Panoramic Ceiling Image Synthesis Method Prioritizing Fixture Outlines using an Omnidirectional Camera

Daiki Kohama*†, Yoshiteru Nagata*, Kazushige Yasutake‡, Kenta Urano*, Shin Katayama*,
Takuro Yonezawa*, Nobuo Kawaguchi*§
*Graduate School of Engineering, Nagoya University
†Email: kohama@ucl.nuee.nagoya-u.ac.jp
‡Kyudenko Corporation
§Institutes of Inovation for Future Society, Nagoya University

***Abstract*—The manual creation of a ceiling plan consumes many human resources to confirm the current state of existing buildings for the renovation. Especially the positions and the types of existing fixtures are important. In this paper, we propose a synthesis method for a panoramic ceiling image using a video shot by an omnidirectional camera. We utilize a Visual SLAM method to map the positions of omnidirectional video frames including the ceiling. We prioritize the clear depiction of fixture outlines for easier identification of the types of ceiling fixtures. To draw clear fixture outlines in the resulting image, each fixture is referred from the single video frame that captures it most clearly. In the experiment, we collected an omnidirectional video in the indoor area of approximately 273 square meters with multiple rooms. We evaluated the synthesized image for positional accuracy of the ceiling fixtures, and our proposed method demonstrated higher accuracy compared to the baseline.**

***Index Terms*—Omnidirectional Camera, Visual SLAM, Image Synthesis, Ceiling Plan**

I. Introduction

The utilization of three-dimensional (3D) models of buildings such as Building Information Modeling (BIM) and Construction Information Modeling (CIM) contributes to the efficiency of design and construction management [1]. There are numerous studies that have been conducted to explore the further utilization of these models after construction [2]–[4]. However, many existing buildings were constructed using two-dimensional (2D) blueprints. Especially in buildings that have been in operation for a long time, there are often cases where the contents of past facility renovations are not reflected or where blueprints are lost. It is rare to obtain blueprints that accurately reflect the current status of existing buildings.

When conducting electrical installation work during building facility renovations, a ceiling plan reflecting the current state is required. A ceiling plan is a blueprint that includes ceiling fixtures such as lighting fixtures, sprinklers, and ceiling access panels. Especially the positions and the types of fixtures are important. It is used for cost estimation, parts ordering, and project planning before construction. If an accurate ceiling plan reflecting the current state does not exist, the worker typically creates it without using tools before the renovation.

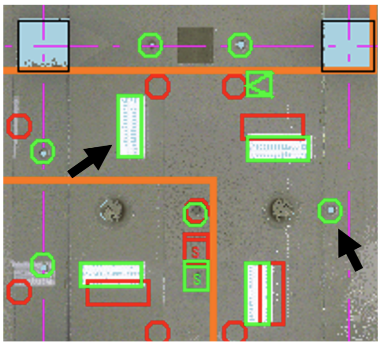

Fig. 1. Accuracy difference of ceiling fixture positions

The manual creation of a ceiling plan consumes many human resources, takes much time, and is not accurate in most cases. Fig. 1 shows the difference in positional accuracy of fixtures between the manual creation and the generation with a 3D laser scanner. When a worker creates a ceiling plan with visual confirmation, a lighting fixture and a sprinkler are overlooked. A 3D laser scanner can accurately measure ceiling fixtures. However, it is used only in limited construction sites because of its high cost and the requirement of 3D reconstruction knowledge to operate it.

In this study, we propose a synthesis method to generate a panoramic ceiling image using a video shot by an omnidirectional camera as shown in Fig. 2. A panoramic ceiling image is synthesized through some computer vision processes detailed in chapter III. For easier identification of the types of ceiling fixtures, the clear depiction of fixture outlines are prioritized. To draw clear fixture outlines in the resulting image, each fixture is referred from a single video frame that captures it most clearly. We use an omnidirectional video shot while walking underneath the ceiling for synthesis. A panoramic ceiling image captures all wide areas of the ceiling surface. Workers can easily identify the types of fixtures on a panoramic ceiling image because of their clearness. They can

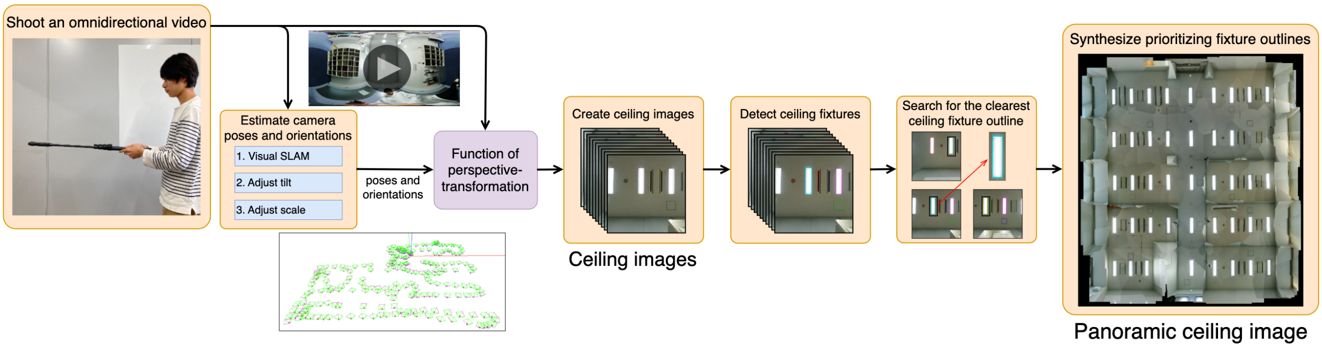

Fig. 2. Process of proposed method

generate a ceiling plan with it more easily.

