ICCEPM 2024 https://dx.doi.org/10.6106/ICCEPM.2024.1019
The 10th International Conference on Construction Engineering and Project Management
Jul. 29-Aug.1, 2024, Sapporo

Location Measurement System for Automated Operation of Construction Machinery Using Visual SLAM

Masaki CHINO[1]*, Atsushi YAMASHITA[2]

[1] *Frontier Research Department, Technical Research Institute, Hazama Ando Corporation, Japan,* E-mail address: chino.masaki@ad-hzm.co.jp
[2] *Department of Human and Engineered Environmental Studies Graduate School of Frontier Sciences, The University of Tokyo, Japan,* E-mail address: yamashita@robot.t.u-tokyo.ac.jp

Abstract: In the construction industry, there is a growing demand for improving productivity, and development of autonomous operation systems for construction machinery is progressing. Autonomous operation of construction machinery requires positioning information because construction must be carried out at planned locations. In this paper, we focused on Visual Simultaneous Localization and Mapping (Visual SLAM) as a method for obtaining location information for construction machinery and proposed an automated operation system using Visual SLAM. For automated driving, the indirect method based on ORB features is used in Visual SLAM, and processes such as mask processing for surrounding moving objects and measurement of initial positions using markers are performed. With the proposed system, it was confirmed that it is possible to perform automated operation in an experimental environment using the location information output by Visual SLAM. In addition, the experiment was conducted to verify the measurement accuracy when using Visual SLAM during construction work at actual construction sites. As a result, the measurement accuracy was less than 500 mm, which is a usable accuracy for actual construction. By using this system, it is possible to obtain the location information of construction machinery even in environments where GNSS cannot be used, and productivity at construction sites can be improved by performing automated operation.

Key words: Visual SLAM, Localization, Automated driving, Civil Engineering Site

1. INTRODUCTION

The construction industry is a field where mechanization has long been delayed due to its characteristics of custom order production and on-site outdoor production. In developed countries, the labor productivity of the construction industry is about 80 % of that of the manufacturing industry, and the low labor productivity is a problem [1]. Therefore, it is required to improve labor productivity. One of the methods is the automation of construction by construction machinery [2]. When construction is automated, it is expected to reduce manpower and improve productivity because construction can be carried out day and night without manpower.

In order to automate the operation of construction machinery, it is necessary to have location information of the construction machinery to move to the planned position and perform work according to each construction machinery. As a method of obtaining location information, there is Simultaneous Localization and Mapping (SLAM), which mounts external sensors on construction machinery and simultaneously creates an environmental map of the surroundings and estimates its own position. Among SLAM, the method of calculating location information from images using a camera as an external sensor is called Visual SLAM.

In this paper, a system for automated operation of construction machinery using location information obtained by Visual SLAM is proposed. It is aimed to verify the applicability of Visual SLAM in civil engineering sites by verifying its accuracy of location information in civil engineering site.

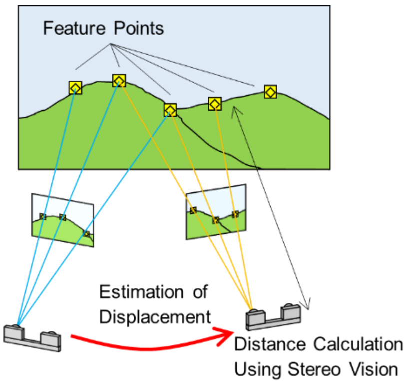


Figure 1: Outline of Visual SLAM

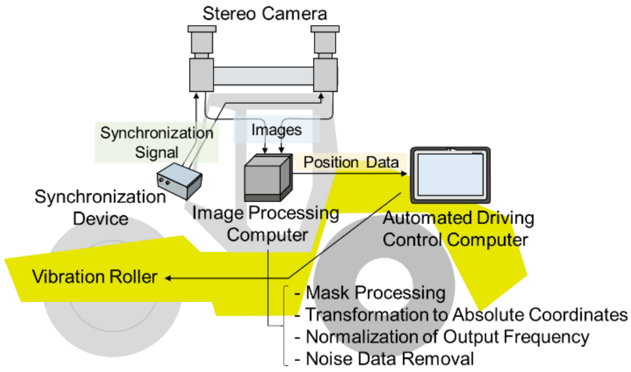


Figure 2: Configuration of automated driving system using Visual SLAM

2. LOCATION MEASUREMENT SYSTEM FOR AUTOMATED OPERATION OF CONSTRUCTION MACHINERY

2.1. Method of Obtaining Location Information of Construction Machinery

As a method of obtaining location information of construction machinery, there is the Global Navigation Satellites System (GNSS) that uses information obtained from artificial satellites [3]. GNSS has been widely used in construction sites because it can obtain location information in real time with centimeter-order accuracy in outdoor environments. However, GNSS which uses signals from artificial satellites cannot be used in mountainous areas where multipath occurs, or in tunnels and underground spaces where signals cannot be received. Also, there are times when the accuracy is not stable due to the changing arrangement of artificial satellites.

