# ACCURACY ASSESSMENT OF VISUAL SLAM BY COMBINATION OF PHOTOGRAMMETRY AND TLS POINT CLOUD

Ren Nitta[1] and Masayuki Matsuoka[2*]
[1]Graduate Student, Graduate School of Engineering, Mie University
1577 Kurimamachiya, Tsu, Mie 514-8507 Japan
Email: 422m521@m.mie-u.ac.jp

[2] Associate Professor, Graduate School of Engineering, Mie University,
1577 Kurimamachiya, Tsu, Mie 514-8507 Japan
Email: matsuoka@info.mie-u.ac.jp

**KEY WORDS:** Visual Simultaneous Localization and Mapping (SLAM), Terrestrial Laser Scanner (TLS), point cloud, accuracy assessment, photogrammetry

**ABSTRACT:** Simultaneous Localization and Mapping (SLAM), which simultaneously conducts location estimation and environment-map generation, has attracted much attention in recent years. Visual SLAM can be conducted by video images only at a low cost. Accuracy verification is necessary to improve the ease of using Visual SLAM. This study aims to evaluate the accuracy of indoor SLAM using a hand-held camera. In the target room, 16 spheres were placed as landmarks, and nine cameras were placed to track the SLAM camera. A three-dimensional (3D) model of the room was generated by the Terrestrial Laser Scanner (TLS) as validation data. We conducted SLAM with a hand-held camera, and we evaluated its accuracy by comparing the trajectory obtained through location estimation with the ground truth acquired by a camera installed in the room. As a result, we estimated the location of the spheres attached to the SLAM and displayed them in a 3D model with the SLAM trajectory. In the future, we will estimate the whole trajectory from the locations of multiple spheres in multiple scenes to compare it with the SLAM-generated trajectory.

## 1. INTRODUCTION

### 1.1 Motivation

In recent years, there has been a lot of research on SLAM. SLAM is a technology that simultaneously estimates the location and attitude of the instrument itself and generates the environment map. Visual SLAM can be done at a low cost using only image data; however, it is generally low in accuracy. Examples of SLAM include autonomous robots in warehouses and work-location recording on farms. For SLAM to be used in industry, it is necessary to know the accuracy precisely. The motivation for this study is to promote the use of Visual SLAM and facilitate its development into an agricultural contributor.

### 1.2 Aims

The aim was to evaluate the accuracy of Visual SLAM conducted with a hand-held camera. SLAM trajectory was compared to the reference trajectory derived from installed cameras to estimate the error.

### 1.3 Related Work

Mur-Artal et al. (2015) have presented ORB-SLAM, a feature-based monocular SLAM system that operates in real-time in large and small, indoor and outdoor environments. The system is robust to severe motion clutter, allows wide baseline loop closing and relocalization, and includes full automatic initialization. Long et al. (2022) introduced a novel dense RGB-D SLAM approach for dynamic planar environments, excelling in multi-object tracking and background reconstruction, even in cases of extensive occlusion. It outperforms state-of-the-art methods in localization, mapping, dynamic segmentation, and object tracking, demonstrating robustness to significant camera motion drift. Helmberger et al. (2022) presented the Hilti SLAM Dataset, a dataset of indoor and outdoor real-world sequences. This contributes to the development of highly accurate and reliable SLAM. Krul et al. (2021) used drone-based Visual SLAM for indoor agriculture. SLAM overcomes the inability to use GPS within a small area and indoors.

## 2. MATERIALS AND DATA

### 2.1 Experimental Site

Visual SLAM was examined in our laboratory. It is somewhat spacious and cluttered with various items. Sixteen spheres were placed as landmarks. These were used for the configuration of the position of the installed cameras. Nine cameras were placed in the corners of the room. They track the location of SLAM for a ground truth. We used GoPro HERO 8 Black for all nine cameras and the Visual SLAM camera. Figure 1 shows an example of a camera showing a SLAM camera.