We utilize a Visual Simultaneous Localization and Mapping (SLAM) method to map the positions of omnidirectional video frames including the ceiling. Omnidirectional cameras can capture a ceiling surface faster and cost lower than 3D laser scanners. The omnidirectional videos capture not only the ceiling surface but also the floor, walls, and everything visible from the camera position. Visual SLAM can estimate the shooting position and orientation more accurately, because of numerous image feature points captured in the frames of the omnidirectional video.

In the experiment, we collected omnidirectional videos in 2 indoor environments. One is an area of approximately 273 square meters with multiple rooms, the other is an open space of approximately 86 square meters. The synthesized image was evaluated for positional and shape accuracy of the ceiling fixtures. Our proposed method demonstrated higher accuracy compared to the commercial products' baseline.

II. Related Work

A. Applied Research of Visual SLAM with Omnidirectional Video

There is a study that applied Visual SLAM to the driving assistance of electric wheelchairs [5]. To perceive the surrounding environment of the wheelchair, camera poses and the sparse 3D model was generated using Visual SLAM based on omnidirectional videos.

Kayukawa et al. conducted a study that generates navigation movies walking in multi-level buildings automatically. [6]. They used Visual SLAM to estimate relative camera positions and orientations for each frame of the omnidirectional movie and prepared for generating the navigation movies.

B. Studies on Modeling Indoor Spaces with Omnidirectional Images

There has been active research in estimating a room layout from a single omnidirectional image [7]–[10]. However, the modeling coverage is limited when using only a single omnidirectional image, and it is not possible to measure large indoor spaces accurately.

Wang et al. conducted a study on layout estimation by considering the camera positions and orientations from two omnidirectional images to improve the accuracy of the estimation [11]. However, measuring large spaces such as commercial facilities or offices is not feasible. Additionally, if estimation is possible in narrow indoor spaces, it is not expected to achieve the panoramic ceiling image which is high-resolution and visually clear.

In the commercial software Matterport [12], it is possible to generate a panoramic ceiling image using multiple omnidirectional images captured at different locations. However, the successful generation of a high-quality image requires knowledge of the 3D reconstruction from multiple images. Additionally, the resulting panoramic ceiling image generated by Matterport may be lower resolution and have discontinuities in ceiling fixtures across captured images. It is difficult to determine the types of ceiling fixtures from it accurately.

Our proposed method is applicable even in a large environment. Shooting an omnidirectional video takes shorter time compared with capturing multiple images at each location. Furthermore, to improve the visibility of ceiling fixtures, our proposed method considers the clear depiction of fixture outlines.

III. Panoramic Ceiling Image Synthesis

In our proposed method, we first shoot a video of an indoor environment using an omnidirectional camera and then estimate the camera positions and orientations for each frame using a Visual SLAM. Perspective-transformed images called ”ceiling images” can be generated from each frame using the estimation results of the camera positions and orientations. Next, we detect rectangular and circular ceiling fixtures in the ceiling images using OpenCV. Then, extraction of the same fixtures across the ceiling images is conducted based on the criteria and we search for the image that captures those fixtures most clearly. Finally, we synthesize the panoramic ceiling image prioritizing the detected fixture outlines using positional relationship of ceiling images. For the estimation process of the orientations, we have to start the shooting from a location where a known-sized rectangular fixture is placed on

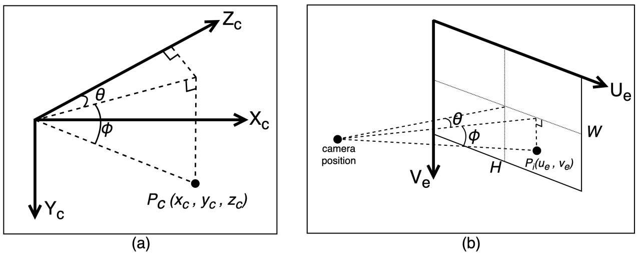

Fig. 3. An azimuth angle θ and an elevation angle in a camera cordinate system (a) and an equrectangular image (b)

the ceiling. We explain a detailed description of our proposed method.

A. Function of Perspective-transformation

We explain the function that generates a perspective-transformed image from an equirectangular image with dimensions $H \times W$ pixels. It is used to estimate camera positions and orientations and in the process of generating ceiling images. The resulting image is generated by mapping each pixel to its corresponding pixel on the equirectangular image.