Another method is to use a total station. A total station is a surveying instrument that simultaneously measures angles and distances by irradiating a prism target with laser light. By using a machine with the function of automatically tracking the target, it is possible to measure the location information of the moving construction machinery in real time and use it as location information for automated driving [4]. However, since the total station uses the transmission and reception of laser light, it cannot measure the position if there is an obstacle between the total station and the construction machinery. It cannot be used generally at construction sites where various construction machinery are around, the construction range is vast, and the terrain is complex.

Since SLAM can determine the relative amount of movement to the surrounding environment, it can be used for automated driving by giving the initial position and orientation in absolute coordinates. Among SLAM, LiDAR SLAM uses LiDAR which uses light such as infrared as an external sensor. LiDAR directly measures the distance to an object by irradiating light, measuring scattered light, and obtains point cloud data. It can measure the distance to surrounding objects with high accuracy and a wide field of view. However, the accuracy may decrease in places such as tunnels and embankment slopes where structural changes are small because it is difficult to correspond point clouds between frames [5].

Visual SLAM measures the relative amount of movement by matching frames using feature points and brightness on the image, as shown in Figure 1. While it has the advantage of not being constrained by the measurement location [6], there is a concern about a decrease in measurement accuracy in places like civil engineering sites where there are few structures around. Research has been conducted on the use of Visual SLAM for the automated operation of omnidirectional AGVs [7]. However, it was conducted in an indoor environment, and the images that can be obtained are different from those at construction sites, so it is difficult to apply it as it is. There are no examples of using Visual SLAM to obtain location information of construction machinery for automated driving.

2.2. Visual SLAM

There are two main methods in Visual SLAM. The first method is the direct method which minimizes photometric errors using image brightness values and matches between images. The second method is the indirect method which minimizes geometric errors using feature points in the image and matches between images. When mounting a camera on construction machinery, there is a concern about vibrations caused by driving and working. The indirect method is less affected by motion blur due to vibration compared to the direct method. Therefore, in this system, stella VSLAM [9], a feature-based method using ORB features [8], is used for Visual SLAM.

Absolute distance is necessary for automated driving. Therefore, a stereo camera using two digital cameras is configured to obtain absolute distance from the information of known baseline length. Grasshopper (Teledyne FLIR LLC, Oregon USA) for the cameras and SyncUsb3 (ViewPLUS, Tokyo JAPAN) for synchronizing the shooting of the left and right cameras are used.

In Visual SLAM, it is assumed that the environment is static, with no moving objects within the camera's shooting range, as self-position estimation is performed by stably tracking landmarks as feature points. However, construction sites are dynamic environments where construction machinery and workers are located close to the camera. Therefore, a process of applying a mask to moving objects step by step according to the ratio they occupy in the image is performed and moving objects from the calculation of self-position estimation is removed [10]. This allows for stable tracking even if landmarks are occluded by moving objects.

The location information obtained by Visual SLAM is the camera coordinate system, which is relative location information with the camera position as the origin. At construction sites, the rectangular coordinate system, which is absolute location information based on latitude and longitude, is used. To obtain the absolute coordinates of the construction machinery, multiple markers whose coordinates have been measured in advance in the rectangular coordinate system are installed at the start position. A picture which captures all the markers is taken, and the absolute coordinates are calculated by the PnP method [11] using the coordinates of the markers and the image that captured the markers, and this is set as the start position.

