Trajectory Recovery of Visual Odometry at Corners by Temporal Stereo Point Cloud Registration

Yusuke Eshima, Kazuha Saito, Masafumi Nakagawa

Department of Civil Engineering, Shibaura Institute of Technology, Japan

ah20034@shibaura-it.ac.jp

Abstract *Mobile mapping systems (MMS) and unmanned aerial vehicles (UAVs) are widely used to acquire 3D data for safe and rapid inspection works of infrastructure, such as bridges, dams, roads, railways, and other structures. In recent years, simultaneous localization and mapping (SLAM) has been applied to image and point cloud acquisition for construction and infrastructure inspection, such as volume measurement of earthworks using handheld SLAM scanners, autonomous patrol robot operation with SLAM, UAV flight control with SLAM for bridge inspection in non-GNSS positioning environments. SLAM can be categorized into SLAM using a laser scanner (LiDAR-SLAM). Representative conventional studies on Visual SLAM are mainly UAV control with ROS-based Visual SLAM and 2D modeling of complex shapes and ortho imaging using UAVs with OpenREALM. The ROS-based Visual SLAM includes a comparison of LiDAR, RGB cameras, and stereo cameras. Our previous results include the development of flight control methodologies to achieve infrastructure inspection using a UAV. We have developed a methodology to improve the stability of positioning visual odometry using multi-directional IMU stereo cameras. We also developed a methodology to improve the availability of positioning and seamless GNSS/non-GNSS positioning by quickly switching of the positioning mode between visual odometry and RTK-GNSS positioning. However, the technical issues of visual odometry remained the high environmental dependency of the processing, such as the presence of lighting and image features, and the error accumulation problem caused by inertial navigation. The additional technical issue is that self-position estimation by visual odometry is not easy when images are blurred due to camera motion and rotation. Therefore, this study focuses on the combination of visual odometry and 3D map matching. We propose a methodology based on temporal point cloud registration to rectify position and attitude estimation errors caused by image blur during when sudden rotations of a moving stereo camera.*

Keywords: *Visual odometry, SLAM, 3D map matching*

Introduction

In recent years, the acquisition of 3D spatial data using mobile mapping systems (MMS) and unmanned aerial vehicles (UAVs) has been widely used as a safe and fast method for inspecting infrastructure facilities such as bridges, dams, roads, and railways. In recent 3D measurements, the use of simultaneous localization and mapping (SLAM) to acquire point clouds is increasing. Examples of SLAM applications in infrastructure inspection include earthwork data acquisition using handheld SLAM scanners. The SLAM can be classified into LiDAR-based SLAM (LiDAR-SLAM). In this paper, we focus on visual SLAM and visual odometry for camera pose estimation. Visual SLAM and visual odometry are intended for autonomous robots and UAV control in non-GNSS (global navigation satellite system) positioning environments. However, the technical challenges of visual odometry include high environmental dependency in processing, such as the presence or absence of lighting and image features, and the error accumulation problem caused by inertial navigation. In addition, visual odometry suffers from self-position estimation error when the captured image is blurred due to camera motion and rotation. Therefore, we propose a method to avoid the problem of image blur problem when the camera moves and rotates. However, when the camera image is blurred, the detection and tracking of the point for camera pose estimation becomes difficult. In this study, we apply a 3D map matching process when images are blurred. We develop a stable camera pose estimation methodology with an integration of 3D map matching and visual odometry to achieve robust camera pose estimation for UAVs.

First, we describe our methodology, which consists of relative camera pose estimation by visual odometry and 3D map matching for the camera pose, correction process. Next, we describe an overview of the indoor walking measurement experiments. In the experiments, we summarize and discuss the data obtained by visual odometry using an IMU stereo camera. Then, we investigate whether errors in camera pose estimation by visual odometry during sharp turns are compensated by 3D map matching.