We consider two coordinate systems: the camera coordinate system, where the Z-axis aligns with the optical axis during the equirectangular image capture, and the zenith coordinate system, which shares the same origin as the camera coordinate system, with the positive Z-direction pointing towards the zenith. The rotation matrix from the zenith coordinate system to the camera coordinate system is denoted as $\mathbf{R}$. We define P_z as the set of points in the zenith coordinate system that represents each pixel of the resulting image. The resulting image maps each pixel to a plane in the zenith coordinate system, where the plane is defined by $Z = d$ in millimeters and each pixel covers an area of a square with M millimeters long sides. The $P_z\left(x_z, y_z, z_z\right)$ satisfies equation (1).

$$\begin{cases} -\dfrac{W}{4}M \leq x_z < \dfrac{W}{4}M \\ -\dfrac{H}{2}M \leq y_z < \dfrac{H}{2}M \\ z_z = d \end{cases} \tag{1}$$

We denote the representation of P_z in the camera coordinate system as P_c. P_c is expressed with the rotation matrix $\mathbf{R}$ by equation (2).

$$P_c = \mathbf{R} P_z \tag{2}$$

We map the $P_c\left(x_c, y_c, z_c\right)$ in the camera coordinate system to corresponding $P_e\left(u_e, v_e\right)$ on the equirectangular image, using the azimuth angle θ and elevation angle ϕ as shown in Fig. 3. The values of each θ and ϕ are expressed by equations (3) and (4).

$$\begin{cases} \theta = \arccos \dfrac{x_c}{z_c} \\ \phi = \arccos \dfrac{y_c}{\sqrt{x_c^2 + z_c^2}} \end{cases} \tag{3}$$

$$\begin{cases} \theta = \left(\dfrac{u_e - W/2}{W}\right) 2\pi \\ \phi = \left(\dfrac{v_e - H/2}{H}\right) \pi \end{cases} \tag{4}$$

We generate the perspective-transformed image with dimensions $H \times W/2$ pixels based on the correspondence between P_e and P_z.

In this paper, we can generate a perspective-transformed image from an omnidirectional image using an equirectangular projection by providing the parameters $(\mathbf{R}, d, M)$.

B. Estimation of Camera Pose and Orientation Using Visual SLAM and Adjustment

In this study, we utilize OpenVSLAM [13] for the estimation of camera positions and orientations in an omnidirectional video. Among the outputs of OpenVSLAM, our proposed method uses the following three values. The index k represents the number of the keyframes inserted when there is a significant change in the field of view.

- The shooting positions of each keyframe p_{c_k} which is represented in the 3D reconstruction coordinate system (SLAM coordinate system)
- The rotation matrices $\mathbf{R_{ck}}$ which represents the transformation from the SLAM coordinate system to the camera coordinate system in each keyframe
- The 3D point cloud PC which is composed of image feature points represented in the SLAM coordinate system

However, the basis of the SLAM coordinate system is determined by the initial camera tilt in the first frame and the scale of the cordinate system and the real world is unknown. The rotation matrix $\mathbf{R}$ from the zenith coordinate system to the camera coordinate system and the distance $d = h_k$ in millimeters from the camera position to the ceiling surface are required for generating a ceiling image. We need to estimate and adjust the tilt and the scale of the SLAM coordinate system.

1) Adjustment of SLAM Cordinate System Tilt: We estimate the zenith direction in the first frame of the omnidirectional video and rotate the shooting positions p_{c_k}, rotation matrices $\mathbf{R_{ck}}$, and 3D point cloud PC. The zenith direction is estimated through two procedures. The first procedure is calcurating the zenith direction using a positional relationship of rectangle vertices on the ceiling surface captured in the first frame. The second procedure is estimation of the ceiling plane using the 3D point cloud PC. We adjust the tilt of the SLAM coordinate system based on its inclination.

Estimating Zenith Direction Using Rectangles in Image: First, we need to generate a perspective-transformed image from the first frame of the omnidirectional video. We do not consider the rotation of the camera coordinate system and set the parameter $\mathbf{R}$ as the identity matrix. Assuming a typical office ceiling height of 2800 millimeters and the camera position height of 1000 millimeters above the floor, we

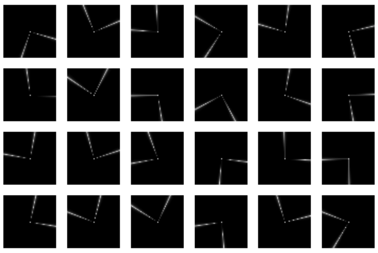

Fig. 4. Vertex filters

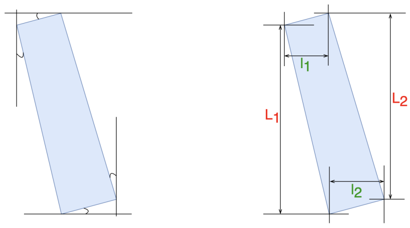

Fig. 5. Criteria of orientation transformation angles

set the parameter $d = 1800$ in millimeters. To ensure that the ceiling fixture of the rectangle is captured, the scale factor M is set as $M = 6000/H$ in millimeters per pixel to capture an area of approximately 6000 millimeters square in the resulting image. H represents the height of the input image in pixels.