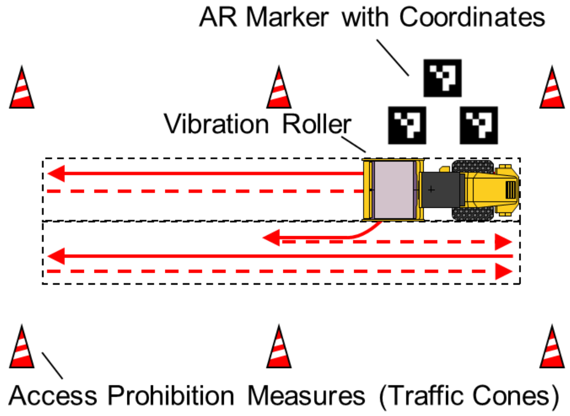
2.3. Automated Operation of Construction Machinery

Existing control programs for automated driving use location information obtained by GNSS as input. The location information that can be obtained by GNSS is latitude and longitude using the geographic coordinate system. On the other hand, the location information output by Visual SLAM is the rectangular coordinate system as mentioned earlier. In different coordinate systems, the automated driving system cannot accept data. Therefore, the coordinate system of the location information by Visual SLAM is converted from the rectangular coordinate system to the geographic coordinate system. Also, the data output frequency is not fixed in the location information calculation of Visual SLAM because there are branching processes. Therefore, a process to regularize the output frequency is performed. If the output interval by Visual SLAM is shorter than the set output frequency, the most recently output location information is used. If the output interval is long, data will be missing, but in this case, the speed from past performance data is estimated and the location information is interpolated. To maintain the robustness of self-position estimation against noise, a process to not output data is performed when data that deviates greatly from the value of the previous location information is output. Figure 2 shows the configuration diagram of a system that adds functions for automated driving to Visual SLAM using a stereo camera.

2.4. Installation of Stereo Camera on Construction Machinery

A vibration roller SV512 for earthwork is used as the construction machinery for automated driving. The vibration roller used in civil engineering work is a construction machine that compacts soil by driving while vibrating the iron wheel. It is believed that embankment work, which can repeatedly perform the same work continuously in the same place, can greatly contribute to productivity improvement by automated driving. In embankment work site, one side is a valley, and the other side is a slope in the direction of the vibration roller's progress.

The stereo camera on the vibration roller is installed on top of the driver's seat, where the impact of vibration is small, and the visibility is good. When obtaining depth information with a stereo camera, the longer the baseline length between the cameras, the larger the disparity between the left and right images, so the error becomes smaller. On the other hand, since the stereo camera is placed on top of the driver's seat of the vibration roller, the baseline length needs to be smaller than the size of the driver's seat. Therefore, the baseline length was set to 1,100 mm. The vibration frequency of SV512 compaction is 33 Hz, and to suppress the impact of vibration, the design is made so that the natural frequency of the camera mount does not match. The natural frequency of the designed camera mount was 61.6 Hz.



(a) Automated driving setting route

(b) Experimental situation

Figure 3: Experimental situation of automated driving

Regarding the method of fixing the stereo camera, most of the construction machinery used on site is rented, so it is difficult to modify the construction machinery itself. It is necessary to fix it robustly against vibration. Therefore, the stereo camera is fixed with neodymium magnets. However, since there is a concern about damage to the camera due to a strong impact at the time of installation, it can be turned ON and OFF using a lever. The strength of the magnet is selected to be strong enough not to come off due to acceleration caused by vibration.

To easily capture the change of feature points between frames, the stereo camera is placed facing the side in the direction of progress. As mentioned earlier, on the side of the vibration roller, one side is a valley and the other side is a slope, and the objects that appear in the image when shooting are greatly different. Good conditions for location measurement with Visual SLAM include being able to take feature points at both near and far places and being able to take feature points evenly throughout the image. On the valley side, close points on the ground and distant points such as the mountain on the opposite bank appear in the image, but it is considered that the distance between these two places becomes large and discontinuous. Also, it is considered that there may be points where feature points cannot be taken because the mountain on the opposite bank are too far away. On the other hand, if the camera is pointed towards the slope side, it is considered that the ground and the slope will be reflected in the image taken. Since the ground and the slope are connected and the slope of the embankment changes evenly, the depth information changes continuously from the front to the back in the image. Therefore, feature points can be obtained at both far and near places in the image. From this, it is considered that the slope side can take feature points evenly in the image and the position measurement accuracy of Visual SLAM will be high. Therefore, the stereo camera is installed facing the slope side. Also, the height of the driver's seat of the vibration roller is about 3 m, and if the camera is pointed horizontally, there is a high possibility that the upper part of the image will become sky. In order to reduce the range where the sky, which is difficult to capture feature points, is reflected, the camera is pointed downwards.

