf{I}_p' \in \mathbb{R}^{h\times w\times 2}$.

As shown in Fig. 4, the images from the first and second captures are integrated to create RGB and polarization images for input into the depth estimation algorithm, $\mathbf{I}_c'' \in \mathbb{R}^{h\times w\times 3}$ and $\mathbf{I}_p'' \in \mathbb{R}^{h\times w\times 2}$, respectively. The following equations describe the integration process:

$$\mathbf{I}_c'' = \mathbf{I}_c \circ \overline{\mathbf{M}}, \tag{5}$$

$$\mathbf{I}_p'' = \mathbf{I}_p \circ \overline{\mathbf{M}} + \mathbf{I}_p' \circ \mathbf{M}, \tag{6}$$

where $\circ$ represents element-wise multiplication, and $\overline{\mathbf{M}}$ is the bit-inverted binary image of $\mathbf{M}$.

The proposed method uses DPSNet [4] to do final depth estimation by replacing the RGB and polarization input images with the generated images $\mathbf{I}_c''$ and $\mathbf{I}_p''$.

D. Multi-Location Depth Integration for Map Generation

The robot moves slightly to ensure that the images taken at the previous and current locations have overlapping regions, and then it stops and captures two images, one with projector illumination and one without. It then moves slightly again and repeats this process.

At each imaging location, the aforementioned single-location depth estimation is applied, allowing the generation of a 3D point cloud from the depth images. Because these point clouds are initially expressed in the coordinate system of the left camera at each location, estimating the camera's position and orientation at each imaging location is necessary to construct a unified environmental map. One approach to estimating the camera's position and orientation from multiple images is Structure from Motion (SfM) [13]. As this method performs map generation as an offline post-processing step after the robot completes its movement, we use COLMAP [14], an offline SfM method.

In the SfM part of the proposed method, the RGB images captured without projector illumination are used. This is because, for self-localization, not only the textures of opaque objects surrounding the glass but also the textures of objects visible through the glass can serve as useful feature points.

Although only the left camera's position and orientation are required for coordinate integration, the RGB images captured by the right camera are also input into the SfM process. This is because the point cloud obtained from SfM is scale-ambiguous. By incorporating images from both the left and right cameras into SfM, it is possible to restore the correct scale. Because the baseline distance between the left and right cameras can be obtained through stereo calibration, comparing it with the distance obtained from SfM allows for scale recovery.

IV. EXPERIMENTS

A. Experimental Settings

An experiment was conducted to demonstrate the feasibility of depth estimation for glass using the proposed method. In this experiment, as shown in Fig. 5, the LUCID VP-PHX050S-Q was used as the polarization camera, and the EPSON EB-X36 was used as the projector. Experiments were conducted on glass placed on a table and glass facing a hallway. The tabletop experiment aimed to demonstrate that selective projection and RGB image masking of the proposed method enhance the S/N of polarization images, enabling depth estimation. The hallway-facing glass experiment evaluated the feasibility of depth estimation and map generation in a hallway environment, which closely resembles real-world conditions where a robot would operate. Both experiments were conducted indoors under fluorescent lighting conditions.

1) Settings of the Tabletop Experiment: This experiment evaluated the S/N improvement of polarization images through selective illumination and the depth estimation of glass. The measurement target was a flat glass panel with dimensions of 295 mm in height, 300 mm in width, and 4.8 mm in tickness. The measurement target was fixed on a table approximately 2 meters away from the measurement device. To evaluate the proposed method for glass depth estimation, we compared three scenarios: (1) without masking or selective illumination, (2) masking only, and (3) both selective illumination and masking (proposed method). Depth estimation using DPSNet was performed for each case, and the resulting depth estimations were compared.

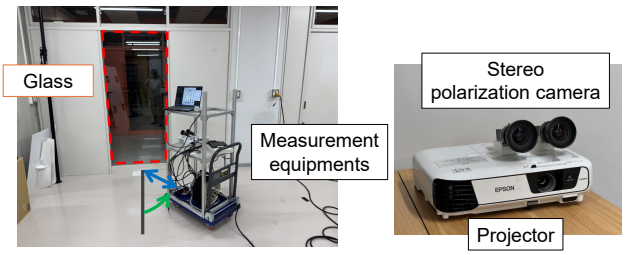


Fig. 5: The experiment was conducted using the measurement device from the right image in the environment shown in the left image. In the left image, the blue arrow represents the distance from the glass, while the green arrow indicates the angle.