Literature Review

A study describes techniques for small UAVs in GPS-denied environments. As an introduction to this paper, UAVs are becoming smaller and more compact. Computing costs and power consumption are also becoming more constrained. Then, there is a need for technologies for self-positioning estimation and environmental mapping technologies without relying on GPS. Under these constraints, the development of SLAM systems that operate with high accuracy and in real time is an important issue. This system uses the bearing-only observations to estimate the position of the vehicle and build a feature map during the flight. An inverse depth method is used in applied in the undelayed feature initialization. A method that combines the Mahalanobis distance and the descriptor match of the SIFT features is used to improve the robustness of the data association. Simulation and real-world experiments were conducted to test the performance of the system. As a result, it can be concluded that the proposed SLAM algorithm can limit the error of vehicle position estimation while building a 3D feature map in GPS-denied environments. Future work will focus on improving the computational complexity of the SLAM. In the extended Kalman filter (EKF) based algorithm, the state is augmented with new observations of features. The computational complexity increases quadratically with the number of features. These drawbacks are unavoidable in EKF-based methods for real-time SLAM for small UAVs. (Chaolei et al., 2012). Monocular visual SLAM for small UAVs in GPS-denied environments. However, there is a need to discuss the development of a Rao-Blackwellized based FastSLAM and the stabilization of the computation as a linear complexity. Another related study describes two tasks in UAVs flying for bridge inspection First, there are no UAVs that can fly seamlessly in indoor and outdoor environments (GNSS/non-GNSS positioning environments) such as bridges, and flight control is difficult under bridges where satellite signals cannot be received. However, autonomous UAV flight is typically enabled by using GNSS positioning, which means that seamless indoor-outdoor flight cannot be realized in non-GNSS environments, such as the interior subspace of bridges. Second, multi-temporal images in surveys must be accurately superimposed, and it is not easy to match the camera position and angle of view with the previous survey when using UAVs for multi-temporal imaging. Therefore, there is a need for a seamless indoor/outdoor flight function that can fly over bridges with mixed GNSS/non-GNSS positioning environments and a seamless indoor/outdoor external positioning function that combines the position and attitude of the UAV-mounted camera with the position and attitude of the previous survey. Therefore, this study proposes a seamless indoor/outdoor positioning method that combines RTK-GNSS positioning with visual odometry to switch positioning modes and suppress the accumulation of errors caused by visual odometry, and a method to match the position and attitude of the UAV-mounted camera with those of the previous survey by detecting and controlling image feature points. This also proposes a method of detecting and controlling image feature points to match the position and attitude of the UAV-mounted camera with the previous survey. In addition, a method is proposed for matching the position and attitude of the UAV-mounted camera by superimposing images taken during multiple periods using image feature point detection and correspondence processing. Prototypes of these functions have been developed and their performance is being verified. (Saito et al., 2024). However, it is necessary to discuss the development of a system to utilize high-resolution images taken by UAVs equipped with the proposed method as building information modeling and civil engineering information modeling data.

Methodology

The proposed method consists of camera pose estimation by visual odometry, and 3D map matching for camera pose correction processing, as shown in Figure 1. First, the relative camera poses are estimated by visual odometry, and 3D map matching is applied to temporal point clouds with iterative closest point (ICP) in parallel. Next, to connect the relative camera pose error correction in visual odometry, a camera turn at a corner is detected using feature detection from images based on motion blur. When the estimated camera poses of visual odometry are discontinuous, relative camera pose correction is applied based on point cloud matching by ICP.

a. Stereo camera calibration

The internal and external parameters of each camera are estimated in a stereo camera calibration with Zhang's methodology. The stereo camera calibration estimates the stereo camera baseline length, focal length, and lens distortion parameters. In our study, the estimated parameters are evaluated with reprojection errors using a checkerboard.