Figure 1. Example of installed camera view

### 2.2 SLAM

Four spheres were attached to the SLAM camera in a tetrahedron shape. The trajectory estimated from these spheres was used as the ground truth. Attitude was also obtained by tracking multiple spheres. The SLAM video was mounted on the dolly to keep the horizontal movement. Figure 2 shows the SLAM equipment. The data was taken from a 47-second video, using a dolly to capture half of the room. We used the monocular mode of ORB-SLAM3 (Mur-Arta et al., 2015) as Visual SLAM. Location and attitude were output. Environment maps were also output as point cloud data.

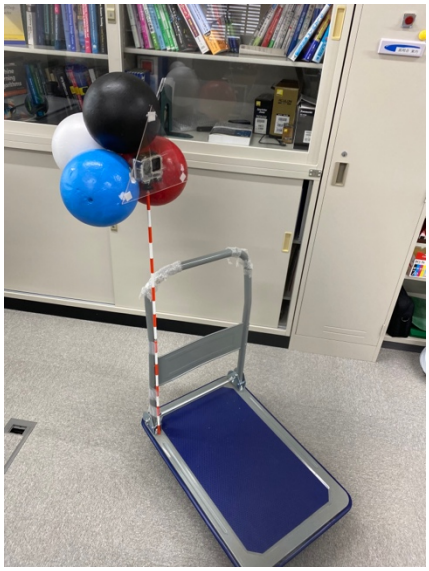

Figure 2. SLAM equipment

### 2.3 TLS

A three-dimensional model of the room was created by TLS. FARO X330 laser scanner was used. We operated TLS at six different locations in a room to capture the entire room. Figure 3 shows the generated 3D model of the target room.

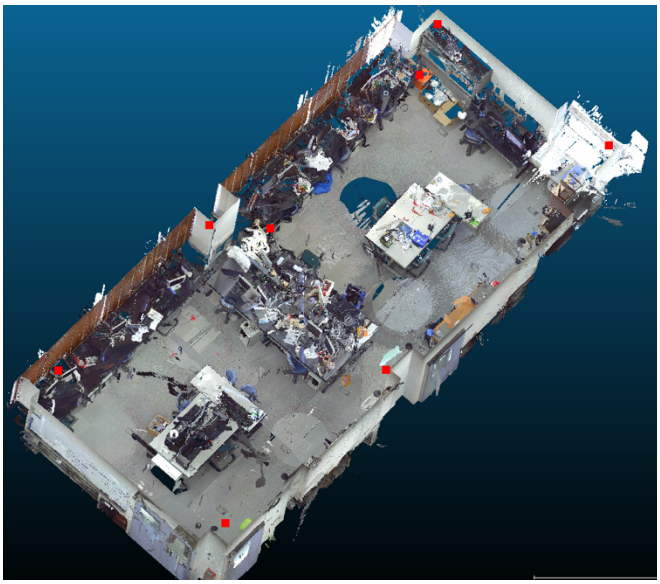

Figure 3. Three-dimensional model of experimental site. Red points show the camera positions

## 3. METHOD

### 3.1 SLAM

The shape of the environment map by SLAM is similar to the 3D model by the TLS point cloud, however, the scale and location were different. Therefore, the environment map was roughly adjusted to the 3D model in scale and location, and then the Iterative Closest Point (ICP) was conducted to align them. The SLAM trajectory was converted by applying the same transformation matrix to align it to the 3D model of the room. The flowchart of the accuracy evaluation is shown in Figure 4.

