Scale Optimization of Structure from Motion for Structured Light-based All-round 3D Measurement

Momoko Kawata[1], Hiroshi Higuchi[1], Sarthak Pathak[2], Atsushi Yamashita[1], and Hajime Asama[1]

Abstract— **In this paper, we propose a novel method for 3D measurement of large structures that have sparse features. The proposed method uses a structured-light approach based on a spherical camera and an omnidirectional ring laser. The spherical camera can capture omnidirectional images which enable it to view all sparse feature points existing in the target environment. The omnidirectional ring laser can enable dense 3D measurement of cross-sections of the environment via the structured light method. Structure from Motion (SfM) is used to measure the motion of the camera to integrate the laser cross sections to obtain a dense 3D model.**

However, the result of SfM does not contain real-world scale information. The novelty of this research lies in a new method to obtain the real-world scale. The real-world scale is determined by comparing a mesh generated from the resultant SfM point cloud and the integrated laser sections in terms of each scale.

In a simulated environment, the proposed method was found to be accurate up to 1 mm. It was also able to accurately measure the 3D shape of a real environment.

I. Introduction

Measuring 3D shape inside large structures is important for maintenance and infrastructures inspection [1], especially for automation of inspections. Various methods have been proposed to measure 3D structures for various target environments. In assembling large industrial products, such as trains and airplanes, the structure needs to be checked to ascertain whether the construction satisfies the requirements. Once the 3D surface structures are obtained, the producer can find if there is a mismatched assembling or wrongly attached parts. This is conducted partly by the human hand currently, which is time-consuming and requires skills. Therefore, the automation of the measurement of large structures is necessary.

Among many 3D measurement methods, the structured light method is known to be able to capture dense point clouds [2]. The structured light method projects a known laser light pattern onto the target surface and can obtain the 3D shape of the target object by viewing the shape of the light via a camera followed by triangulation. Line structured light method is one of the commonly used light patterns for such an approach. It projects a ring laser onto the target environment and take images of the target environment which contain the projected ring laser. There have been some research works on 3D measurement via line structured light for large structures [3] [4]. In order to capture the shape of the entire body of the target object via line structured light,

[1]Department of Precision Engineering, The University of Tokyo, Japan
[2]Department of Precision Mechanics, Chuo University, Japan
Contact: kawata@robot.t.u-tokyo.ac.jp

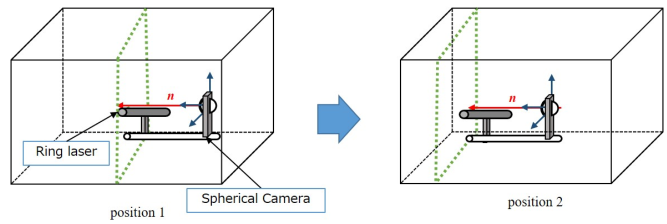


Fig. 1: The apparatus used in this research. It consists of two devices: an omnidirectional ring laser and spherical camera. By moving them around together, the images of the target environment that include the laser can be captured.

it is necessary to move the apparatus and integrate multiple cross-sections at different points, since one shot can only capture a limited part of the target object.

There are mainly two ways to integrate cross-sections. One is through the integration with known movement. In other words, cross-sections are integrated mechanically by moving around the equipment along predetermined poses. This is often used for quality inspection of relatively small products in controlled environments. However, this is unsuitable for large, uncontrolled environments.

The other integration method is through motion estimation of the apparatus via Structure from Motion (SfM) [5], which is the method applied in this research. SfM can estimate the camera's pose based on 2D feature points in the images. Figure 1 shows a diagram of the apparatus used in this research. It consists of two devices: an omnidirectional ring laser and spherical camera. While moving the equipment, spherical images that include the projected laser can be obtained [3] [4] [6]. SfM's motion estimation is done via the camera images along, and hence, does not include real-world scale information. Thus, it is necessary to determine the real-world scale when integrating cross-sections via SfM.

It is possible to solve this problem by directly using the points where the laser is projected in the motion estimation, as they can be measured directly by knowing the distance between the camera and the laser transmitter [3] [4] [6]. However, this requires the presence of textures in the environment around the projected laser points. Therefore, this approach cannot be applied to sparsely textured environments which don't have many feature points on their surface. In this paper, such environments typically contain only structural features such as 3D edges and corners of pillar or door.

Therefore, in this research, we focus on 3D measurement via line structured light method and SfM mainly for large, sparsely textured target objects and environments. The originality of our proposed method is that we can estimate the real-world scale without the need for the presence of dense textures in the regions illuminated by the laser. Instead, we choose to optimize the real-world scale via direct alignment of the laser cross-sections and the resultant SfM point cloud. Thus, we require no correspondence between measured 3D points from the light section method and the points of SfM for getting the real-world scale. The strength of our research is two points: One is that **it can work in sparsely textured environments**. The second is that **it can measure large structures and environments like entire rooms efficiently**.

II. Proposed System and its advantages

This section explains the proposed method for measuring 3D structure of the target environment and its design's concept.

