

Acceleration of 3D Measurement of Large Structures with Ring Laser and Camera via FFT-based Template Matching

Momoko Kawata, Hiroshi Higuchi, Hiromitsu Fujii, Atsushi Taniguchi,
Masahiro Watanabe, Atsushi Yamashita and Hajime Asama

Abstract—For quality inspection and maintenance in the manufacturing industry, it is extremely important to accurately measure the 3D shape of various structures in an efficient manner. In this paper, we propose an accelerated light-section method to accurately measure the 3D shape of large structures. The light-section method is a popular method to measure the 3D shapes accurately, which uses a laser and a camera that can observe the shape of the laser cross-sections. This method consists of two important steps: first, images that include the laser cross-section are captured. Next, the cross-sections are integrated by calculating the transformations between these images. For want of high accuracy, area-based matching methods, such as template matching, which match pixels in an area one-by-one, are used to calculate these transformations. Hence, the processing time is too high and speed/efficiency is quite important for real-life 3D measurement applications. Therefore, we focus on developing a faster template matching method. In this process, the presence of the laser cross-section on the image texture can induce errors in the integration and lower accuracy. Thus, it is required to mask the laser regions inside each image. To solve this problem and make the calculation faster, a masked Fast Fourier Transform (FFT) based template matching is proposed. Via experimental evaluation, we show that the 3D measurement process can be made almost 1.6 times faster, with similar accuracy as compared to previous methods.

I. INTRODUCTION

Measurement of 3D shape is an important technology in the manufacturing industry, for the purpose of the quality inspection and maintenance. Accurate measurement of large structures such as escalators and railway vehicles is an active area of research and various methods have been proposed for this, so far. Such technology can make the currently manual process of measurement more precise and efficient [1]. Methods to measure large structures can be broadly divided into two categories: time-of-flight (ToF) methods and triangulation methods [2] [3].

One typical ToF method for 3D measurement involves SLAM by using Laser Range Finders (LRF) [4]. In this method, the distance from the LRF to the target object is observed via the time-of-flight of the reflected laser beams. Such methods can be applied for generating maps of indoor environments. However, it is not suitable for accurate measurement for the manufacturing industry because the number of laser beams is fixed. Thus, further the target object, sparser

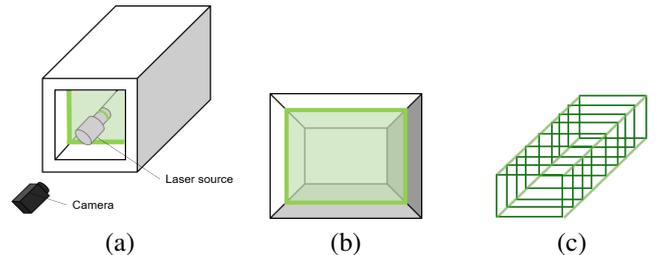


Fig. 1. Overview of 3D measurement of large structures by the light-section method. (a) shows the capture of the image including the laser inside the structure. (b) shows an example of an image captured. The green light from the laser is captured in the image. The 3D position of the area irradiated by the laser can be computed by triangulation. The cross-sectional shape can be obtained by observing the laser. (c) shows the output of the integration of multiple such cross-sectional shapes.

the measurement points. Therefore, they are not the best for the inspection of industrial products, where accuracy is required. In addition, in order to measure large structures, it is necessary to acquire data from different positions to capture the whole shape. After point clouds are generated based on the measured points, they are usually integrated by a process known as Iterative Closest Point (ICP). However, ICP requires unevenness in the surface of the structure to work properly. Thus, it is difficult to apply such methods to measure large structures that lack 3D surface features.

As for the other type of methods based on triangulation, the popular light-section method is able to measure the cross-sectional shape of the target object with high precision, and in a dense manner. Therefore, it is suitable to be used for inspection in the manufacturing industry. For measuring the shape of the target object by the light-section method, it is necessary to integrate the different cross-sectional shapes. Usually, the light-section method is performed by mechanically constraining the motion of the sensor and using the known motion for integration, which is to measure the large structures because of the limitation of the motion. In contrast, there are methods which directly use the texture information present inside the images of the target environment [5] [6]. With this idea, it is possible to measure the 3D shape of large structures more efficiently as we can obtain both, the cross-sectional shape and texture information for the integration, at the same time. Due to its usefulness as a practical measurement method of large structures, we focus on this type of light-section method.