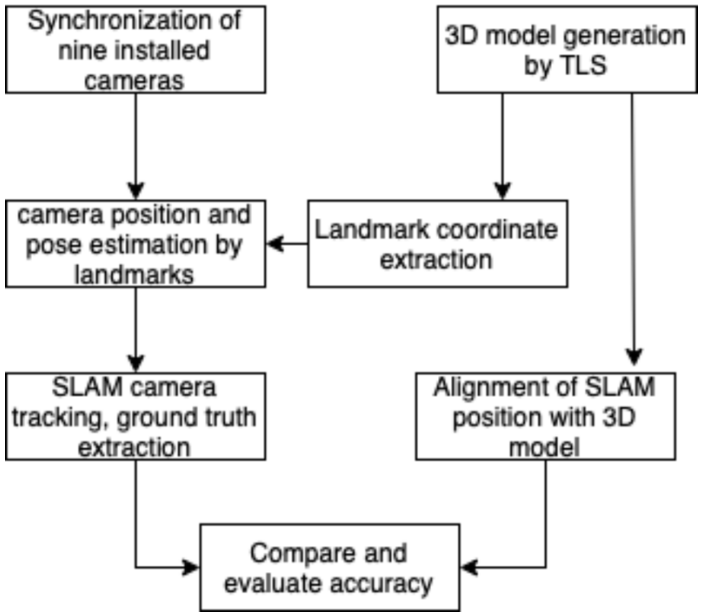


Figure 4. Flow chart of the analysis

### 3.2 Camera Synchronization

To synchronize the time of the videos recorded by nine cameras, we showed one stopwatch to all cameras before SLAM operation. During the analysis, we synchronized the videos using video-captured time.

### 3.3 Position and Attitude of Cameras

Positions and attitudes of the nine fixed cameras were estimated based on a photogrammetric scheme. Pixel positions were measured for the landmark spheres captured by the fixed camera. Then, pixel positions of the same spheres were estimated using the position and attitude of the camera, and we compared these pixel positions. The positions should be the same if the positions and attitudes of the camera are correct. Before this process, we estimated the camera parameters using a checker flag image. After a rough initial position and attitude were given, these were optimized by minimizing the difference between the captured and estimated position of spheres (Hasegawa et al., 1995).

### 3.4 Tracking of the SLAM Camera

If the positions and attitudes of fixed cameras, and the positions of the spheres attached to the SLAM camera are known, we could derive the 3D vectors from the fixed cameras to the SLAM spheres. Vectors were derived for multiple fixed cameras, and then, the location of the SLAM sphere was determined at the intersection of these vectors. By deriving the position of each SLAM sphere, the attitude and location of the camera used for SLAM could be calculated.

### 3.5 Evaluation

SLAM trajectory and environment maps were combined with the 3D model. On the other hand, the movement of the SLAM was estimated from cameras installed at the same time when the SLAM was conducted. These two trajectories were compared and evaluated.

## 4. RESULTS

### 4.1 SLAM Trajectory and Environment Map

Figure 5 shows the point clouds of the trajectory and the environmental map derived by SLAM. The shape of this trajectory was similar to the actual trajectory taken. The environmental map also shows the edges of windows and other features.

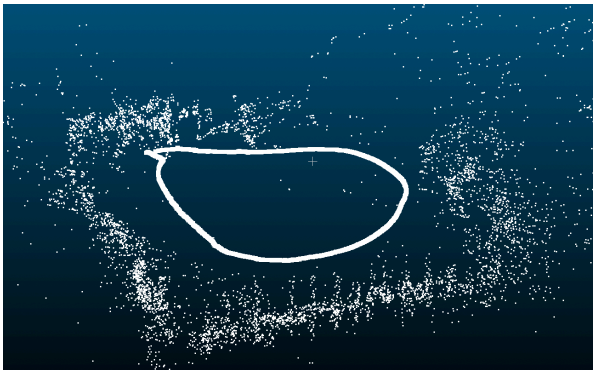

Figure 5. SLAM trajectory and environments map point cloud

### 4.2 Camera Location and Pose Estimation

The red points in Figure 4 are the estimated positions of the camera. They were approximately the same as the location where the camera was installed.

Figure 6. Left: camera view simulated from TLS point cloud. Right: actual camera view

Figure 6 shows the camera's field of view reproduced from the 3D model and the estimated camera location and pose. The views are almost the same.