A. Proposed System

As previously mentioned, in this research, the structured light method is used for getting dense 3D points. There are three characteristics of the proposed setup: **1. Integration via SfM**, **2. The use of a Spherical Camera 3. Real-world scale adjustment via global optimization by comparing two different generated point clouds**. In the following paragraphs, these three characteristics will be explained.

B. Design Concept of Proposed System

Integration via SfM: In a structured light approach, a laser is used at a disparity from a camera and the 3D distances of points illuminated by the laser are calculated by triangulation using the calibrated extrinsic parameters between the laser and the camera. In this research, an omnidirectional ring laser is used, which can illuminate a single cross-section of the environment. The apparatus is moved and SfM is used to integrate multiple cross-sections.

Cross-sections can be integrated by known mechanical movement. However, this requires large equipment and is only applicable for relatively small products like in a controlled, factory environment. It limits the range of applications of this method. Therefore, we use the camera images directly to obtain the motion of the camera via SfM to avoid the restriction. SfM can be conducted even in case of a limited number of textures in the environment.

Spherical Camera: In this paper, a spherical camera is used to capture images of the environment and the laser-illuminated regions. Spherical camera is a camera which can take omnidirectional images. It can obtain 360 degrees of view around the camera. There are three reasons to use the spherical camera based on the field of view.

For dealing with sparsely textured environments, it is necessary to have a wide field of view to include as many features as possible to calculate camera motion. Spherical camera can make this possible with its wide field of view.

Secondly, the width of the field of view can also solve a problem which exists in a normal perspective camera. In the case of a normal perspective camera, it is required to adjust the location of the laser source carefully depending on the target structure's size to include whole illuminated light. Otherwise, the laser may illuminate regions outside of the camera's field of view. This limits the baseline between the camera and the laser, limiting the accuracy and range of 3D measurement. In the case of a spherical camera, this limitation is overcome due to the fact that the laser illuminated regions will always be within sight of the camera.

Thirdly, the wide field of view can provide robust rotation estimation since it is easier to distinguish between translation and rotation in SfM.

Global Scale Adjustment: There have been some previous works which use SfM pose estimation to integrate structured light cross-sections. In the case of integration of cross-sections via SfM pose estimation, there is always a problem with real-world scale. The estimated pose by SfM does not contain real-world scale and only gives a relative scale.

Therefore, it is necessary to add real-world scale information by some other method in order to accurately integrate cross-sections.

Previous studies [3] [4] use the points illuminated by the laser to calculate this scale, as their absolute 3D distance is known via triangulation. However, these regions need to be matched as the camera moves and the laser prevents the use of feature points. Hence, block matching was employed in the region around each laser-illuminated point to find corresponding points in order to integrate the cross-sections.

However, this requires rich textures in the environment. Specifically, this requires a correspondence between the feature points used for SfM-based pose estimation and the 3D points reconstructed by the structured light approach. This is not suitable for many artificial environments which have a small number of features. Moreover, they rely on finding the absolute motion between consecutive frame pairs in the sequence. This can lead to the accumulation of errors.

Other studies [6] [7] integrate cross-sections light by reprojecting SfM 3D points to each frame and find corresponding laser illuminated points from these feature points. However, this also requires rich textures in the target environment.

In this paper, global scale optimization is adopted in order to make the measurement possible even **in sparsely textured environment**. The basic idea of this scale optimization is to minimize the difference between the resultant SfM point cloud and point cloud obtained from integrated cross-sections. This is done by converting the SfM point cloud to a triangular mesh and minimizing the distance between the 3D laser-illuminated point clouds and the mesh triangles. This is one of the originalities of our research.

By our method, the best scale can be found using global information about the whole structure, and not just consecutive frames. Moreover, we require no correspondence between measured 3D points from the light section method and the points used to calculate camera pose. The details are covered in the following sections.

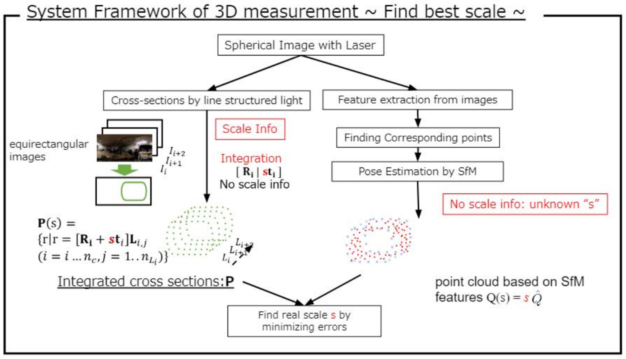


Fig. 2: Overall Framework: the framework can be divided into three parts. One is for obtaining 3D cross-sectional measurements by line structured light. The next is for estimating camera's poses $[\mathbf{R}|s\hat{\mathbf{t}}]$ by SfM and for obtaining the point cloud consisting of feature points ($s\hat{\mathbf{Q}}$) via SfM. Finally, after combining cross-sections by SfM pose estimation ($P(s)$), the real-world scale s is determined by minimizing the error between $P(s)$ and $s\hat{\mathbf{Q}}$ about the scale s.