We utilize OpenCV to detect the rectangular objects in the generated image and estimate the zenith direction using one of them. We apply the Prewitt filter [14] to the vicinity area of each vertex $v\,(= 0, 1, 2, 3)$ of the rectangle, and denote the resulting image as $I_v\,(x, y)$. To accurately capture the vertices of the rectangle, we make 3600 vertex filters $E_n\,(n = 1, 2, ..., 3600)$ with dimensions $H_e \times W_e$ pixels, as shown in Fig. 4. The scores representing how the pixel is the likelihood of a vertex are calculated in each pixel of $I_v\,(x, y)$ according to equation (5). We consider the pixel $(\hat{x}, \hat{y})$ with the highest score as the vertex, as stated in equation (6).

$$f\left(I\left(x,y\right),E_n\right)=\sum_{i=-\frac{W_e}{2}}^{\frac{W_e}{2}}\sum_{j=-\frac{H_e}{2}}^{\frac{H_e}{2}} I_v\left(x+i,y+j\right)E_n\left(i,j\right) \quad (5)$$

$$\hat{x},\hat{y},\hat{n}=\underset{x,y,n}{\text{argmax}} f\left(I_v\left(x,y\right),E_n\right) \quad (6)$$

We repeatedly determine the orientation transformation angles (Euler angles) that make the rectangular object a true rectangle, based on the positions of its vertices. The orientation transformation angles are calculated each time based on the following criteria as shown in Fig. 5.

- The average difference between the angles of the four vertices and their corresponding ideal angles.
- The ratio of the two corresponding edges on the left and right sides ($L_1/L2$).
- The ratio of the two corresponding edges on the top and bottom sides (l_1/l_2).

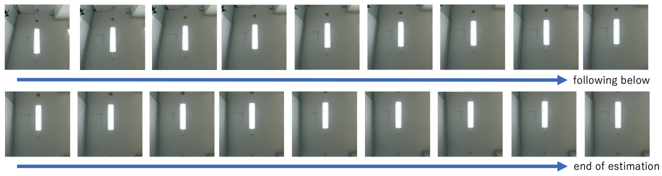

Fig. 6. Progress of the orientation transformation

This process continues until the difference in transformation angles becomes smaller than a predefined threshold or a certain number of iterations is reached. At that point, we consider the orientation as the zenith direction. Each calculated angle is multiplied by $-0.98^{-n} + 2$ to achieve convergence in the orientation transformation ($n(0, 1, ..., 34)$ represents the number of iterations). We use this factor to increase the transformation angles when the number of iterations is small. Fig. 6 shows the progress of the orientation transformation at each iteration.

We denote the rotation matrix from the initial orientation to the final orientation after the iterative process as $\mathbf{R_i}$. For further processing, we rotate the shooting positions p_{ck}, rotation matrices $\mathbf{R_{ck}}$, and 3D point cloud PC using $\mathbf{R_i}$ as follows:

$$p_{cik} = p_{ck}\mathbf{R_i}, \quad (7)$$

$$\mathbf{R_{cik}} = \mathbf{R_{ck}}\mathbf{R_i} \quad (8)$$

$$PC_i = PC\mathbf{R_i} \quad (9)$$

Estimating Zenith Direction Using Point Cloud: There are cases where it is difficult to detect the accurate rectangular vertices in variable lighting in image processing. Additionally, image distortion can also make it challenging to determine the precise zenith direction. Therefore, we estimate the zenith direction using the 3D point cloud PC from Visual SLAM. Ceiling fixtures, such as lighting elements, often show different brightness from the ceiling surface. These differences can be detected as feature points in the image. We can extract the ceiling plane from the 3D point cloud PC_i by using numerous feature points on the ceiling surface.

Before the extraction process, Radius Outlier Removal is applied to all points in the 3D point cloud PC_i to eliminate noise. The removal is based on the number of points within a sphere centered at each point. For the extraction process, we utilize Random Sample Consensus (RANSAC) [15]. RANSAC randomly selects initial samples, which may lead to the extraction of surfaces other than the ceiling plane. To address this issue, we apply RANSAC with a 3D plane model to a subset of the 3D point cloud PC_i, specifically the upper 1/4 portion representing the ceiling side. This approach helps ensure the accurate extraction of the ceiling plane.

The SLAM coordinate system is rotated based on the estimated ceiling plane so that its Z-axis becomes parallel to the normal of the ceiling plane. The rotation matrix for this transformation is denoted as $\mathbf{R_p}$. For further processing, we rotate the shooting positions p_{ci_k}, rotation matrices $\mathbf{R_{ci}}_k$, and 3D point cloud PC_i using $\mathbf{R_p}$ as follows:

$$p_{cp_k} = p_{ci_k}\mathbf{R_p} \tag{10}$$

$$\mathbf{R_{cp_k}} = \mathbf{R_{ci_k}}\mathbf{R_p} \tag{11}$$

$$PC_p = PC_i\mathbf{R_p} \tag{12}$$

2) Adjustment of SLAM Cordinate System Scale: The scale between the SLAM coordinate system and the real world is unknown. We estimate the scale based on the corresponding distances of the SLAM cordinate system and the real world. From the result of the estimated scale, shooting positions p_{cp_k} in the SLAM coordinate system are adjusted so that one unit in the coordinate system corresponds to 1 millimeter.