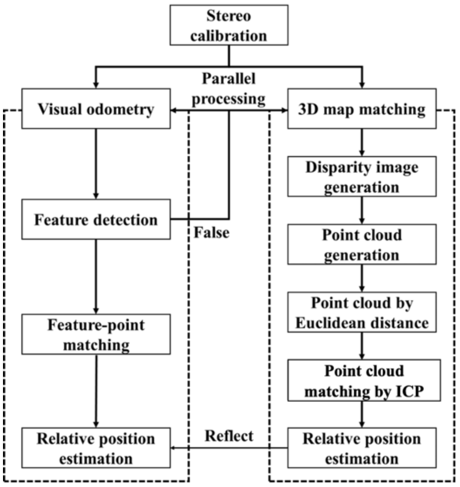

Stereo calibration
Parallel processing
Visual odometry
3D map matching
Disparity image generation
Feature detection
False
Point cloud generation
Point cloud by Euclidean distance
Feature-point matching
Point cloud matching by ICP
Reflect
Relative position estimation
Relative position estimation

Figure 1: Proposed methodology.

b. Visual odometry

Visual odometry is a camera pose estimation using images. For visual odometry to work effectively, the positioning environment must have sufficient light and dark and a textured surface to allow feature point extraction during movement measurements. It is also necessary to ensure that the scene changes overlap in a sufficiently continuous manner to capture camera frames continuously. It is also useful as a complement to other types of sensors such as GPS, inertial measurement units (IMU), and LiDAR. Literature studies have shown the use of multi-directional IMU stereo cameras to improve the accuracy of camera pose estimation by visual odometry, and the robust control of UAVs with multi-directional stereo cameras. In our study, visual odometry with a single IMU stereo camera is used as the basic process, focusing on the advantages of lightweight 3D measurement systems and scale factor determination methods.

c. 3D map matching

3D map matching is the process of correcting for off-track camera positions in the acquired 3D map (3D data). Map matching is the correction of errors in position information acquired by GNSS or SLAM to a considered optimal position by map data. Literature studies include a method of map matching by GPS positioning results of automobiles, automobiles can estimate their approximate absolute position by GPS and drive only as specified on the road. In our study, the camera pose estimation by visual odometry and the point cloud generation from stereo images are processed in parallel, and the camera pose estimation and posture during cornering are corrected by the results of the camera pose estimation and posture before and after cornering (Figure 2). A point cloud is generated from the clear stereo images before and after the blurred image. The correction is then reflected in the camera pose estimation and posture of the visual odometry by 3D map matching.

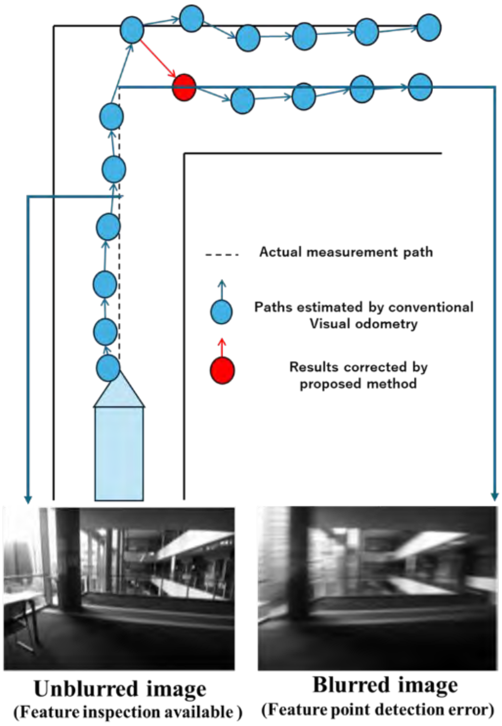

Actual measurement path
Paths estimated by conventional Visual odometry
Results corrected by proposed method
Unblurred image
(Feature inspection available)
Blurred image
(Feature point detection error)

Figure 2: 3D map matching conceptual diagram

d. Point cloud generation by semi-global matching and point cloud segmentation based on Euclidean distance

We describe our proposed semi-global matching for point cloud generation and point cloud segmentation based on Euclidean distance. Semi-global matching is a stereo matching algorithm to obtain a disparity map from two image pairs (left camera image and right camera image). The advantages of using semi-global matching include its resistance to noise and its ability to smoothly determine distances even for objects with indistinct contours. A disparity map is a process over the entire image using the disparity, i.e. The positions where the object is misaligned between the left and right images. From this disparity map, the distance per pixel can be calculated. The Euclidean distance is a method to measure the distance between two points by applying the Pythagorean theorem to find the shortest linear distance. In our study, the disparity map is generated by semi-global matching within the depth thresholds, and point clouds matching the thresholds are generated from the disparity map. Then, the generated semi-global matching point cloud is segmented into clusters by calculating the modal frequency of labels by the minimum Euclidean distance between points of different clusters.