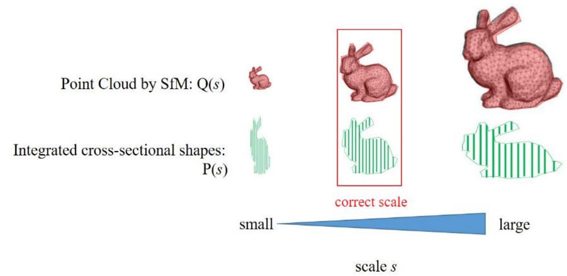


Fig. 3: Problem setting to determine the real-world scale in this research. How the shapes of the two point clouds (the SfM point cloud and the laser point cloud) change depending on the scale. When the scale s is correct one, their shapes should look the same. Using that concept, we can optimize the real-world scale s by comparing the similarity of the two point clouds.

III. Overview of Method

The apparatus consisting of the omnidirectional ring laser and the spherical camera is moved in the environment in order to make sure that the laser covers all areas as much as possible. Simultaneously, images are captured containing the laser cross section and other areas of the environment. This forms the input for our proposed method. The laser light in the image frames is used to extract 3D cross sections in the environment, and the rest of the information in the image frames is used to find the camera poses for integrating these cross sections to provide for dense 3D measurement. Figure 2 shows the framework of the proposed method.

The proposed method can be divided into three parts. One is for obtaining 3D cross-sectional measurements by line structured light. This part will be covered in Sec. III-A. The next involves SfM, which can provide the camera pose for each frame and generate a sparse point cloud of the surroundings. This part will be covered in Sec. III-B. The last part is the determination of the real-world scale. The real-world scale is determined by minimizing the distance of two different point clouds - the SfM point cloud used for camera pose estimation, and the laser point cloud - obtained from the light section method. Figure. 3 shows the main idea of the determination of the real-world scale. This part will be covered in Sec. IV.

A. Extraction of Cross-Sections

Figure 4 shows how the 3D laser cross section is projected on an omnidirectional image frame. In the process of obtaining the 3D cross-section, its projection on the omnidirectional image must be extracted. These extracted 2D points are called l. These 2D laser points can be converted into 3D points, $\mathbf{L}$, using the calibrated relationship between camera and ring laser plane. It should be noted that these

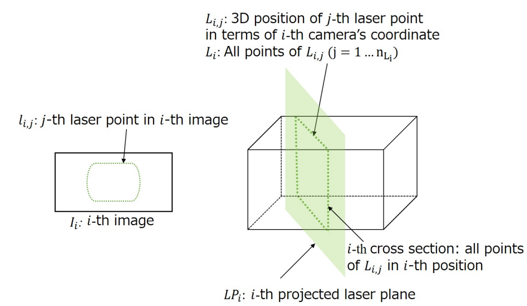


Fig. 4: The 3D laser cross section projected on the omnidirectional image captured from the spherical camera.

extracted 3D laser points $\mathbf{L}$ are present in the coordinate system of the image frame they were extracted from. Once the 3D cross sections are extracted in each image frame, they must be integrated and brought to the world coordinate system in order to obtain the dense point cloud of the environment. To integrate these independent cross-sections, it is necessary to find the relative camera poses of all image frames, with the correct scale information.

B. Structure from Motion

SfM is used for finding the camera poses for integration of the cross-sections. After finding corresponding points in all image frames, the camera's poses are calculated. AKAZE feature [8] are used in this. As output, the following are obtained. The first is the set of estimated camera poses for each image frame (i.e. for each laser cross section), $[\mathbf{R}|s\hat{\mathbf{t}}]$, where $\mathbf{R}$ is a rotation matrix and $\hat{\mathbf{t}}$ is a translation vector, both relative to the first image frame's frame of reference. The other is the 3D point cloud $s\hat{\mathbf{Q}}$ of the feature points used for SfM. Here, s, the scale of the translational distance of the camera poses is relative. Only the shape of the point cloud and the camera pose trajectory is accurate and the real-world scale and the absolute size need to be determined somehow.

C. Integration of Cross-Sections

The rotation matrix and translation vector of the i-th image frame pose can be written as Eq. (1), where $\|\mathbf{t}_1\| = 1$. $\|\hat{\mathbf{t}_i}\|$ is relative and normalized according to $\|\mathbf{t}_1\|$, and can be expressed as $\hat{\mathbf{t}_i} = \mathbf{t}/\|\mathbf{t}_1\|$. Since $\hat{\mathbf{t}_i}$ is scaled relative to $\|\mathbf{t}_1\|$, it can be determined uniquely.

$$[\mathbf{R}_i|s\hat{\mathbf{t}_i}] \quad (i = 1, \ldots, n). \tag{1}$$

$P(s)$ is the integrated result of all 3D cross-sections with scale s and can be expressed as Eq. (2).

$$P(s) = \{\mathbf{r} \mid \mathbf{r} = [\mathbf{R}_i|s\hat{\mathbf{t}_i}]\mathbf{L}_{i,j} \ \ (i = 1, \ldots, n, j = 1 \ldots m_i)\}, \tag{2}$$

where n is the position where the images are captured and m_i is the number of laser points in i-th image. $Q(s)$ is the SfM point cloud with scale s expressed as Eq. (3), where $\hat{\mathbf{Q}}$ is the normalized point cloud.

$$Q(s) = s\hat{\mathbf{Q}}. \tag{3}$$

As explained, s, the real-world scale has to be determined correctly in order to obtain the integrated cross-sections. The details are covered next in Sec. IV.