We denote the distance from the origin (the shooting position of the first frame) to the estimated ceiling plane in the VSLAM coordinate system as h_v. In the real world, the distance from the shooting position of the first frame to the ceiling plane is denoted as h_1 in millimeters. To calculate h_1, we need to generate a perspective-transformed image from the first frame of the omnidirectional video. For this process, the parameters are set as follows: $\mathbf{R} = \mathbf{R_{cp_k}}$, $d = 1800$ in millimeters, and $M = 6000/H$ in millimeters per pixel, similar to the settings in estimation of the zenith direction using image processing(section III-B1). We measure the length D_p in pixels of the ceiling fixture in the generated perspective-transformed image and the actual length D_r in millimeters and calculate h_1 from equation(13).

$$h_1 = d\frac{D_r}{D_pM} \tag{13}$$

The scale S is calculated as follows:

$$S = \frac{h_1}{h_v} \tag{14}$$

For the process of generating the ceiling image, the estimated shooting positions p_{cp_k} are normalized based on the scale S as follows:

$$p_{cs_k} = p_{cp_k} \times S \tag{15}$$

C. Generating Ceiling Images

For each keyframe, a zenith-directed perspective-transformed image called ”ceiling image” are generated to synthesize the panoramic ceiling image. The generation of ceiling images use the estimated shooting position $p_{cs_k}\,(x_{cs_k}, y_{cs_k}, z_{cs_k})$ and the rotation matrix $\mathbf{R_{cp_k}}$. The distance h_k from the shooting position to the ceiling surface is represented by equation(16).

$$h_k = h_1 - z_{cs_k} \tag{16}$$

We set the length per pixel in the resulting ceiling image as M_c. The zenith-directed perspective-transformed images are generated with parameters: $\mathbf{R} = \mathbf{R_{cp_k}}$, $d = h_k$, and $M = M_c$ in each keyframe. This process allows us to generate ceiling images with dimentions $H_c \times W_c$ pixels.

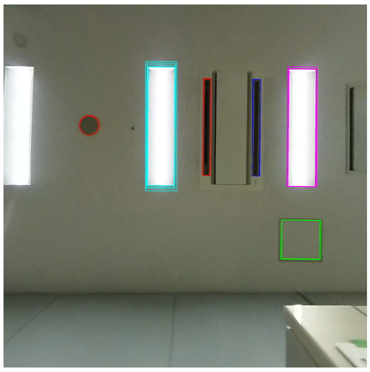

Fig. 7. Example of detected fixture image

D. Detecting Rectangular and Circular Ceiling Fixtures

For each ceiling image, we perform the detection of rectangular and circular ceiling fixtures to draw clear fixture outlines in the panoramic ceiling image. The ceiling images are converted to an 8-bit grayscale image using OpenCV. Then, the grayscale images are applied thresholding using 21 different threshold values ranging from 30 to 230 in increments of 10. We utilize OpenCV to extract contours and find the minimum bounding rectangles and minimum enclosing circles in each binary image obtained from thresholding. From the obtained rectangles and circles, extraction of fixture candidates is needed. To determine the fixtures from the candidates, we consider the object that is consistently identified across multiple thresholded binary images as a fixture. By using multiple threshold values, we enhance the robustness of the recognition process against variations in lighting conditions and color tones.

For the minimum bounding rectangles, we calcurate the area S_r, the perimeter length l_r, the length of the contour l_{rc}, the area inside the contour S_{rc}, the number of black pixels inside the rectangles N_{rb} and the number of white pixels inside the rectangles N_{rw}. We denote the area of the ceiling images as $S_{ci} = H_cW_c$. Any candidate that meets at least one of the following conditions is excluded, and the remaining ones are considered as rectangular fixture candidates.

- $0 \le S_r/S_{ci} \le 0.005$ or $0.6 \le S_r/S_{ci} \le 1.0$ (to remove noise).
- Rectangles that may be partially out of the image (to remove fixtures that are not fully captured).
- $l_{rc}/l_r \le 0.8$ or $1.2 \le l_{rc}/l_r$ (to remove non-rectangular objects).
- $S_{rc}/S_r \le 0.8$ (to remove non-rectangular objects).
- $\max(N_{rb}, N_{rw})/S_r \le 0.8$ (to remove rectangles with non-uniform inside).

For the minimum enclosing circles, we calcurate the area S_c, the perimeter length l_c, the length of the contour l_{cc}, the area inside the contour S_{cc}, the number of black pixels inside the rectangles N_{cb} and the number of white pixels inside the rectangles N_{cw}. Any candidate that meets at least one of the following conditions is excluded, and the remaining ones are considered as circular fixture candidates.