Figure 1 provides a brief explanation of 3D measurement by the light-section method using texture information as

M. Kawata, H. Higuchi, A. Yamashita and H. Asama are with the Department of Precision Engineering, The University of Tokyo, Japan. kawata@robot.t.u-tokyo.ac.jp

H. Fujii is with Chiba Institute of Technology, Japan.

A. Taniguchi and M. Watanabe are with Hitachi, Ltd., Japan.

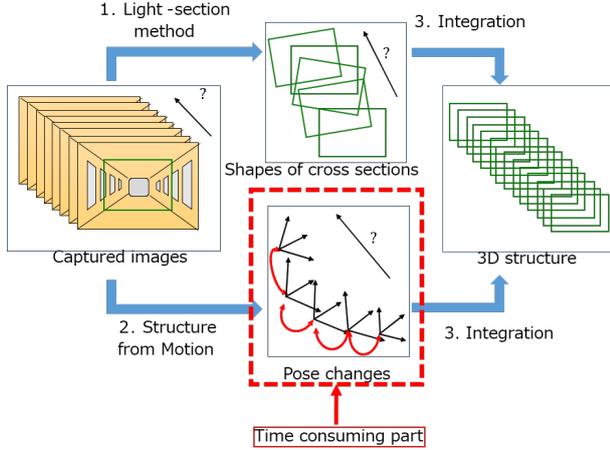


Fig. 2. Overview of 3D shape measurement by light-section method. After capturing images that include the laser irradiated area, 1. Obtain cross-sectional shape by observing the laser irradiated area. 2. Estimate the pose of the camera by Structure from Motion (SfM), using the texture information in the image. 3. Integrate all cross-sectional shapes using the estimated poses. The pose estimation step, which involves finding corresponding points in different images, takes up most of the calculation time. Therefore, in this paper, we mainly focus on accelerating this part.

done by [5]. However, the method explained in [5] is not practical for real measurement applications as the processing time is simply too long. Particularly, finding corresponding points between images for integration takes up most of the total calculation time. Therefore, in this paper, we propose a method for faster 3D shape measurement by the light-section method. We focus on the accelerating the process of finding corresponding points. Specifically, we use a Fast Fourier Transform (FFT) based similarity calculation to make it faster.

II. OVERVIEW OF METHOD

Figure 2 shows the outline of the method. First, images that include the target objects irradiated with a laser are captured at different positions. The 3D cross-sectional shapes can be obtained by observing the irradiation pattern. The camera positions and orientations of all the images are estimated using Structure from Motion (SfM). Finally, different cross-sectional shapes are integrated to form the whole 3D shape using the estimated camera positions and orientations. In the following subsections, this process will be explained.

A. 3D Cross-Sectional Shape Measurement

This section explains how to obtain a cross-sectional shape by the light-section method. We capture a picture that includes the target object which is irradiated with a laser, as shown in Fig. 1. By the idea of triangulation, we can calculate the 3D coordinate of each point of the laser projection in the camera's coordinate system. Figure 3 describes briefly how to calculate the 3D position of a laser projection point.

In order to measure the cross-sectional shape from each image, it is first necessary to extract the laser-irradiated cross-section inside each image. A green filter with a very narrow

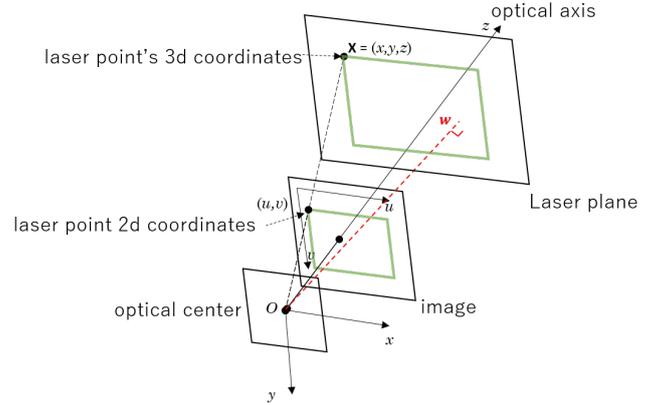


Fig. 3. Relationship between camera coordinates and the 3D position of the laser projection. By assuming the laser projection to be planar, we can get its 3D position in the camera coordinate system via triangulation.

wavelength is used to filter out all green light other than the laser. This results in the green channel of each image to contain only the laser-irradiated cross-section. Meanwhile, the red and blue channels are unaffected, providing texture information for the integration process.