### 4.3 Tracking SLAM Camera

Table 1 shows the number of vectors used for the position estimation of a single SLAM sphere, and the root mean square errors (RMSEs) in the position estimation. These five points are shown in Figure 7 with red points. Except for Point 5, RMSEs were roughly from 2 cm to 4 cm, independent of the number of vectors.

Table 1. Number of vectors used in the estimation and the RMSE of each combination

| Point | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| Number of vectors | 4 | 4 | 3 | 3 | 4 |
| RMSE (m) | 0.036 | 0.043 | 0.028 | 0.023 | 0.097 |

### 4.4 Display on 3D Model

The trajectory output by the SLAM and the points estimated from the installed cameras are shown in Figure 7. The green line shows the trajectory of the SLAM, and the red, blue, white, and black points show the positions of spheres attached to the SLAM. The trajectory was similar to the one that actually followed, and the location of the spheres was correct in some places.

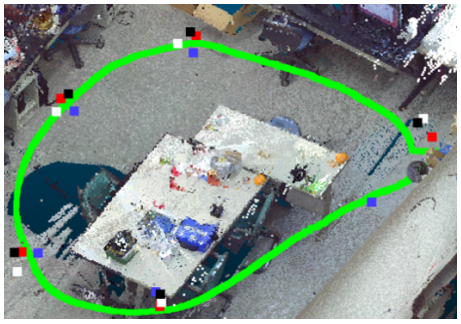

Figure 7. SLAM trajectory of the green line and spheres attached to red, blue, white, and black SLAM camera of each color point

### 4.5 Discussions

The four spheres were attached to the SLAM camera with tetrahedron shape as shown in Figure 2, however, the estimated positions of spheres on the left side of Figure 7 were largely dispersed. Some of the sets of spheres were out of alignment and some were far out of alignment. Regarding Table 1, taking the red sphere as an example, the RMSE of point 5 was large. The reason seems that the camera was far away from the sphere, making it difficult to accurately extract the center of the sphere from the camera, and the vectors from the camera to the center of the sphere were misaligned. Possible solutions are to exclude vector combinations with large misalignments and to capture more accurately the center of the sphere from the image.

## 5. CONCLUSIONS

We introduced the outline of the accuracy evaluation of Visual SLAM and current results. We conducted SLAM and TLS to generate a 3D model of the target room. We also synchronized the time of each video camera and estimated their attitude and position. Using the cameras, we obtained vectors for the spheres attached to the SLAM camera and estimated their positions. In the future, we will track multiple spheres and produce the accuracy of the Visual SLAM pose and location. We will also generate results for multiple trajectories and analyze them.

## REFERENCES

Hasegawa, H., et al., 1995. *Analytical photogrammetry (Kaiseki shasin-sokuryo)*. Japan Society of Photogrammetry and Remote Sensing, Tokyo, pp. 46-56. (in Japanese)

Helmberger, M., Morin, K., Berner, B., Nitish Kumar, Cioffi, G., and Scaramuzza, D., 2022. The Hilti SLAM Challenge Dataset. *IEEE Robotics and Automation Letters*, 7(3), pp. 7518-7525. https://doi.org/10.1109/LRA.2022.3183759

Krul, S., Pantos, C., Frangulea, M., and Valente**,** J., 2021. Visual SLAM for Indoor Livestock and Farming Using a Small Drone with a Monocular Camera: A Feasibility Study. *Drones*, 5(2), 41. https://doi.org/10.3390/drones5020041

Long, R., Rauch, C., Zhang, T., Ivan, V., Lam, T., and Vijayakumar, S., 2022. RGB-D SLAM in Indoor Planar Environments with Multiple Large Dynamic Objects. *IEEE Robotics and Automation Letters*, 7(3), pp. 8209-8216. https://doi.org/10.1109/LRA.2022.3186091

Mur-Artal, R., Montiel, J., and Tardós, J., 2015. ORB-SLAM: A Versatile and Accurate Monocular SLAM System. *IEEE Transactions on Robotics,* 31(5), pp. 1147-1163. https://doi.org/10.48550/arXiv.1502.00956