- $0 \leq S_c/S_{ci} \leq 0.001$ or $0.6 \leq S_c/S_{ci} \leq 1.0$ (to remove noise).
- Circles that may be partially out of the image (to remove fixtures that are not fully captured).
- $l_{cc}/l_c \leq 0.8$ or $1.2 \leq l_{cc}/l_c$ (to remove non-circular objects).
- $S_{cc}/S_c \leq 0.8$ (to remove non-circular objects).
- $\max(N_{cb}, N_{cw})/S_c \leq 0.8$ (to remove circles with non-uniform inside).

We determine if the fixture candidate is a fixture based on the distance between the centroids of fixture candidates in the multiple binary images. If the distance is within 50 millimeters, fixture candidates are considered as the same candidates and decided as fixtures. Fig. 7 shows an example of the detected fixture image.

E. Extract and Search the Clearest Ceiling Fixture Outlines

We want to draw clearest outline each of fixtures from a single ceiling image in the synthesized panoramic ceiling image. Firstly, we extract the identical ceiling fixtures from the detected fixtures across different ceiling images, since the ceiling images overlap. Next, we search for the ceiling image that captures the fixture outline most clearly from the images which show the identical one.

1) Extract Ceiling Fixtures Across the Ceiling Images: We extract the same ceiling fixtures from different ceiling images. The center coordinates of each ceiling image are (x_{cs_k}, y_{cs_k}) based on the shooting position $p_{cs_k}(x_{cs_k}, y_{cs_k}, z_{cs_k})$. The positions of detected ceiling fixtures are calculated based on the positions of the ceiling images. We consider the pairs of ceiling fixtures that meet the following conditions to be the identical ceiling fixture.

- The distance between the centroids of two ceiling fixtures is within 300 millimeters.
- The area ratio between the two ceiling fixtures is between 0.8 and 1.2.

If the centroid of one ceiling fixture is enclosed within the outline of another ceiling fixture, we compare the areas of them. The one with the smaller area is excluded from the set of extracted ceiling fixtures. This process ensures that the set includes only the detected ceiling fixtures that are fully captured. We need to exclude the detected ceiling fixtures that are partly captured as shown in Fig. 8.

2) Search for the Clearest Ceiling Fixture Outlines: We search for the ceiling image that captures the ceiling fixture outline most clearly from the ceiling images capturing the same ceiling fixture. The closer the ceiling fixture is to the center of the ceiling image, the less distortion it is affected by. Therefore, we consider that the ceiling fixture outline captured near the center of the ceiling image is clearer. Among the ceiling fixtures considered as the same one, we search for the ceiling image where the distance from the center of the image to the centroid of the ceiling fixture is the shortest. In other words, for each ceiling fixture, the image where the fixture is closer to the center is searched.

(a) (b)

Fig. 8. Angles of view that capture a fixture partly (a) and wholly (b)

F. Synthesize Panoramic Ceiling Image Prioritizing Fixture Outlines

We generate a panoramic ceiling image that shows the clear depiction of fixture outlines, based on the positions of each ceiling image and the searched ceiling fixture outlines. Firstly, the dimensions (width and height) in pixels of the resulting panoramic ceiling image is calculated. Next, we perform Voronoi partitioning on each pixel based on the center positions of the ceiling images. For each region generated by the partitioning, the pixel values are copied from the corresponding ceiling image. However, for the area of detected ceiling fixtures and their surroundings, the pixel values are copied from the corresponding searched ceiling images. Through this process, we can synthesize a panoramic ceiling image that shows clear fixture outlines.

IV. Experiment and Evaluation

We collected data for the proposed method and baseline in 2 indoor environments. We evaluated the synthesized panoramic ceiling image for positional and shape accuracy of the ceiling fixtures.

A. Baseline

We used Matterport, the commercial software developed by Matterport Inc., as the baseline for comparison. Matterport enables us to capture indoor spaces using 3D laser scanners, omnidirectional cameras, smartphones, and other devices. It requires fundamental knowledge of 3D reconstruction for accurate capture. In the experiment, we used the same omnidirectional camera as the proposed method. However, the input for Matterport is images, not videos.

We captured indoor spaces with continuity in capturing positions. While capturing, we needed to connect the omnidirectional camera to a smartphone app via Wi-Fi. After capturing the omnidirectional images, we uploaded them from the smartphone app to the cloud for processing. The outputs of the processing are walkthroughs of the measured indoor space, 3D models, panoramic images of the floor and ceiling surfaces, etc.

B. Collecting Data

We collected data in two different indoor environments. The first environment is a workspace area of approximately

(a) Environment 1

(b) Environment 2

Fig. 9. Experiment environments

TABLE I
DETAIL OF COLLECTED DATA

	Environment	Environment 1	Environment 2
Ours	Video length [seconds]	211	91
	Frame rate [fps]	29.97	
	Resolution [pix]	3840×1920	
	Collection time [minutes]	3.5	1.5
Matterport	Number of images	100	45
	Resolution [pix]	6720×3360	
	Collection time [minutes]	57	27

273 square meters, which contains multiple individual rooms, desks, monitors, chairs, and other furnishings, as shown in Fig. 9(a). The second environment is an open space of approximately 86 square meters, which is a computer room with aligned desks with computers on top, as shown in Fig. 9(b).