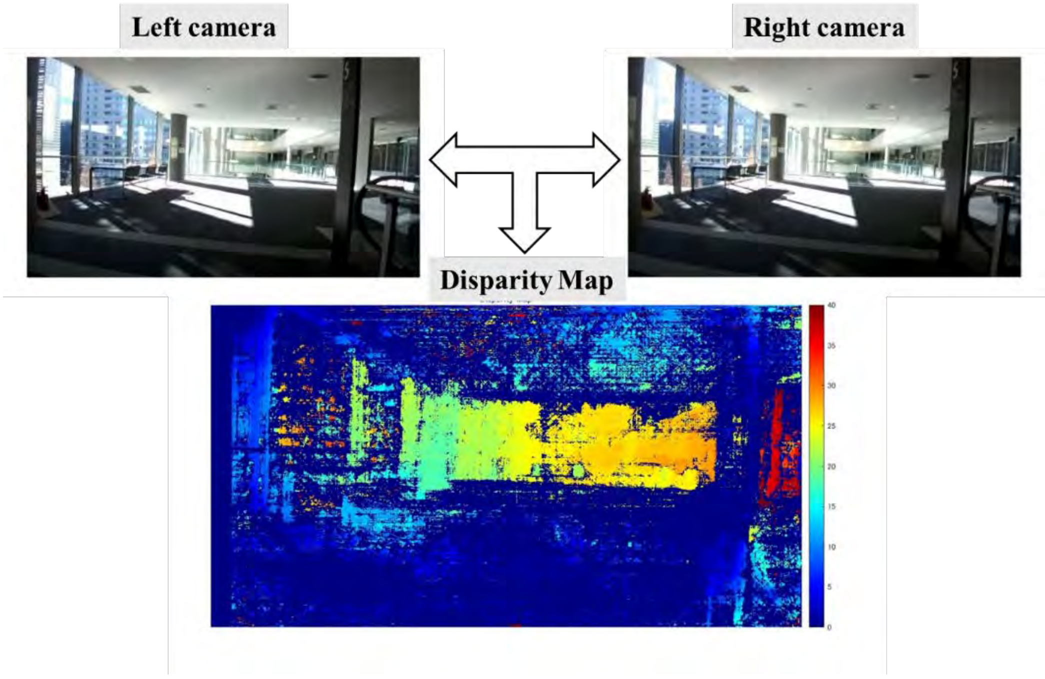


Figure 3: Stereo image and disparity map (left image and right image)

e. Point cloud matching with ICP

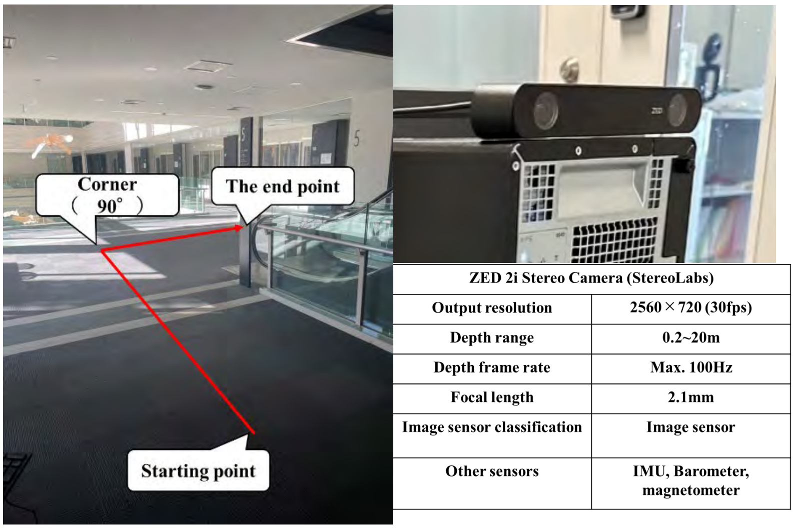
The ICP algorithm is a method for aligning two point clouds by iterative computation. The nearest neighbor points between the reference and input point clouds are called matching points, and the matching point is the point where the distance between the corresponding points is calculated to be the shortest. Then, the two point clouds are aligned by repeating the process between mapping of the two point clouds and estimating the translation vector and rotation matrix. In our study, the point cloud matching process by ICP is used as a specific method of 3D map matching. The point clouds from disparity images by the semi-global matching process for each frame in a time series are applied the ICP algorithm process between frames.