IV. Determination of real-world scale

A. Overview and Challenging Point of Scale Determination

As explained in Sec.III-B, it is necessary to determine the real-world scale in order to integrate cross-sections by line structured light. Figure 3 shows the problem setting when determining the real-world scale in the case of this paper. Even though both point cloud $Q(s)$ and $P(s)$ depend on the real-world scale s, they change differently. This is because in $Q(s)$, s is applied to 3D position of each point as Eq. (3) shows. $Q(s)$ retains its shape on increasing the scale s, but changes its size. On the other hand, in case of $P(s)$, the scale s is only applied to the direction of translation vector between consecutive frames $\hat{\mathbf{t}}_i$ as Eq. (2) shows. Thus, while the laser cross sections do not change their size and shape, the resultant point cloud on integration shows a different shape depending on how the camera was moved.

If the scale s is incorrect, the two point clouds have a very different shape. Only the correct real-world scale leads to their shapes aligning. In this research, we use this novel idea to find the correct real-world scale, by comparing the similarity of two point clouds.

The following three points summarize the challenging points in finding the real-world scale.

1) The density of both point clouds differs a lot
2) Two point clouds $P(s)$ and $Q(s)$ change differently depending on the scale s

These challenging points 1) and 2) cause problems in using the typical point cloud registration approach. 3D feature point descriptors [9] and ICP [10] cannot be used due to the changing shapes over varying scales.

Therefore, it is necessary to optimize s by constructing a new method to evaluate point cloud similarity of $Q(s)$ and $P(s)$ over changing shapes and scales.

We take advantage of the fact that $Q(s)$ and $P(s)$ share the same frame of reference - that of the first image frame, and develop a novel method to estimate the real-world scale.

The core idea is to compare the two point cloud by the following: **Either one of the point clouds is converted into a mesh firstly. Then, the mesh and the other point cloud are compared over different scales**.

This cannot be implemented in regular point cloud registration due to different frames of reference.

B. Definition and Minimization of Intra-Point Cloud Distance

In order to compare the two point clouds $P(s)$ and $Q(s)$, it is necessary to define a distance metric between them. In this subsection, the definition of the distance is explained.

As explained in Sec. IV-A, two point clouds, $P(s)$ and $Q(s)$, cannot be compared easily due to the difference in density and changeability of the shape of $P(s)$'s depending on scale s.

In this paper, the distance is calculated as the distance between mesh $M(s)$, which is generated from $Q(s)$ by Poisson Reconstruction [11], and $P(s)$. The point cloud obtained from SfM is chosen for meshing as it is easier to scale the mesh size due to its fixed shape.

The brief overview is as follows:

1) Find the ray vector $\mathbf{r} \in P(s)$ which crosses the triangle $M_j \in M(s)$.
2) Calculate the distance d_j for mesh M_j and Calculate $D(s)$ by summing all d_i.
3) Find the scale s which minimizes $D(s)$.

In the following sections, the details are explained.

Firstly, we explain the method to find the ray $\mathbf{r} \in P(s)$ which crosses each triangle $M_j \in M$. Next, the definition of distance $D(s)$ is explained.

C. Find Rays Crossing Mesh

First, we define the set of ray vectors, $\mathbb{R}_j$, which cross M_j as follows.

$$\mathbb{R}_j = \{\mathbf{r} \mid \mathbf{r} \in P(s) \wedge \ \mathbf{r} \text{ crosses } M_j\}. \tag{4}$$

Later in this section, it is explained how $\mathbf{r} \in \mathbb{R}_j$ is estimated. Figure 5 shows the definition of distance d_j between ray vector $\mathbf{r} \in \mathbb{R}_j$, which crosses the mesh triangle M_j.

By summing up d_j, $D(s)$ can be obtained as shown below in Eq. (5).

$$D(s) = \sum_{j}^{n_M} d_j^2, \tag{5}$$

where n_M is the number of triangles in the mesh. In the end, s_b which gives the smallest $D(s)$ is selected as the best estimate of the real world scale scale s.

The generated mesh $M(s)$ has triangles M_j which consist of the points in the SfM point cloud with a scale variable s. Each M_j consists of three corner points, $\mathbf{a}_j$, $\mathbf{b}_j$, $\mathbf{c}_j$.