Within the green channel of the image, we find the 2D “laser projection points” that make up the laser-irradiated cross-section by searching along lines drawn through the center of the image in an angular fashion. At each angle, the line from the center is searched and the point of maximum luminance in the green channel is chosen. Thus, the 2D laser projection points inside the image can be extracted from the image. How the laser projection points are captured in the image can be seen in Fig. 4.

Next, these 2D laser projection points are projected to 3D coordinates. The 3D coordinates are found by using the camera's intrinsic parameters matrix \mathbf{A} , which can be obtained by camera calibration. Figure 3 shows how these parameters are used to project these points to 3D. \mathbf{A} can be expressed as shown in Eq. (1), where (f_x, f_y) is the camera focal length and (c_x, c_y) is the center of the image.

$$\mathbf{A} = \begin{pmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{pmatrix}. \quad (1)$$

Thus, the 3D ray vector \mathbf{r} of a laser projection point on an image $(u, v)^T$ is defined as follows.

$$\mathbf{r} \equiv \mathbf{A}^{-1} \begin{pmatrix} u \\ v \\ 1 \end{pmatrix}. \quad (2)$$

The 3D coordinate $\mathbf{X} = (x, y, z)^T$ in the camera coordinate system can be expressed as Eq. (3), scaled by a variable a .

$$\mathbf{X} = a\mathbf{r}. \quad (3)$$

From Eq. (3), the 3D coordinate of a laser projection point can be estimated by determining the variable a . In case of the ring laser used in this research, as shown in Fig. 3, all 3D

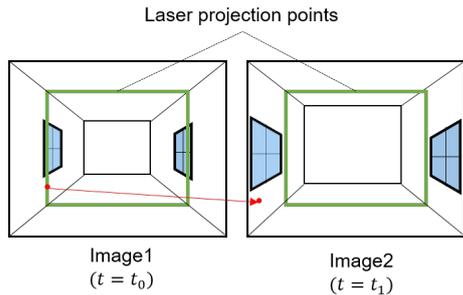


Fig. 4. Overview of pose estimation. If image1 and image2 are the images captured in different locations within the same target structure: 1. Find laser corresponding points for each laser projection points of image1 in image2. Left image's red point shows one example of the laser corresponding points. 2. Calculate translation and orientation between images using laser corresponding point and SfM.

laser projection points lie on the same plane, which is made by irradiated laser light. Therefore, using \mathbf{w} , which is the vector from the optical center of the camera, perpendicular to the laser plane, as shown in Fig. 3, we can obtain Eq. (4).

$$\mathbf{w}^T(\mathbf{X} - \mathbf{w}) = 0. \quad (4)$$

By solving the Eq. (4) for \mathbf{X} , we can obtain each laser projection point's 3D coordinates in the camera coordinate system. The next step is to estimate the pose of the camera and convert these 3D coordinates to the global coordinate system, as described in the next section.

B. Camera Pose Estimation

This section describes the method to estimate a camera position and orientation in each image by SfM (Structure from Motion). In our case, for each frame, since the 3D coordinates of laser projection points are already known in the previous frame (as was explained in Section II-A), the camera's position and orientation can be calculated on an actual scale by using these points. Thus, first, the laser projection points need to be tracked between multiple images via estimation of corresponding points. By doing so, the camera poses in each of the images can be estimated. Figure 4 shows the outline of this process. However, due to the effect of the laser on the texture of the image, this is a difficult process.

C. The Detection of Corresponding Points of Laser Projection Point

The points in other images, which correspond to the laser projection points in n -th image, are named as the "laser corresponding point". These points are used for camera pose estimation as we know their actual distance, making it is possible to estimate the camera motion according to real-world scale using SfM. Red points in Fig. 5 show the "laser corresponding point".