We collected data using an omnidirectional camera RICOH THETA Z1 for both the proposed method and the baseline. Table I provides details about the collected data. We directed one of the camera lenses towards the ceiling to ensure that one lens captured the ceiling surface independently when we shot the omnidirectional video for the proposed method. There might be differences in brightness and color tone between the lenses. We used a standalone tripod with a length of approximately 126 centimeters to ensure that the camera lenses were oriented horizontally while capturing the omnidirectional images for the baseline.

As ground truth data, we used a FARO FocusM70 3D laser scanner to obtain a 3D point cloud of the indoor space. The ceiling part was extracted from the acquired 3D point cloud to generate the ground truth panoramic ceiling image.

C. Evaluation Method

1) Positional Accuracy: We evaluate the positional accuracy of the panoramic ceiling image based on the positions of the four corners of rectangular light fixtures on the ceiling. The positions of the corners of rectangular light fixtures are marked manually. We perform rotation, scaling, and translation operations on the marked positions of each panoramic ceiling image. The accuracy of the positions are evaluated that minimize the average positional error with the ground truth data.

2) Shape Accuracy: We evaluate whether the fixtures in the panoramic ceiling image have the correct shape. We calcurate the interior angles of all rectangular light fixtures, then calculate the mean and standard deviation of the set of all angles. All interior angles of the rectangular light fixtures are 90 degrees. If the standard deviation is small, it indicates that the panoramic ceiling image reflects the shapes of the fixtures outlines accurately.

TABLE II
EVALUATION RESULTS

Environment	Environment 1		Environment 2	
Method	Ours	Baseline	Ours	Baseline
Positional Error [mm]	**53.09**	126.24	**14.47**	57.55
Angle Mean [deg]	90.00	90.00	90.00	90.00
Angle Standard Deviation [deg]	**0.95**	4.14	**0.68**	3.17

D. Results

Fig. 10 shows the panoramic ceiling images synthesized using the proposed method or the baseline. In the proposed method, we set $M_c = 2$ to achieve a scale of 2 millimeters per pixel. The image synthesized by the proposed method represents the indoor space more accurately, and the outlines of the fixtures are clearer compared to the baseline.

Table II provides the evaluation results. The proposed method demonstrated higher positional accuracy compared to the baseline. We used videos as input in the proposed method, while the baseline used high-resolution images. The use of continuous video likely contributed to its higher accuracy compared to the use of high-resolution images.

The standard deviation of the shape accuracy is smaller for the proposed method as shown in table II. This indicates that it accurately captures the shapes of the rectangular light fixtures in the proposed panoramic ceiling image. Synthesizing the fixture outlines from a single ceiling image while the synthesis of the panoramic ceiling image contributed to the accuracy of the shapes.

Fig. 11 shows the panoramic ceiling image with the seams between ceiling images indicated by red lines. The proposed method is achieved to capture the fixture outlines seamlessly in the panoramic ceiling image.

V. CONCLUSION

In this paper, we proposed a panoramic ceiling image synthesis method that prioritizes the fixture outlines using omnidirectional videos. In our proposed method, we first estimated the shooting positions and orientations for each keyframe using Visual SLAM and then adjusted the estimation results. Based on the shooting positions and orientations, the panoramic ceiling image was synthesized from the ceiling images. We prioritized the positions of the detected fixture outlines for easier identification of the types of ceiling fixtures.

In the experiment, we synthesized panoramic ceiling images in two different indoor environments. As a result, the panoramic ceiling images synthesized with our proposed method are more accurate than those with the baseline. Moreover, the workers at the construction site can create a ceiling plan in a shorter time using our method, because the time for shooting is very short.

We confirmed that it is possible to synthesize panoramic ceiling images prioritizing the fixture outlines. However, it