Any arbitary $\mathbf{r} \in P(s)$ can be expressed using these three vectors as shown in Eq. (6) with real numbers α, β, γ,

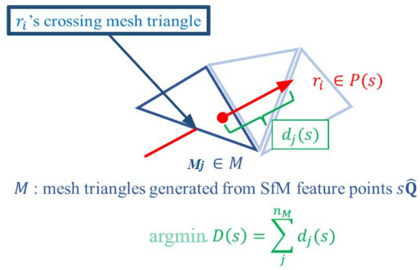


Fig. 5: Calculation of the distance between two point clouds. The SfM point cloud is converted to a mesh and its distance to the other point cloud is calculated.

since $\mathbf{a}_j$, $\mathbf{b}_j$, $\mathbf{c}_j$ are linearly independent. Here, $\mathbf{a}_j$, $\mathbf{b}_j$, $\mathbf{c}_j$ have scale $s = 1$. Therefore, when considering scale s, the equation for $\mathbf{r}$ is expressed as follows.

$$\mathbf{r} = s \begin{bmatrix} \mathbf{a}_j & \mathbf{b}_j & \mathbf{c}_j \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \\ \gamma \end{bmatrix}, \tag{6}$$

$$\begin{bmatrix} \alpha \\ \beta \\ \gamma \end{bmatrix} = \frac{1}{s} \begin{bmatrix} \mathbf{a}_j & \mathbf{b}_j & \mathbf{c}_j \end{bmatrix}^{-1} \mathbf{r}. \tag{7}$$

The condition that the $\mathbf{r} \in \mathbb{R}_j$, which satisfies Eq. (4), is shown below in Eq. (8).

$$\alpha \geq 0,\ \beta \geq 0,\ \gamma \geq 0. \tag{8}$$

D. Definition and Minimization of the Distance

Next, the definition of the distance between two point clouds, $Q(s)$ and $P(s)$, is described. We can define a distance d_j for a mesh M_j as Eq. (11) with the following equations.

$$D_j = \{d \mid d = f_s(\mathbf{r}, M_j), \mathbf{r} \in \mathbf{R}_j\}, \tag{9}$$

$$f_s(\mathbf{r}, M_j) = \|\alpha\mathbf{a}_j + \beta\mathbf{b}_j + \gamma\mathbf{c}_j - \frac{\alpha\mathbf{a}_j + \beta\mathbf{b}_j + \gamma\mathbf{c}_j}{\sqrt{\alpha^2 + \beta^2 + \gamma^2}}\|, \tag{10}$$

As a side note, Eq.(10) depends on the constant variable, scale s, according to Eq.(6).

$$d_j = \text{mean}(D_j). \tag{11}$$

The distance between mesh M_j and $\mathbf{r} \in \mathbf{R}_j$ is defined as the mean of D_j as shown in Eq. (11), since there can be multiple rays $\mathbf{r}$ which are crossing M_j, and the number of such $\mathbf{r}$ is different depending on M_j. For all $j =, \ldots, n_M$, $D(s)$ can be defined as the sum of d_j, where n_M is the number of mesh triangles. In practice, there are some mesh triangles which do not have any crossing rays $\mathbf{r}$, therefore, $D(s)$ can be described as Eq. (12).

$$D(s) = \frac{\sum_{j=1}^{n_M} x_j {d_j}^2}{\sum_{j=1}^{n_M} x_j}, \tag{12}$$

where x_j is expressed as follows.

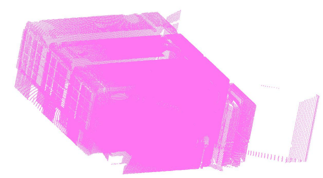


Fig. 6: Structure of classroom model in Blender

$$x_j = \begin{cases} 1 & (M_j \text{ has crossing ray}) \\ 0 & (M_j \text{does not have crossing ray}) \end{cases}. \tag{13}$$

Finally, s_b can be obtained as the best scale as shown in Eq. (14).

$$s_\text{b} = \arg\min D(s). \tag{14}$$

E. Scale Estimation by Distance Minimization

Before finding the bast scale s_b, it is necessary to determine the rough scale in order to search more efficiently.

The initial scale s_init is determined by Eq. (15). This equation uses the average norm of 1-*st* cam's cross-section points $\mathbf{L}_1$ and the normalized SfM point cloud $\hat{\mathbf{Q}}$ as Eq. (17) shows.

$$s_\text{init} = \frac{OD(\text{mean}(\mathbf{L_1}))}{OD(\text{mean}(\hat{\mathbf{Q}}))}. \tag{15}$$

Here,

$$OD(x) = 10^{i-1}, \tag{16}$$

where i is the number of digits of x, and

$$\text{mean}(\mathbf{x}) = \|\frac{\sum_{i=1}^{n_\mathbf{x}} \mathbf{x}_i}{n_\mathbf{x}}\|, \tag{17}$$

where, $\mathbf{x}$ is $n_\mathbf{x} \times 3$ is the size array.

Algorithm 1 shows the procedure to calculate the smallest D and the best scale s_b. After the initial scale s_init via Eq. (15) is found, the distances depending on the scale s can be computed. Starting from the s_init, the distances D are computed with different scales while updating the minimum D, D_min. The step size of the scales updates in every iteration and the estimation is stopped when the step is small enough - 0.001.

V. Experiments

Two kinds of experiments were conducted to verify the proposed method. One is an experiment in a virtual environment. The other was the real room, which the calibration's impact needs to be considered.

A. Virtual Environment

A classroom model was used as the target environment in Blender [12], which can render realistic images. Inside the model, the equipment was moved in a circle and the 3D structure was measured.