For determining the laser corresponding points between images, block matching can be used. By calculating the

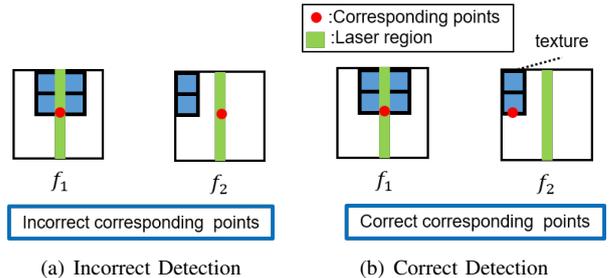


Fig. 5. (a) Incorrect detection of corresponding point. The red points represent the corresponding points between the images. (b) Correct detection of corresponding point.

similarity of images for the matching, the laser corresponding points can be detected. Zero Mean Normalized Cross-Correlation (ZNCC) is used for the similarity calculation. ZNCC can be defined as shown in Eq. (5). Let f_1 and f_2 be the target images, and \bar{f}_1 and \bar{f}_2 represent the average luminance of f_1 and f_2 , where k and h represent the number of horizontal and vertical pixels.

$$R_{zncc} = \frac{\sum_{i,j=0}^{k-1,h-1} ((f_1(i,j) - \bar{f}_1)(f_2(i,j) - \bar{f}_2))}{\sqrt{\sum_{i,j=0}^{k-1,h-1} (f_1(i,j) - \bar{f}_1)^2} \sqrt{\sum_{i,j=0}^{k-1,h-1} (f_2(i,j) - \bar{f}_2)^2}}. \quad (5)$$

D. Difficulty in Detection of Corresponding Points

The texture of the laser, which is always present in each image, will have a major effect on the ZNCC calculation for detecting the laser corresponding points using Eq. (5). Figure 5 shows the case of incorrect detection of the laser corresponding point, due to the effect of the laser texture. In each image, the laser projection points are at different positions in the images. However, the texture induced by the laser is similar between the images. Therefore, there is a possibility of resulting in a high degree of similarity at the incorrect point.

Accordingly, it is necessary to mark the laser projection and calculate similarity by excluding it, in order to eliminate the influence of the laser on the image texture. The previous method of [5] processes Eq. (5) for each pixel in order to mask the laser region. However, when computing Eq. (5) directly in the spatial domain in this manner, the computational complexity is $O(n^2)$, which is high [7]. Thus, we focus on speeding up this similarity calculation in order to find corresponding points.

III. FASTER MATCHING CONSIDERING MASK REGION

The detection of corresponding points takes most of the total calculation time. Therefore, for faster 3D measurement, detecting corresponding points is playing an important role. In this section, the method for faster detection of corresponding points considering mask region, whose difficulty is explained in Section II-D, is proposed. Figure 6 shows the overview of this faster detection of the laser corresponding points. Firstly, the image which represents the mask regions is generated. Next, the faster similarity calculation of ZNCC

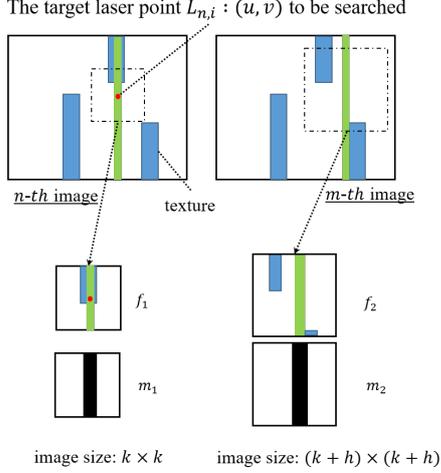


Fig. 6. Overview of fast corresponding point's detection. 1. Crop images around the target laser projection point (x, y) . Cropped images are called as f_1 and f_2 . 2. Create mask images from f_1 and f_2 . Name it as m_1 and m_2 . 3. Find the laser corresponding point of the target laser projection point in f_1 from f_2 , while considering m_1 and m_2 , and using the faster ZNCC computation. 4. From the the distribution of ZNCC, calculate the laser corresponding point with subpixel precision.

is applied considering the mask region. Then, the subpixel precision of the laser corresponding points is shown. The detail is described in the following sections respectively.