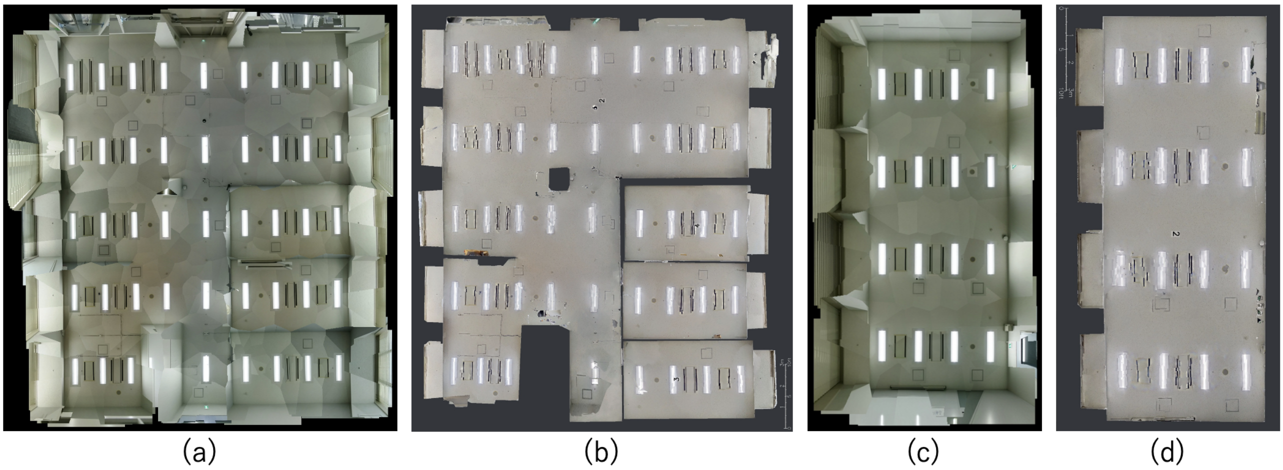

Fig. 10. (a) Ours, (b) Matterport in environment 1, (c) Ours, (d) Matterport in environment 2

Fig. 11. Seams between ceiling images

is necessary to establish recognition technology for ceiling fixtures. There are many kinds of ceiling fixtures and various types with different sizes and shapes. It is a challenge to establish a recognition technology that can handle unknown ceiling fixtures. We focused on the ceiling surface in this paper. We will apply this technology to other surfaces such as floors and walls.

Acknowledgment

This research is partially supported by JST, CREST Grant Number JPMJCR22M4, commissioned research (No.22609) by National Institute of Information and Communications Technology (NICT), JSPS KAKENHI Grant Number JP22K18422 and a project, JPNP23003, commissioned by the New Energy and Industrial Technology Development Organization (NEDO) in Japan.

References

[1] L. Doumbouya, G. Gao, and C. Guan, "Adoption of the Building Information Modeling (BIM) for Construction Project Effectiveness: The Review of BIM Benefits," *American Journal of Civil Engineering and Architecture*, vol. 4, no. 3, pp. 74–79, 2016.

[2] Y. C. Lin, Y. T. Hsu, H. T. Hu *et al.*, "BIM Model Management for BIM-Based Facility Management in Buildings," *Advances in Civil Engineering*, vol. 2022, 2022.

[3] J. V. Moreno, R. Machete, A. P. Falcão, A. B. Gonçalves, and R. Bento, "Dynamic Data Feeding into BIM for Facility Management: A Prototype Application to a University Building," *Buildings*, vol. 12, no. 5, p. 645, 2022.

[4] L. Pinti, R. Codinhoto, and S. Bonelli, "A Review of Building Information Modelling (BIM) for Facility Management (FM): Implementation in Public Organisations," *Applied Sciences*, vol. 12, no. 3, p. 1540, 2022.

[5] F. Morbidi, L. Devigne, C. Teodorescu, B. Fraud, É. Leblong, T. Carlson, M. Babel, G. Caron, S. Delmas, F. Pasteau *et al.*, "Assistive Robotic Technologies for Next-Generation Smart Wheelchairs," *IEEE Robotics and Automation magazine*, 2023.

[6] S. Kayukawa, K. Higuchi, S. Morishima, and K. Sakurada, "3DMovieMap: An Interactive Route Viewer for Multi-Level Buildings," in *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems*, 2023, pp. 1–11.

[7] C. Sun, C.-W. Hsiao, M. Sun, and H.-T. Chen, "HorizonNet: Learning Room Layout With 1D Representation and Pano Stretch Data Augmentation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 1047–1056.

[8] Z. Jiang, Z. Xiang, J. Xu, and M. Zhao, "LGT-Net: Indoor Panoramic Room Layout Estimation With Geometry-Aware Transformer Network," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 1654–1663.

[9] P. Fayyazsanavi, Z. Wan, W. Hutchcroft, I. Boyadzhiev, Y. Li, J. Kosecka, and S. B. Kang, "U2RLE: Uncertainty-Guided 2-Stage Room Layout Estimation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 3561–3569.

[10] M. Tabata, K. Kurata, and J. Tamamatsu, "Shape-Net: Room Layout Estimation From Panoramic Images Robust to Occlusion Using Knowledge Distillation With 3D Shapes As Additional Inputs," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 3551–3560.

[11] H. Wang, W. Hutchcroft, Y. Li, Z. Wan, I. Boyadzhiev, Y. Tian, and S. B. Kang, "PSMNet: Position-Aware Stereo Merging Network for Room Layout Estimation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 8616–8625.

[12] Matterport, https://matterport.com/.

[13] S. Sumikura, M. Shibuya, and K. Sakurada, "OpenVSLAM: A versatile visual SLAM framework," in *Proceedings of the 27th ACM International Conference on Multimedia*, 2019, pp. 2292–2295.

[14] J. M. Prewitt *et al.*, "Object enhancement and extraction," *Picture processing and Psychopictorics*, vol. 10, no. 1, pp. 15–19, 1970.

[15] M. A. Fischler and R. C. Bolles, "Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography," *Communications of the ACM*, vol. 24, no. 6, pp. 381–395, 1981.