3. EXPERIMENT

In Chapter 3, experiments were conducted with two purposes. The first purpose is to verify whether it is possible to automatically drive a vibration roller using the mechanism of the Visual SLAM system discussed in Chapter 2. The second purpose is to confirm the practicality by installing the Visual SLAM system on a vibration roller that works at construction site and measuring the location.

3.1. Application to Automated Driving

An experiment was conducted to verify whether it is possible to automatically drive a vibration roller using the mechanism of the Visual SLAM system discussed in Chapter 2.

The experiment was conducted in an outdoor experimental field within the research institute. Multiple markers whose coordinates were measured by a total station were installed at the start position to obtain the coordinates of the start position of the vibration roller's run. The markers were taken by the camera installed on the vibration roller, and the start position of the vibration roller's run was calculated by the

(a) Embankment site

(b) Stereo camera on vibration roller

Figure 4: Experimental situation of actual construction work

Table 1: Experimental conditions

Shooting direction	Valley side, Embankment slope side
Downward angle [deg]	20, 30, 40

PnP method. In automated driving, as in actual construction, the vibration roller was driven automatically with a setting to move forward and backward 20 m and then change lanes. Figure 3 (a) shows the route driven in the experiment, and Figure 3 (b) shows the experimental situation. During the experiment, an operator was in the driver's seat of the vibration roller to safely stop it in case of an unexpected situation.

As a result of the experiment, it was confirmed that the vibration roller could be driven automatically to the end of the set route using the location information obtained by Visual SLAM. Although the impact of vibration was a concern, there was no damage or detachment of the device due to vibration, and there was no significant difference in error due to the vibration or no-vibration.

3.2. Location Measurement Accuracy in Actual Construction

In Section 3.1, it was confirmed that the vibration roller can be operated automatically using the location information by Visual SLAM. In order to verify its practicality, an experiment was conducted to check the location measurement accuracy when the Visual SLAM system was installed on a vibration roller performing construction work.

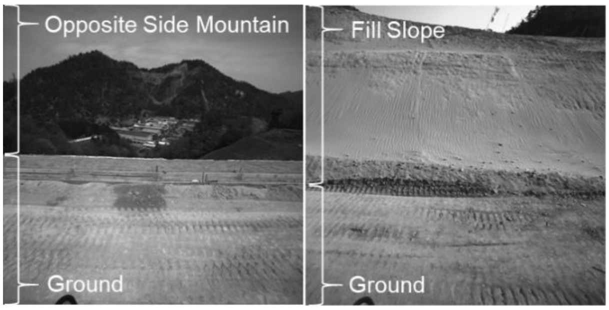
The experiment was conducted on the compaction work at the embankment site shown in Figure 4 (a). The vibration roller was driven for about 40 m with the iron wheel vibrating in manned operation, and several lanes were driven. Location information was calculated by Visual SLAM from the images obtained during driving. Figure 4 (b) shows a photo of the camera installed.

The experimental conditions are shown in Table 1. The shooting direction of the camera was set to the valley side and the slope side, and the downward angle of the camera was changed from 20 degrees to 40 degrees.

The location information of the vibration roller during construction was measured with an automatic tracking total station. The coordinates by Visual SLAM and by the total station were compared, and the difference of two coordinates was used as the evaluation item. The experiment was conducted during the day. Since the camera was pointed downward from the horizontal and there was no backlight, so there was no change in illuminance due to sunlight.

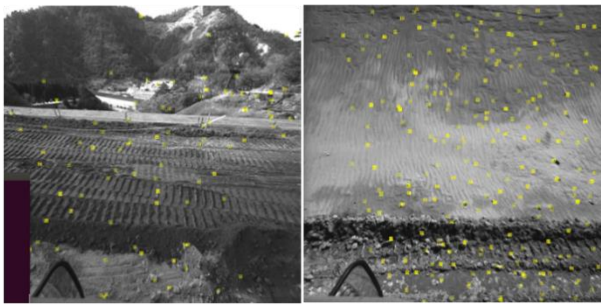
3.3. Experimental Results and Discussion

Figure 5 shows images taken towards the valley side and the slope side. On the valley side, the lower half is the ground, and the upper half is the mountain on the opposite bank. In the image of the slope, the lower half is the ground, and the upper half is the slope. Figure 6 shows images where ORB feature points have been extracted from the captured images. The yellow squares in the image are the extracted feature points. As expected, in the valley side image, feature points can be uniformly obtained in the lower half of the ground, but the number of feature points in the upper half of the mountain on the opposite bank was sparse. On the other hand, in the image of the slope, feature points appeared evenly throughout the image.