A. Generation of Mask Image

First, a mask image is created for each image, as shown in Fig. 6. The area around the laser projection points is expanded and this region is used as a mask for eliminating the influence of the laser. In this binary mask, the laser projection point area expanded, has the value of 0, and the rest has the value of 1. As for the target i -th laser projection point in n -th image, $L_{n,i}$, let us consider the laser corresponding point of $L_{n,i}$ in m -th image. In the n -th image, the image around $L_{n,i}$'s coordinate (u, v) is cropped with the size of $k \times k$ and denoted as f_1 . m_1 is the mask image corresponding to f_1 . Next, in a similar manner, f_2 is obtained by cropping the m -th image around (u, v) with the size of $(k+h) \times (k+h)$. m_2 is the mask image corresponding to f_2 . The ZNCC similarity score considering the mask region is calculated according to Eq. (6). This shows the ZNCC similarity where moves by (s, t) from the origin of the image in the f_2 image. And, \bar{f}_1 and $\bar{f}_{2,s,t}$ are the average luminance of f_1 and $f_{2,s,t}$.

$$R_{zncc}(s, t) = \frac{\sum((f_1(i, j) - \bar{f}_1)(f_2(i-s, j-t) - \bar{f}_{2,s,t}))}{\sqrt{\sum(f_1(i, j) - \bar{f}_1)^2} \sqrt{\sum(f_2(i-s, j-t) - \bar{f}_{2,s,t})^2}}, \quad (6)$$

$$\sum \equiv \sum_{i,j=-k/2}^{k/2} m_1(i, j)m_2(i, j), \quad (7)$$

$$-\frac{h}{2} \leq s \leq \frac{h}{2}, -\frac{h}{2} \leq t \leq \frac{h}{2} \quad (8)$$

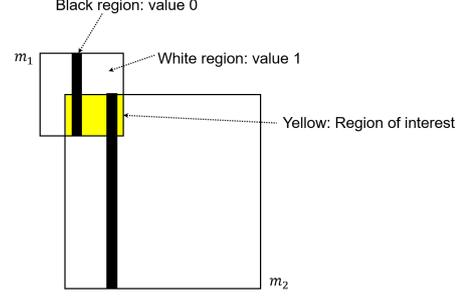


Fig. 7. Region where Fourier transform is applied. m_1 and m_2 are the mask images. Considering only the regions where both mask images have value 1, i.e. the yellow region.

Equation (6) can consider the mask region by ignoring mask region's pixels, however, it takes time because it calculates one pixel by one in spatial domain, whose computation complexity is $O(n^2)$.

B. Acceleration of Masked Block Matching

Next, we compute the distribution of ZNCC over all overlapped positions of the two images f_1 and f_2 . The point where the value of ZNCC is the highest is chosen as the laser corresponding point. In order to speed up this process, methods that match in the frequency domain like [8] and [9] can be used. However, these methods cannot account for the existence of the the mask region. Hence, the method of [7] is applied to our process, since it can compute ZNCC in the frequency domain considering masked regions. Therefore, in the proposed method, [7] is used for fast computation. Compared to spatial domain matching whose computational complexity is $O(n^2)$, the computation complexity of frequency-domain matching is only $O(n \log n)$.

For this computation, a convolutional integral, expressed in Eq. (9) is used for transforming Eq. (5) while considering the mask region. Simply put, Eq. (5) is modified into the calculation in frequency domain by Eq. (9), where the Fourier transform is only applied in the region of interest, wherever the masks have value of 1 in both images, shown as yellow in Fig. 7. Equation (10) is the equation used in the proposed method. In the computation, FFT method is used.

$$\int_{-\infty}^{\infty} f_1(\tau)f_2(t-\tau)d\tau = \mathcal{F}^{-1}(\mathcal{F}(f_1)\mathcal{F}(f_2')), \quad (9)$$

$$R_{zncc} = \frac{a}{b \cdot c}, \quad (10)$$

$$a \equiv \mathcal{F}^{-1}(F_1 \cdot F_2^*) - \frac{\mathcal{F}^{-1}(F_1 \cdot M_2^*)\mathcal{F}^{-1}(M_1 \cdot F_2^*)}{\mathcal{F}^{-1}(M_1 \cdot M_2^*)}, \quad (11)$$

$$b \equiv \sqrt{\mathcal{F}^{-1}(\mathcal{F}((f_1 \cdot f_1) \cdot M_2^*)) - \frac{(\mathcal{F}^{-1}(F_1 \cdot M_2^*))^2}{\mathcal{F}^{-1}(M_1 \cdot M_2^*)}}, \quad (12)$$

$$c \equiv \sqrt{\mathcal{F}^{-1}(M_1 \cdot \mathcal{F}((f_2' \cdot f_2')) - \frac{(\mathcal{F}^{-1}(M_1 \cdot F_2^*))^2}{\mathcal{F}^{-1}(M_1 \cdot M_2^*)}}. \quad (13)$$