f. Camera pose correction processing

The location of the corner is identified by the number of feature points as the corner error. Next, the camera pose estimation of visual odometry is corrected by the camera pose estimation process calculated by the point cloud matching in 3D map matching at the frame number identified as the error. After the correction, the camera pose estimation of visual odometry is performed using the translation vector and rotation matrix.

Experiment

We conducted a walking measurement experiment on the 5th floor of the classroom building on our campus (Toyosu Campus) as an indoor experimental site (Figure 4). The measurement line consisted of a straight section of 4.3 [m] from the starting point to a corner, and a straight section of 4.3[m] from the corner to the end point. In the experiment, we acquired temporal stereo images at 30 [fps] with a speed of 1.5 [m/s] using an IMU stereo camera (ZED 2i, Stereolabs) connected to a desktop PC (Intel Core-i7, 2.50GHz). The stereo camera was mounted a forward direction along the direction of motion. The stereo camera was also rotated 90 degrees horizontally at the corner to represent a turn and measurement by an autonomous mobile robot or an indoor drone. We evaluated the proposed method in the walking measurement experiment. We also evaluated the horizontal deviation between the actual path and the camera trajectory data estimated by visual odometry and 3D map matching using point clouds before and after the corner. The position of the corner was determined from the images in each scene. In visual odometry, self-position estimation was started to match the length of the experimental path. The conditions for position estimation by visual odometry in this study assumed an autonomous mobile robot and a UAV. Although the position estimation by visual odometry included accumulated errors and spike errors due to motion blur at the corner, the estimated trajectory was recorded as 3D position data without trajectory rectification. Self-position estimation by 3D map matching was performed by generating point clouds before and after the turn. During the experiment, the walking speed was kept constant along the straight path, and we stopped to turn the camera at the corner. At the corner, the image was blurred during the turn, thus, feature points were not detected and tracked in the images.



ZED 2i Stereo Camera (StereoLabs)	
Output resolution	2560 × 720 (30fps)
Depth range	0.2~20m
Depth frame rate	Max. 100Hz
Focal length	2.1mm
Image sensor classification	Image sensor
Other sensors	IMU, Barometer, magnetometer

Figure 4: Experimental site and equipments

Results

a. Stereo camera calibration results

The results of the stereo camera calibration result are shown in Figure 5 and Figure 6. and the calculation of the re-projection error is shown in Figure 6. Figure 5 shows no obvious errors in the visualization of the external parameters. Figure 6 shows the reprojection errors. Although the 12th stereo pair had the maximum error value, the overall average reprojection error was 0.81 [pixel].