Data: $M(s)$, $P(s)$
Result: Best scale s_b, minimum D_{min}
initialization;
$s \leftarrow s_{init}$;
$s_{max} \leftarrow 10 s_{init}$;
$s_{min} \leftarrow s_{init}/10$;
$g \leftarrow s_{max}/10$;
$D_{min} \leftarrow \infty$;
$s_b \leftarrow 0$;
while $g \geq 0.001$ **do**
  for $s = s_{min};\ s < s_{max};\ s = s + g$ **do**
    $D \leftarrow$ calcD(s);
    if $D < D_{min}$ **then**
      $D_{min} \leftarrow D$;
      $s_b = s$;
    end
  end
  $s_{min} \leftarrow s_b - g$;
  $s_{max} = s_b + g$;
  $g \leftarrow g/10$;
end

Algorithm 1: Scale Estimation by Distance Minimization

Setup: Figure 6 shows an outer view of the environment. The rough dimensions of the room were 3[m]×6[m]×8[m]. The camera was rotated in the room at the center. The radius of the rotation was 0.1 m and the height was 1.25 m from the floor.

In this virtual environment, the spherical images and depth images were recorded. Spherical images were used for pose estimation by SfM. Depth images show the depth information on each location in an image.

The depth image was used for obtaining 3D laser cross-sections. This was done because it was not possible to simulate laser illumination in Blender. In this experiment, cross-sections were obtained by assuming that a plane at a fixed distance from the camera. The laser points at each fixed plane were calculated from the depth images. These formed the 3D laser cross section.

Figure 7 shows one virtual cross section.

In order to estimate poses via SfM, **openMVG** was used since it could handle spherical images [13]. The resolution of each spherical image was 1000 × 500 and feature descriptor used for SfM was AKAZE [8]. Incremental SfM [14] was conducted.

Results and Evaluation: Figure 8(a) shows the result of 3D measurement of room by the proposed method. Figure 8(b) shows the integrated cross-sections with ground truth camera poses. Ground truth poses were obtained directly from Blender. Figure 9 shows the a point cloud generated by SfM with scale 1 and ground truth integrated cross-sections. This shows us that the initial scale of SfM is far from the true scale.

Figure 10 shows the plot of $D(s)$ while calculating of the best scale by the proposed method. The smallest $D(s)$ for

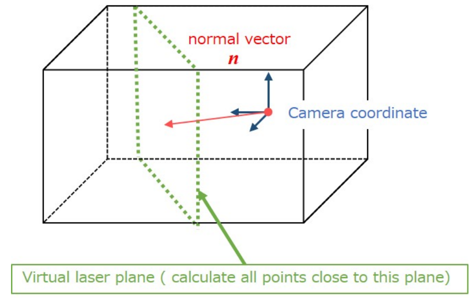
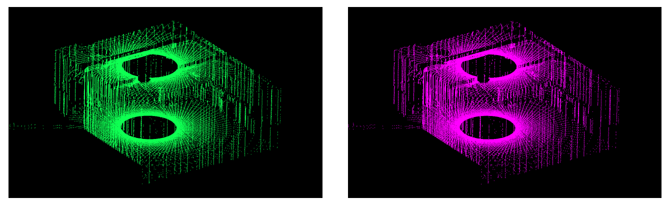


Fig. 7: Virtual Laser Plane: Since it was not possible to simulate laser illumination in Blender, a virtual cross section was used. The cross section was obtained by calculating all the points which were on a plane at a fixed distance from each camera.



(a) proposed method (b) ground truth

Fig. 8: Integrated cross-sections (a) the result by the proposed method (b) the ground truth result by integrating each cross section with the true camera poses. This shows that the proposed method can generate an accurate 3D shape.

the correct scale can be seen. The estimated best scale in this experiment was 0.09. The proposed method was successfully able to find the best real-world scale. When comparing the result of the integrated cross-sections of the proposed method and ground truth, the average 3D measurement per-point error was found to be 0.0276 mm per point.

B. Real Environment

Setup: Next, an experiment with the real environment was conducted. A classroom with dimensions of about 7[m] × 8[m] × 3[m] was used as the target environment. Figure 11(a) shows the used equipment. On top of the rotatable panhead as shows Fig. 11(b), the ring laser and spherical camera were mounted. By rotating the panhead, the images with laser illuminated cross sections were captured. A commercial spherical camera, RICOH THETA Z1, was used in this experiment. The resolution was 3840 × 1920. A ring laser was used, which could generate green laser light in all directions. The power of the laser was 30mW and the current was 530mA. In this experiment, two different images in each location were captured in order to extract the laser cross sections more precisely for the line structured method. The panhead was moved in steps of 10 degrees.