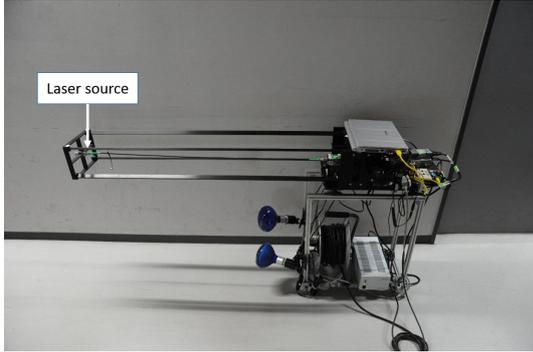


Fig. 8. Experimental equipment



Fig. 9. Experimental environment in a hallway. A textured paper was applied on the wall.

The Fourier transforms of f_1 , f_2 , m_1 and m_2 are represented as F_1 , F_2 , M_1 and M_2 , and M^* is a complex conjugate of M . In addition, f' refers to a 180 degree rotated f . After obtaining the laser corresponding point in the $L_{n,i}$ in m -th image, it is refined to subpixel accuracy using parabola fitting [10] in order to further increase the accuracy.

IV. EXPERIMENT

A. Experiment Data

In the experiment, images were obtained in an indoor hallway environment. 2851 images were captured in steps of 5 mm each. Figure 8 shows the experimental equipment, and Fig. 9 shows the experimental environment. Figure 10 shows one of the images captured in the experiment, used for the computation. The mask image is created by expanding a 30 pixel radius around the laser projection points. The image size of the region to be searched was set as 200×200 pixels and the size of the reference image was set as 280×280 pixels, both of which were cropped around the target laser projection points. In the implementation, we only used the laser projection points which had texture around them. We applied masks to the textureless region manually since the region is fixed in each frame. In addition, we compared the speed of calculation for different sizes of template and searched region.



Fig. 10. Example of an image captured during the experiment. The green line in the picture is the projection from the ring laser. Textures were attached to the surface of wall and floor.

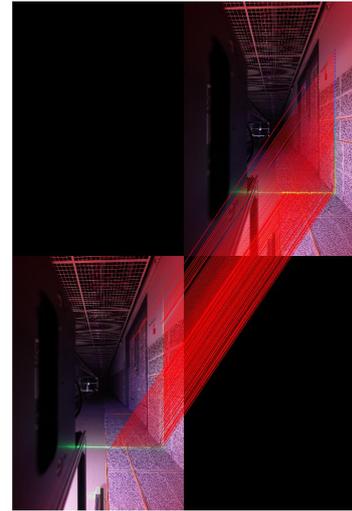


Fig. 11. An example of finding the laser corresponding points between two different images. The left image shows the reference image and the right image shows the image in which the corresponding points are estimated. Each corresponding point pair is connected with red lines shown in the images.

B. Result

Figure 12 shows the 3D measurement result of the experiment. Figure 11 shows the state of the detection of the laser corresponding point in one pair of images. Some detected points as laser projection points are not on the laser line. It is because they are wrongly located in the part of extracting the laser. However, it didn't have the impact on the result in this case.

1) *Calculation Time for Different Image Sizes:* We compared the calculation time for different image sizes. The pair of images used were randomly selected from 2851 images. We computed average of 10 pairs of images for the calculation time. Table I shows the time it took for one pair of images. The results show that the larger image size, the

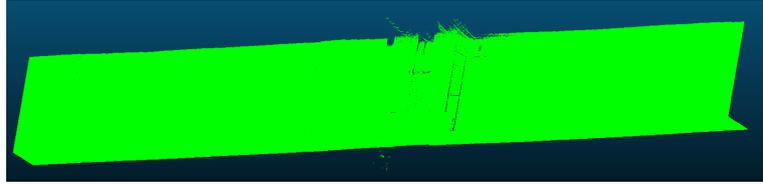


Fig. 12. The 3D measurement result from 2851 images. This confirms that the shape of the hallway, the target environment, can be restored from the calculation.

greater the speed difference between the two methods.

2) *Overall Calculation Time:* We compared the overall calculation time taken by [5], which uses masked spatial domain ZNCC, to that of our proposed method, which uses masked, frequency-domain ZNCC. The calculation times are summarized in Table II. It can be noticed that the speed of matching for our proposed method is about 1.7 times faster when using 2851 images. Overall, the speed increased by about 1.6 times.