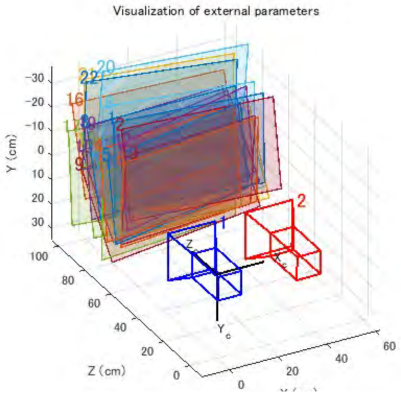


Figure 5: Stereo camera calibration result

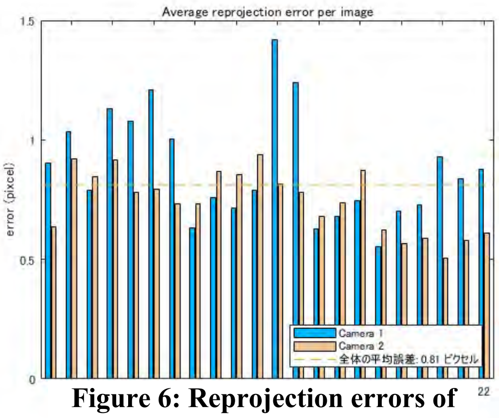


Figure 6: Reprojection errors of stereo camera calibration

b. Results and discussion of relative mobility

The results of the visual odometry processing are shown in Figure 7. The pre-corner trajectory was within approximately 0.9[m] of the X-axis deviation and within approximately 0.1[m] of the Y-axis deviation. The accuracy pre-corner of the motion measurement before turning was approximately 0.3[m], and the accuracy of the movement measurement after the turn was approximately 1.0[m].

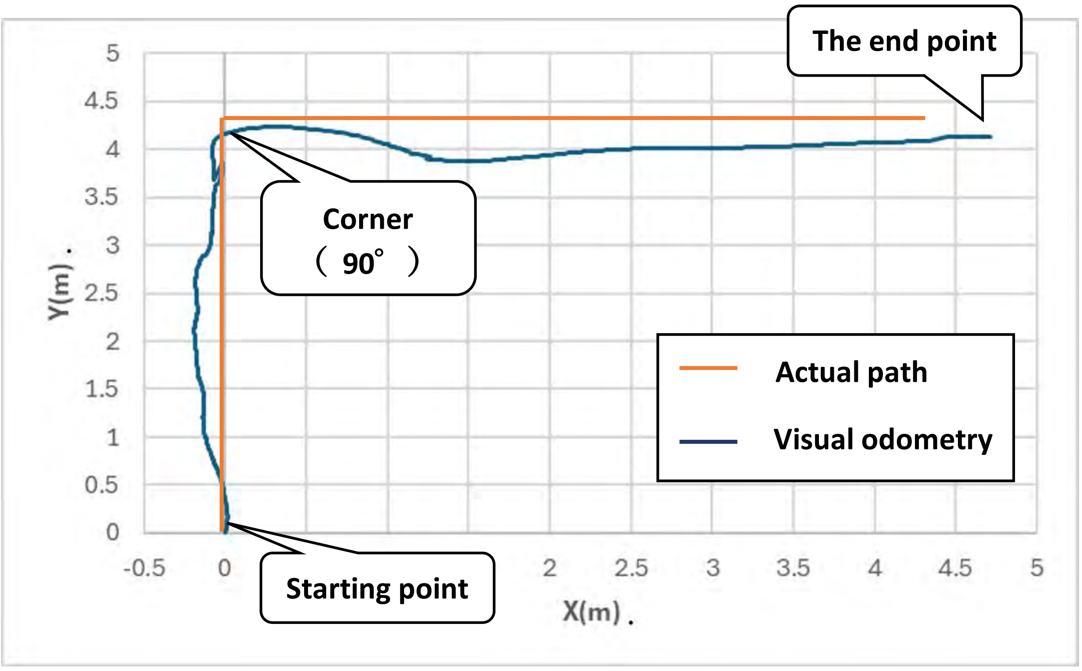


Figure 7: Visual odometry results

Figure 8 shows the amount of movement between frames. The maximum displacement was 0.18[m] and the minimum was 0.00[m]. Although the motion speed was 1.5 [m/s] (0.05 [m/frame]), there were several spike noises.