2) Settings of the Hallway-Facing Glass Experiment: This experiment evaluated depth estimation accuracy from a single viewpoint and map generation results from multiple viewpoints. In the evaluation of depth estimation accuracy from a single viewpoint, the measurement target was a flat glass panel with dimensions of 2.1 m in height and 0.9 m in width.

Because the proposed method involves projection using a projector during measurement, it is assumed that the measurement accuracy depends on the distance from the target due to light intensity attenuation over distance. Additionally, because the degree and angle of polarization of light reflected and transmitted through glass depend on the angle of incidence, the measurement angle is also considered to influence accuracy. Therefore, measurements were conducted from 12 different positions with distances of 2 m, 3 m, and 4 m from the glass, and angles of 0°, 30°, 45°, and 60° between the glass normal and the polarization camera's optical axis.

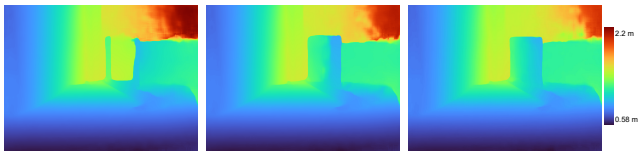
AR markers were attached to the four corners of the glass to obtain ground truth depth values for the glass plane, and the accuracy of the depth estimation results was evaluated. The ArUco library [15] was used for the generation and detection of AR markers. For comparison, depth estimation was also performed using DPSNet without applying the proposed method, inputting RGB and polarization images captured directly through the glass.

The accuracy of depth estimation using the proposed method was evaluated. The mean absolute error (MAE) for each distance at angles of 0°, 30°, 45°, and 60° is shown in Table I. Depth maps for an incident angle of 0° are shown in Fig. 8, using a turbo colormap to represent depth, with cool colors indicating closer regions and warm colors indicating farther regions. The color scale varies depending on the measurement distance.

In evaluation of map generation from multi viewpoints, the measurement device was fixed on a cart to perform measurements from multiple viewpoints. The device was mounted at a height of 0.75 m above the ground, facing diagonally forward to capture the hallway walls and glass within the field of view of both the polarization camera and projector. Measurements were performed by manually moving the cart. The measurement area included glass with dimensions of 2.1 m in height and 0.9 m in width. The measurements were conducted in an indoor hallway at night. Data from six positions along a 2.0 m section were integrated.

(a) RGB image (w/o projection). (b) DoP image (w/o projection). (c) DoP image (w/ projection).

Fig. 6: Images captured using polarization camera. In DoP images, the whiter the color, the higher DoP.



(a) w/o masking. (b) Only masking. (c) Proposed.

Fig. 7: Depth maps in the tabletop experiment. The depth is colored with a turbo colormap and ranging from 0.58 m to 2.2 m.

B. Results

1) Tabletop Experiment: The RGB image obtained from the polarization camera in the tabletop experiment is shown in Fig. 6(a). Figures 6(b) and (c) shows the degree of polarization (DoP) image captured without and with projection, recpectively. The degree of polarization in the glass region is higher with projection than without. This result indicates that the proposed projection algorithm improves the S/N ratio of polarization images in the glass region. Figure. 7(a) shows the depth map without masking or selective illumination, where the depth inside the glass region corresponds to the distant wall behind it. Figure. 7(b) shows the depth map estimated with only masking in the RGB image, where the edges of the glass are correctly estimated, but unintended edges appear within the glass, despite it being a planar surface. Figure. 7(c) shows the depth estimation result of the proposed method (masking + selective illumination), the depth inside the glass region shows a gradual gradient, indicating that the actual depth of the glass is successfully estimated.