(a) Valley side (b) Slope side

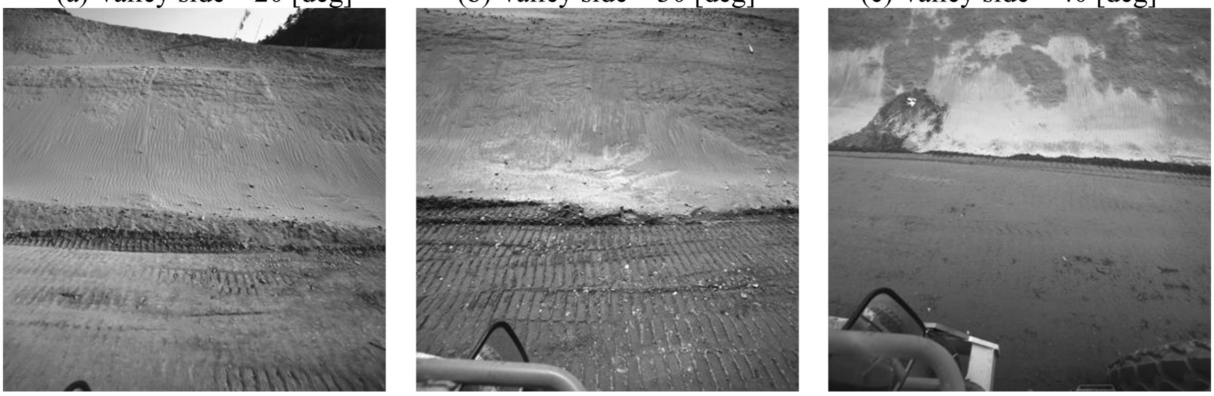
Figure 5: Captured images



(a) Valley side (b) Slope side

Figure 6: Detection of feature points

(a) Valley side – 20 [deg] (b) Valley side – 30 [deg] (c) Valley side – 40 [deg]



(a) Slope side – 20 [deg] (b) Slope side – 30 [deg] (c) Slope side – 40 [deg]

Figure 7: Captured images due to camera downward angle

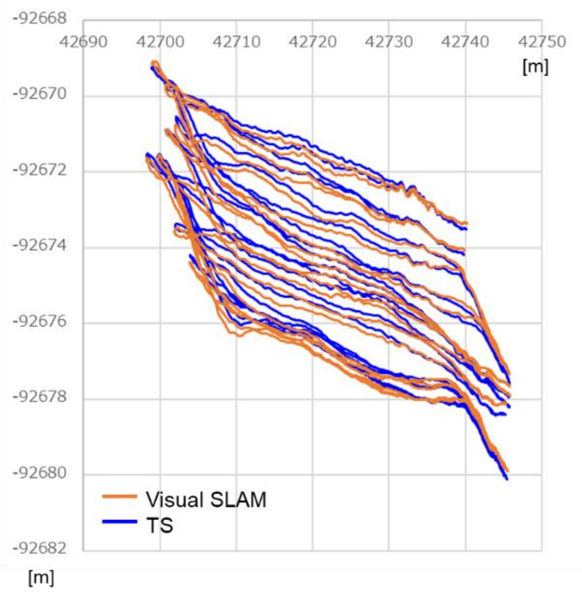


Figure 8: Trajectories of location measurement by Visual SLAM and total station

Table 2: Difference between the measurements by Visual SLAM and total station

Shooting direction	Downward angle [deg]	Average [mm]	Maximum [mm]
Valley side	20	139	1,479
Valley side	30	478	987
Valley side	40	170	2,013
Slope side	20	178	487
Slope side	30	203	980
Slope side	40	230	750

Figure 7 shows the difference in the images obtained depending on the downward angle of the camera. As the angle increased from 20 degrees to 40 degrees, the proportion of the ground in the image increased on both the valley side and the slope side.