Result and Evaluation: Figures 12(a) and 12(b) shows the captured image for one location. Figure 12(a) was used for pose estimation by SfM and Fig. 12(b) was used for laser cross section extraction for the line structured light method. Figure 13 shows the 3D measurement result by the proposed

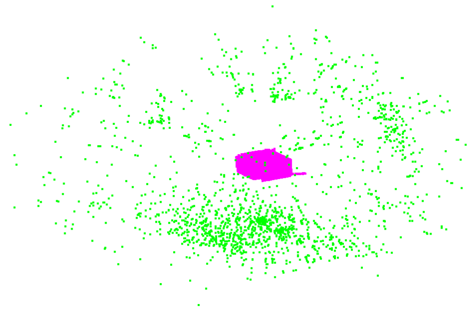


Fig. 9: The integrated cross sections of room with ground truth (pink) and SfM point cloud with scale 1 (green). It can be seen that the initial scale of SfM was far from the true scale.

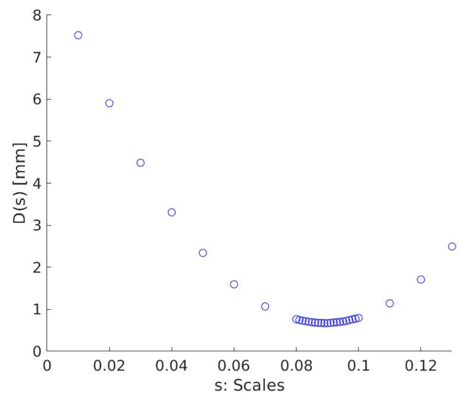


Fig. 10: Virtual environment: Plot the distance D vs. scale s

method. The estimated real-world scale was 582.63 for this experiment.

Figure 14 is the plot of distance D vs. scale s. The real-world scale could be determined successfully. In this experiment, accuracy was evaluated by measuring the flatness of each wall, the ceiling, and the floor. We manually selected the flat regions and evaluated the planarity.

Tables I, II, III and IV show the flatness of four planes, ceiling, floor, two walls. Figures 15(a), 15(b), 15(c) and 15(d) show how far each point in four planes, ceiling, floor and two walls was from the fitted plane.

TABLE I: Flatness of ceiling

number of points	46011
mean error [mm]	22.23

TABLE II: Flatness of floor

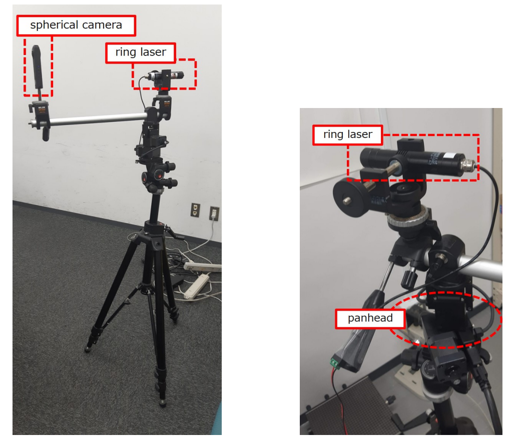
number of points	32020
mean error [mm]	31.47

TABLE III: Flatness of wall 1

number of points	2810
mean error [mm]	65.71

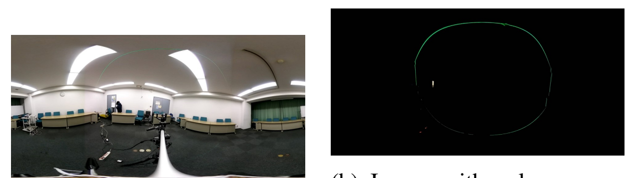
TABLE IV: Flatness of wall 2

number of points	2267
mean error [mm]	143.52



(a) Experimental setup (b) Panhead

Fig. 11: Captured images



(a) Captured image for SfM (b) Image with a laser cross section

Fig. 12: Captured images: normal images were used from SfM, and the other image with the laser light was were used for laser cross section extraction.

C. Discussion

Through the experiment with the virtual environment above, it was shown that the proposed method obtained the real-world scale. 3D measurement via the proposed method was successfully achieved and an accurate result was obtained. The error per point was less than 1mm.

In the experiment in a real environment, it was shown that the shape of the room could be measured properly by the proposed method. This can be noted from the flatness values reported.

There was a drop in accuracy in reconstructing the walls due to problems in the extraction of cross-section via line structured light method. The main difference between the Blender environment and the real environment was the line structured method. In Blender, virtual cross-sections were used instead of obtaining cross sections by line structured light method. In this paper, the extraction of points of line structured light from the spherical image was implemented by treating the image as a regular perspective camera image. The noise of points of the wall came from not considering the distortion of the spherical images. In other words, when extracting laser points from 2D image, the extraction method in the perspective image cannot be used as it is for spherical image. Another reason for the drop in accuracy while measuring walls could be a fault in laser plane to camera calibration.