3) *Accuracy:* Considering that the images were captured in steps of 5 mm, the groundtruth translation of the camera in each image was calculated. The cross-sectional structure was integrated by using the groundtruth coordinate transformations and the integration result obtained by doing so was used as the groundtruth 3D measurement result for accuracy evaluation. The measurement error between the groundtruth 3D measurement result and the calculated 3D measurement result was computed and compared to the error obtained by [5]’s method. The procedure for evaluating the measurement error was as follows. First, the groundtruth 3D measurement result and the calculated 3D measurement result were aligned using ICP. The total distance between the groundtruth and calculated result was computed and converted to an average error per point. Then, it was compared to the per-point error obtained by [5]’s method. Table III shows both errors. It can be seen that the accuracy is roughly the same in both methods, while our proposed method is around 1.6 times faster, overall.

V. CONCLUSION

In this research, we proposed a method to accelerate high-accuracy 3D measurement for large structures using a masked, frequency-domain matching method. It was experimentally shown that measurement could be accelerated

TABLE I
THE CALCULATION SPEED FOR ONE PAIR IN DIFFERENT SIZES

template size	reference size	proposed [sec]	[5] [sec]
200 × 200	600 × 600	0.35	5.6
200 × 200	400 × 400	0.24	1.6
200 × 200	280 × 280	0.12	0.25

TABLE II
OVERALL CALCULATION SPEED RESULT

	block matching [days]	total time [days]
proposed	5.28	5.91
previous [5]	8.87	9.38

TABLE III

ACCURACY EVALUATION AND COMPARISON TO THE PREVIOUS METHOD.

	error [mm/point]
proposed	47.16
previous [5]	75.28

up to 1.6 times overall, while keeping similar accuracy as compared to the previous, slower method [5]. In this research, we focused on the detection of corresponding points, which occupied almost 90% of the entire calculation. In future, acceleration of other steps in the process will also be considered. Moreover, it still takes some days to complete all procedures. Therefore, there is also room for the acceleration of the detection of corresponding points. As the other future works, the automation of detection of textureless region in the images and eliminating wrong the laser projection points also can be mentioned.

REFERENCES

- [1] M. Golparvar-Fard, J. Bohn, J. Teizer, S. Savarese, and F. Pea-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] R. Lange and P. Seitz, “Solid-state time-of-flight range camera,” *IEEE Journal of Quantum Electronics*, no. 3, pp. 390–397, 2001.
- [3] S. V. der Jeught and J. J. Dirckx, “Real-time structured light profilometry: a review,” *Optics and Lasers in Engineering*, vol. 87, pp. 18–31, 2016.
- [4] C. Cadena, L. Carlone, H. Carrillo, Y. Latif, D. Scaramuzza, J. Neira, I. Reid, and J. J. Leonard, “Past, present, and future of simultaneous localization and mapping: Toward the robust-perception age,” *IEEE Transactions on Robotics*, vol. 32, no. 6, pp. 1309–1332, 2016.
- [5] H. Higuchi, H. Fujii, A. Taniguchi, M. Watanabe, A. Yamashita, and H. Asama, “3d measurement of large structures with a ring laser and a camera using structure from motion for integrating cross sections,” *Proceedings of the 6th International Conference on Advanced Mechatronics (ICAM2015)*, vol. 2015.6, pp. 233–234, 2015.
- [6] 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 omnidirectional rangefinder,” *Proceedings of the 2011 IEEE International Conference on Robotics and Biomimetics (ROBIO2011)*, pp. 1017–1023, 2011.
- [7] D. Padfield, “Masked object registration in the fourier domain,” *IEEE Transactions on Image Processing*, vol. 21, no. 5, pp. 2706–2718, 2012.
- [8] J. P. Lewis, “Fast template matching,” *Vision Interface*, pp. 120–123, 1995.
- [9] K. Takita, T. Aoki, Y. Sasaki, T. Higuchi, and K. Kobayashi, “High-accuracy subpixel image registration based on phase-only correlation,” *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, vol. 86, no. 8, pp. 1925–1934, 2003.
- [10] M. Shimizu and M. Okutomi, “Precise subpixel estimation on area-based matching,” *Systems and Computers in Japan*, vol. 33, no. 7, pp. 1–10, 2002.