Figure 8 shows the results of the trajectory of location measurement by Visual SLAM and the total station. The graph is for when the shooting direction is towards the slope and the camera is tilted downward by 20 degrees. The vertical axis of the graph is the north-south position, the horizontal axis is the east-west position, the red is the trajectory of Visual SLAM, and the blue is the trajectory of the total station. The difference from the total station was small, indicating that the location measurement was performed correctly. Table 2 shows the difference between the measurement results by the total station and Visual SLAM. The average values were all less than 500 mm. The maximum value of the difference increased when the shooting direction was towards the valley side. This may be occurred because the extracted feature points became intermittent, and an error occurred in the matching. When the camera was pointed towards the slope side and the downward angle was changed, the result at 20 degrees was the best. While only a part of the slope is reflected at 40 degrees, the entire slope is reflected at 20 degrees. It is thought that the result was better at 20 degrees because continuous feature point detection is possible by reflecting the slope. The compaction management block size of the vibration roller in the management guidelines for embankment compaction work using ICT equipment in Japan [12] is 500 mm. Since the difference between the location by the total station and Visual SLAM was less than 500 mm at most when the camera was pointed towards the slope, it can be said that this system is practical.

4. CONCLUSION

In this paper, a Visual SLAM system for automated driving of construction machinery used in civil engineering work was proposed. It was demonstrated that it is possible to perform automated driving using the location information obtained by Visual SLAM in the experimental environment. In addition, it was found that effective images for performing Visual SLAM can be obtained by pointing the camera at the slope where feature points can be uniformly obtained in an actual embankment site. As a result of the experiment, it was shown that the location measurement accuracy of Visual SLAM is less than 500 mm at most, indicating that it has an accuracy that can be used for quality control.

REFERENCES

[1] The Japan Institute for Labour Policy and Training: Databook of International Labour Statistics 2023, 2023.

[2] Satoru Miura, Izuru Kuronuma, Kenniti Hamamoto: “Next Generation Construction Production System: On Automated Construction Machinery”, Proceedings of the 7th Civil Engineering Conference In the Asian Region, pp. 1-11, 2016.

[3] Ying Jiang, Xiangyu He: “Overview of Applications of the Sensor Technologies for Construction Machinery”, IEEE Access, Vol. 8, pp. 110324-110335, 2020.

[4] Nobuaki Kurihara, Hiromichi Miyazaki, Hiroaki Aoki, Saburo Katayama: “Demonstration of Vibrating Roller of Autonomous Control Type”, Technical Report of Taisei Advanced Center of Technology, Vol. 47, pp. 1-6, 2014.

[5] Julian Nubert, Etienne Walther, Shehryar Khattak, Marco Hutter: “Learning-based Localizability Estimation for Robust LiDAR Localization”, Proceedings of the 2022 IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 1-8, 2022.

[6] Runqiu Bao, Ren Komatsu, Renato Miyagusuku, Masaki Chino, Atsushi Yamashita, Hajime Asama: "Cost-effective and Robust Visual Based Localization with Consumer-level Cameras at Construction Sites", Proceedings of the 2019 IEEE 8th Global Conference on Consumer Electronics, pp. 1007-1009, 2019.

[7] Cosimo Patruno, Roberto Colella, Massimiliano Nitti, Vito Reno, Nicola Mosca, Ettore Stella: "A Vision-Based Odometer for Localization of Omnidirectional Indoor Robots", Sensors, Vol. 20, Issue 3, pp. 1-25, 2020.

[8] Ethan Rublee, Vincent Rabaud, Kurt Konolige, Gary Bradski: "ORB: An Efficient Alternative to SIFT or SURF", Proceedings of the 2011 International Conference on Computer Vision, pp. 2564-2571, 2011.

[9] Shinya Sumikura, Mikiya Shibuya, Ken Sakurada: "OpenVSLAM: A Versatile Visual SLAM Framework", Proceedings of the 27th ACM International Conference on Multimedia, pp. 2292-2295, 2019.

[10] Runqiu Bao, Ren Komatsu, Renato Miyagusuku, Masaki Chino, Atsushi Yamashita, Hajime Asama: "Stereo Camera Visual SLAM with Hierarchical Masking and Motion-state Classification at Outdoor Construction Sites Containing Large Dynamic Objects", Advanced Robotics, Vol. 35, No. 3-4, pp. 228-241, 2021.

[11] Vincent Lepetit, Francesc Moreno-Noguer, Pascal Fua: "EPnP: An Accurate O(n) Solution to the PnP Problem", International Journal of Computer Vision, Vol. 81, No. 2, pp. 155-166, 2009.

[12] Ministry of Land, Infrastructure, Transport and Tourism: "TS GNSS wo mochiita morido no shimekatame kannri youryou (in Japanese) (Management Procedure for Compaction of Embankments Using TS and GNSS)", 2020.