2) Experiment in Hallway: For an incident angle of 0°, the experimental results are presented in Table I and Fig. 8. At all distances, the MAE of the proposed method was smaller than that of the comparison method. Notably, at distances of 2 m and 3 m, the comparison method exhibited an MAEs close to 2 m, indicating that the depth of objects behind the glass was measured. This is evident in Fig. 8, where objects beyond the glass are visible. In contrast, the proposed method successfully estimated the depth of the glass itself without outputting the depth of objects behind it.

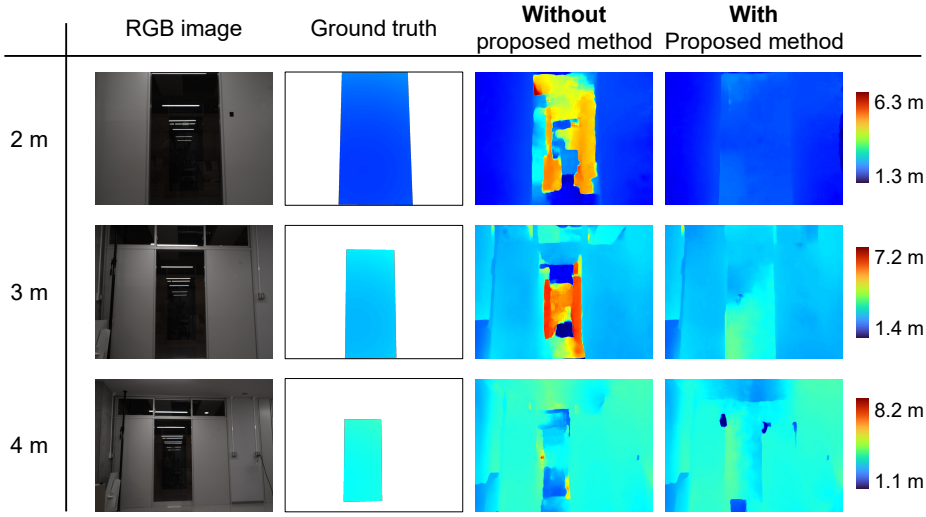


Fig. 8: Depth map at 0°. The depth of the glass was measured using an AR marker and used as the ground truth. Without the proposed method, the depth estimation corresponds to objects behind the glass, whereas with the proposed method, the estimated depth is closer to the ground truth.

For an incident angle of 30°, the results are shown in Table I. At all distances, the MAE of the proposed method was smaller than that of the comparison method. Although the comparison method also output the depth of the glass rather than the background at this angle, the proposed method provided depth estimates closer to the ground truth. Under the 3 m condition of the comparison method, the depths smaller than the glass depth were output within the glass region, likely due to reflections from fluorescent lights causing erroneous depth estimates, which were successfully corrected by the proposed method.

For an incident angle of 45°, the results are shown in Table I. At this angle, the MAE at 3 m was larger for the proposed method than for the comparison method. The degradation in estimation accuracy is attributed to overestimated depths near the glass edges in the center of the image.

For an incident angle of 60°, the results are shown in Table I. At all distances, the MAE of the proposed method was smaller than that of the comparison method. Under this condition, the texture behind the glass was barely captured in the RGB images, allowing the comparison method to output glass depths as well. However, the proposed method achieved higher accuracy in glass depth estimation through the combination of segmentation and selective projection.

Figure 9(a) shows the experimental environment. Figure 9(b) shows the environmental map generated using the proposed method. Figure 9(c) shows the map generation results using the comparison method, where depth estimation was performed solely with DPSNet from a single viewpoint. The planar shapes of the walls on either side of the glass were preserved, and point clouds existed in the regions where glass was present. This result indicates successful measurement of the hallway structure, including the glass. However, due to errors in self-position estimation and depth estimation, some opaque objects appeared as layered point clouds. Additionally, compared to opaque regions, the glass regions exhibited greater variability, with some significantly erroneous estimates.