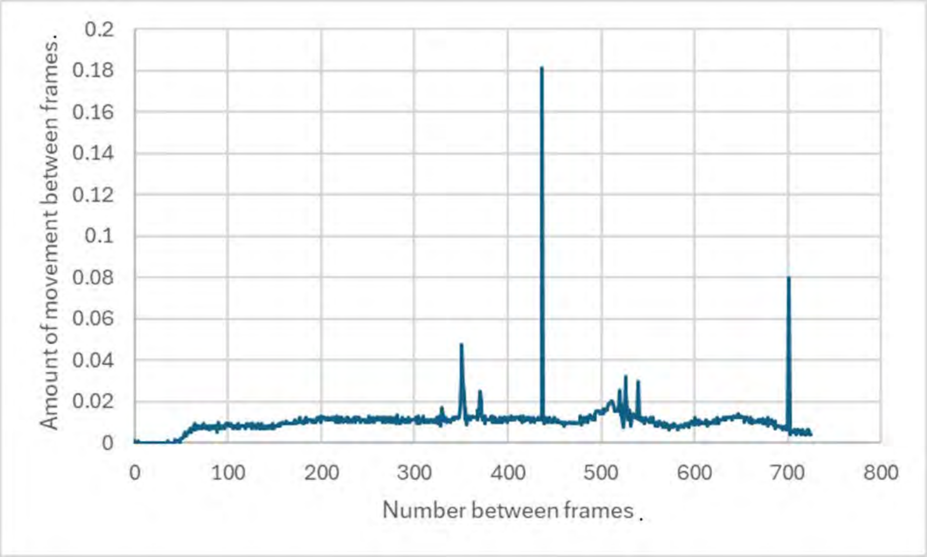


Figure 8: Amount of motion between frames

Figure 9 shows the directional vectors of the camera rotation parameters.

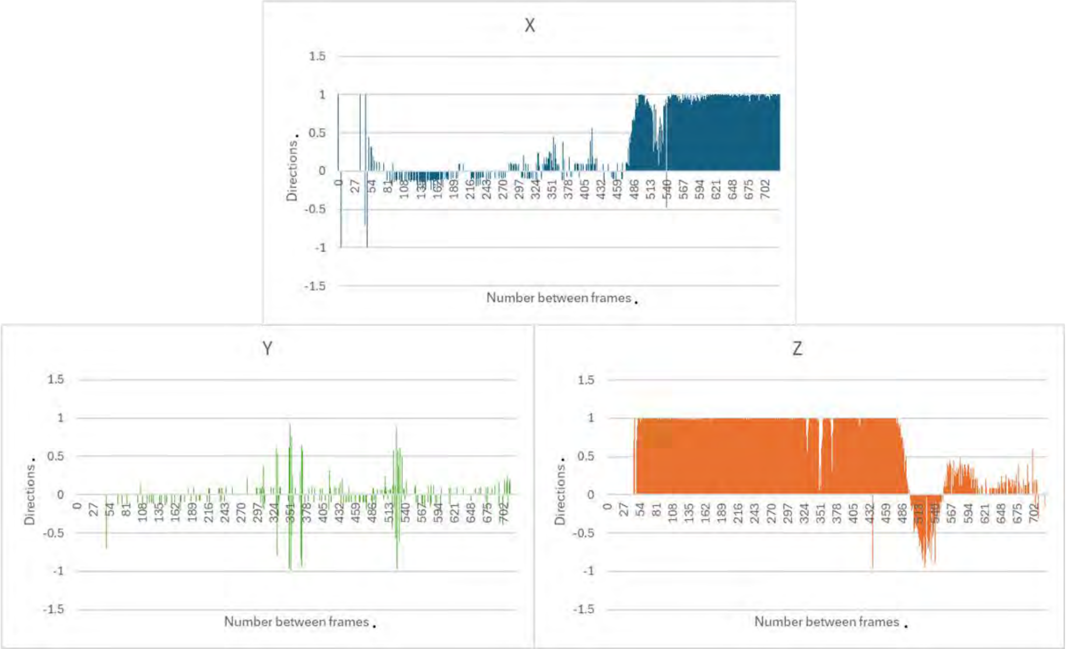


Figure 9: Camera rotation vectors

c. Point cloud generation based on Euclidean distance

Figure 10 shows the results of processing based on Euclidean distance based processing for the point clouds generated from the stereo images acquired in the experiments. We confirmed that noise around the measured objects can be removed while preserving the geometry of the features.