In this paper, the core concept was the optimization

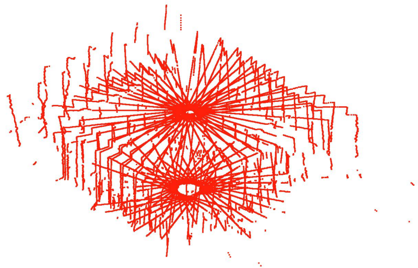


Fig. 13: Estimated result

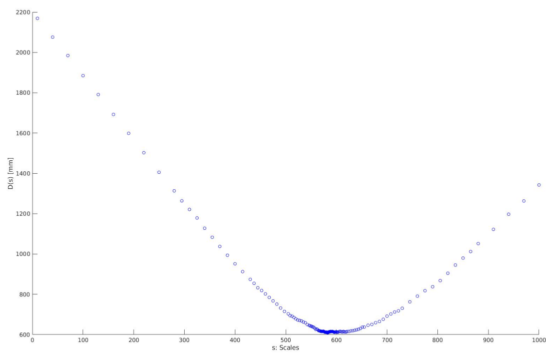


Fig. 14: Real environment: plot of distance D vs. scale s

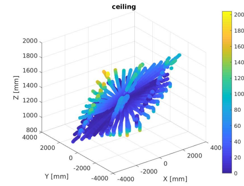
of the real-world scale using the global structure of the environment. Due to errors existing in laser plane calibration and laser cross section extraction in spherical images, the walls were measured with lower accuracy as compared to the ceiling and the floor. A suitable line structured light method specific to spherical cameras will be considered for future work.

VI. Conclusion

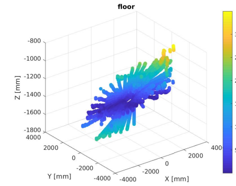
In this paper, we proposed a novel method for 3D measurement in a less textured environment with a low number of feature points. It was conducted via line structured light method based on scale optimization of SfM. textureless environment meant that it had only structural features such as edges of objects as feature points and didn't have high-frequency textures essential to feature point extraction. This was achieved by global scale optimization to determine the real-world scale of SfM. By comparing the point cloud of SfM and integrated laser cross-section point cloud, the real-world scale was obtained by minimizing the distance between two point clouds. The fact that this did not require the presence of high-frequency textures and relied only on a low number of feature points in the environment made this different from other related work.

Through experiments in both, virtual and real environments, the validity of the proposed method was shown. It was able to measure large structures and environments like entire rooms efficiently.

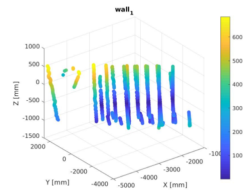
It achieved less than 1mm accuracy in the virtual environment. In the real environment, it successfully measured the 3D shape of the target environment.



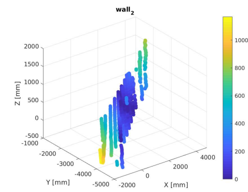
(a) Flatness profile of ceiling



(b) Flatness profile of floor



(c) Flatness profile of wall 1



(d) Flatness profile of wall 2.

Fig. 15: Flatness of the ceiling, floor and walls. The error bar shows the distance from fitted plane.

References

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

[2] J. Geng, "Structured-light 3d surface imaging: A tutorial," *Advances in Optics and Photonics*, vol. 3, no. 2, pp. 128–160, 2011.

[3] A. Yamashita, K. Matsui, R. Kawanishi, T. Kaneko, T. Murakami, H. Omori, T. Nakamura, and H. Asama, "Self-localization and 3-d model construction of pipe by earthworm robot equipped with omni-directional rangefinder," in *Proceedings of the 2011 IEEE International Conference on Robotics and Biomimetics (ROBIO2011)*, Phuket Island, Thailand, 2011, pp. 1017–1023.

[4] H. Higuchi, H. Fujii, A. Taniguchi, M. Watanabe, A. Yamashita, and H. Asama, "3d measurement of large structure by multiple cameras and a ring laser," *Journal of Robotics and Mechatronics*, vol. 31, no. 2, pp. 251–262, April 2019.

[5] 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.

[6] A. Duda, J. Schwendner, and C. Gaudig, "Srsl: Monocular self-referenced line structured light," in *Proceedings of the 2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS2015)*, Hamburg, Germany, 2015, pp. 717–722.

[7] B. Zheng, T. Oishi, and K. Ikeuchi, "Rail sensor: A mobile lidar system for 3d archiving the bas-reliefs in angkor wat," *IPSJ Transactions on Computer Vision and Applications 7*, pp. 59–63, 2015.

[8] P. F. Alcantarilla, J. Nuevo, and A. Bartoli, "Fast explicit diffusion for accelerated features in nonlinear scale spaces," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 7, pp. 1281–1298, 2011.

[9] H. Kim and A. Hilton, "Evaluation of 3d feature descriptors for multi-modal data registration," in *2013 International Conference on 3D Vision - 3DV 2013*, 2013, pp. 119–126.

[10] P. Besl and N. D. McKay, "A method for registration of 3-d shapes," pp. 239–256, 1992.

[11] M. Kazhdan, M. Bolitho, and H. Hoppe, "Poisson surface reconstruction," in *Proceedings of the 4th Eurographics Symposium on Geometry Processing*, vol. 7, 2006.

[12] Blender. [Online]. Available: https://www.blender.org/

[13] P. Moulon, P. Monasse, and R. Marlet, "Openmvg: Open multiple view geometry," in *Proceedings of International Workshop on Reproducible Research in Pattern Recognition*, 2016, pp. 60–74.

[14] ——, "Adaptive structure from motion with a contrario model estimation," in *Proceedings of Asian Conference on Computer Vision (ACCV2012)*, 2012, pp. 257–270.