TABLE I: Mean absolute error of depth estimation [m]

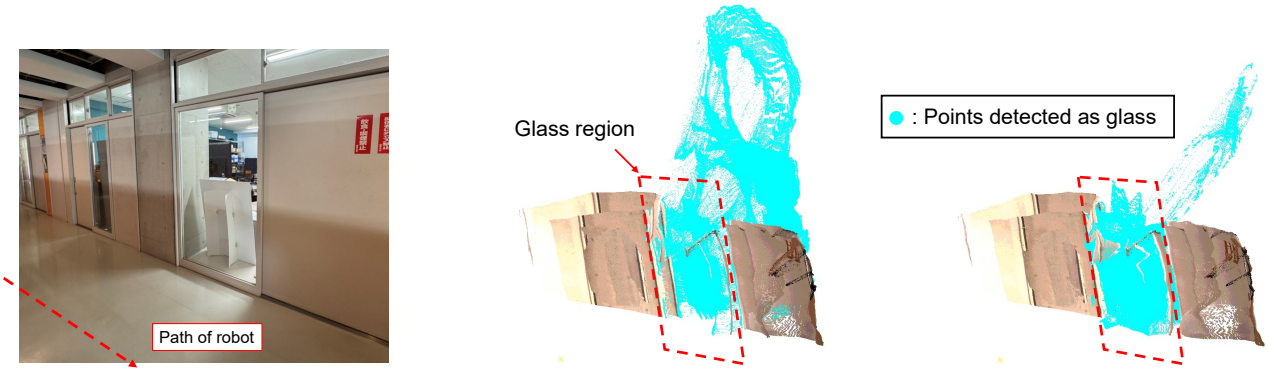
Incident angle	Distance	DPSNet	Proposed method
0°	2 m	1.37	0.09
	3 m	1.41	0.39
	4 m	0.60	0.37
30°	2 m	0.40	0.23
	3 m	0.47	0.27
	4 m	0.76	0.39
45°	2 m	0.57	0.26
	3 m	0.54	0.57
	4 m	0.93	0.24
60°	2 m	0.24	0.17
	3 m	0.30	0.28
	4 m	0.35	0.31

C. Discussions

The tabletop experiment indicated that the DoP of the glass region is improved using the proposed projection algorithm and both proposed projection algorithm and RGB image masking contribute to the accurate depth estimation of glass. In addition to enhancing the accuracy of masking by applying morphological closing to the segmentation result obtained by PGSNet, it was found that the presence of projection light resulted in higher accuracy compared to masking alone. The improvement in the S/N due to the projection was beneficial.

In the hallway experiment, the proposed method estimated accurate depth of glass from a single viewpoint. In the proposed method, images obtained from two captures are combined separately for RGB and polarization to generate the input for the depth estimation algorithm. In glass region, RGB information leads to estimate the depth of objects behind the glass. Therefore, by masking RGB information, the precision of glass depth estimation was improved.

In the environmental map, we confirmed that the self-localization results from SfM, using RGB images from the



(a) Experimental environment. (b) DPSNet. (c) Proposed method.

Fig. 9: Results of map generation. The cyan-colored points represent the region where the presence of glass is detected. The region containing glass is outlined with a red dashed line. In the proposed method, it was confirmed that the point cloud is present near the actual glass regions.

polarization camera, could be effectively utilized to integrate depth estimation results from each capture point. While we successfully measured the planar shape of hallway walls, including glass, we observed that the accuracy of the estimated glass point cloud was lower than that of the opaque object regions. This discrepancy is likely because opaque objects are measured with absolute depth using RGB information, whereas for glass regions, depth is estimated relatively based on surface normals from surrounding objects.

V. LIMITATIONS AND FUTURE WORKS

In proposed method, segmentation and coordinate conversion need to be calculated between 1st capture and 2nd capture. Mapping experiment is conducted by repeatedly stopping at viewpoint and move. Future work should develop a continuous measurement system that allows robots to perform depth estimation without stopping at each viewpoint.

VI. CONCLUSIONS

This paper proposed a novel 3D map generation method that accounts for transparent objects in indoor environments. By utilizing a stereo polarization camera and a projector, the system selectively illuminated non-glass regions to enhance polarization-based depth estimation. Experimental results demonstrated the effectiveness of the proposed method in generating accurate depth and environmental maps, including glass structures. In a tabletop experiment, the proposed algorithm increased the DoP in the glass region. Also, masking of the glass region in RGB images removed the influence of the objects behind the glass in depth estimation. The accuracy of the depth estimation was improved using the two methods. The hallway expriment ensured that the proposed single-viewpoint depth estimation algorithm worked for actual glass windows. Also, in the map generation, the proposed method generated an accurate point cloud of glass.

REFERENCES

[1] P. Foster, Z. Sun, J. J. Park, and B. Kuipers, "Visagge: Visible angle grid for glass environments," in *Proceedings of the 2013 IEEE International Conference on Robotics and Automation*, 2013, pp. 2213–2220.

[2] X. Li, Z. Liu, Y. Cai, C. Pan, J. Song, J. Wang, and X. Shao, "Polarization 3d imaging technology: a review," *Frontiers in Physics*, vol. 11, p. 1198457, 2023.

[3] M. Shao, C. Xia, Z. Yang, J. Huang, and X. Wang, "Transparent shape from a single view polarization image," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023, pp. 9277–9286.

[4] C. Tian, W. Pan, Z. Wang, M. Mao, G. Zhang, H. Bao, P. Tan, and Z. Cui, "Dps-net: Deep polarimetric stereo depth estimation," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023, pp. 3569–3579.

[5] J. Jiang, R. Miyagusuku, A. Yamashita, and H. Asama, "Online glass confidence map building using laser rangefinder for mobile robots," *Advanced Robotics*, vol. 34, no. 23, pp. 1506–1521, 2020.

[6] E. Yamaguchi, H. Higuchi, A. Yamashita, and H. Asama, "Glass detection using polarization camera and lrf for slam in environment with glass," in *Proceedings of the 2020 21st International Conference on Research and Education in Mechatronics*. IEEE, 2020, pp. 1–6.

[7] X. Zhao, Z. Yang, and S. Schwertfeger, "Mapping with Reflection - Detection and Utilization of Reflection in 3D Lidar Scans," in *Proceedings of the 2020 IEEE International Symposium on Safety, Security, and Rescue Robotics*, 2020, pp. 27–33.

[8] H. Tibebu, J. Roche, V. De Silva, and A. Kondoz, "LiDAR-Based Glass Detection for Improved Occupancy Grid Mapping," *Sensors*, vol. 21, no. 7, p. 2263, 2021.

[9] P. Foster, C. Johnson, and B. Kuipers, "The reflectance field map: Mapping glass and specular surfaces in dynamic environments," in *Proceedings of the 2023 IEEE International Conference on Robotics and Automation*, 2023, pp. 8393–8399.

[10] D. Miyazaki and K. Ikeuchi, "Shape estimation of transparent objects by using polarization analyses," *Information and Media Technologies*, vol. 1, no. 2, pp. 1073–1093, 2006.

[11] K. Berger, R. Voorhies, and L. H. Matthies, "Depth from stereo polarization in specular scenes for urban robotics," in *Proceedings of the 2017 IEEE International Conference on Robotics and Automation*, 2017, pp. 1966–1973.

[12] L. Yu, H. Mei, W. Dong, Z. Wei, L. Zhu, Y. Wang, and X. Yang, "Progressive Glass Segmentation," *IEEE Transactions on Image Processing*, vol. 31, pp. 2920–2933, 2022.

[13] E. p. Herrera-Granda, J. C. Torres-Cantero, and D. H. Peluffo-Ordóñez, "Monocular visual slam, visual odometry, and structure from motion methods applied to 3d reconstruction: A comprehensive survey," *Heliyon*, vol. 10, no. 18, p. e37356, 2024.

[14] J. L. Schönberger and J.-M. Frahm, "Structure-from-Motion Revisited," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 4104–4113.

[15] S. Garrido-Jurado, R. Muñoz-Salinas, F. Madrid-Cuevas, and M. Marín-Jiménez, "Automatic generation and detection of highly reliable fiducial markers under occlusion," *Pattern Recognition*, vol. 47, no. 6, pp. 2280–2292, 2014.