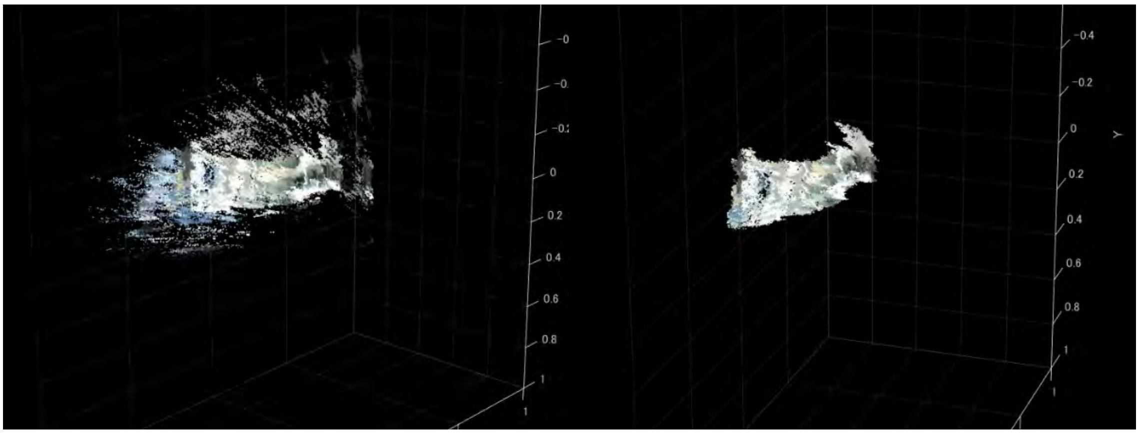


Figure 10: Generated point clouds (left: all acquired point clouds, right: filtered point clouds)

d. Corner detection from temporal images

Figure 11 shows the result of the corner detection. The number of feature points decreased from frame number 400, which coincided with the position where blur appeared in the image.

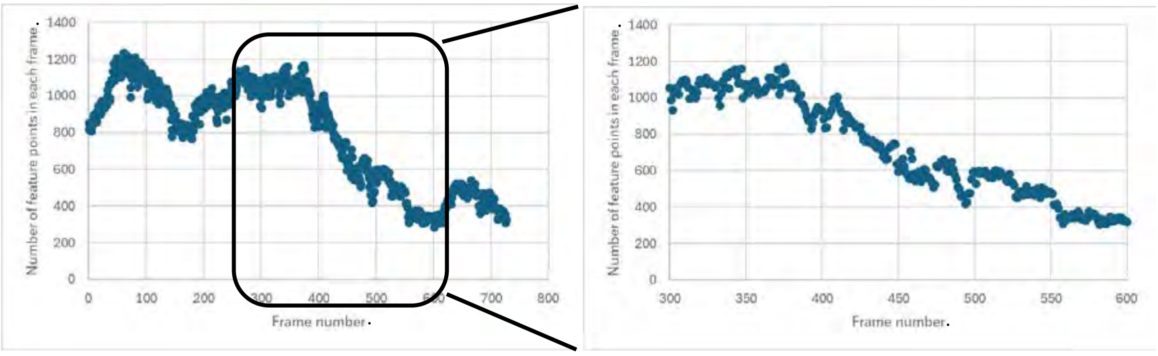


Figure 11: Corner detection result

e. Correction processing result

The results of applying self-position estimation by 3D map matching are shown in Figure 10. The orange line shows the actual path and the blue line shows the correction processing result. The maximum error on the X-axis was approximately 1.3 [m], and the maximum error on the Y-axis was approximately 0.1 [m].

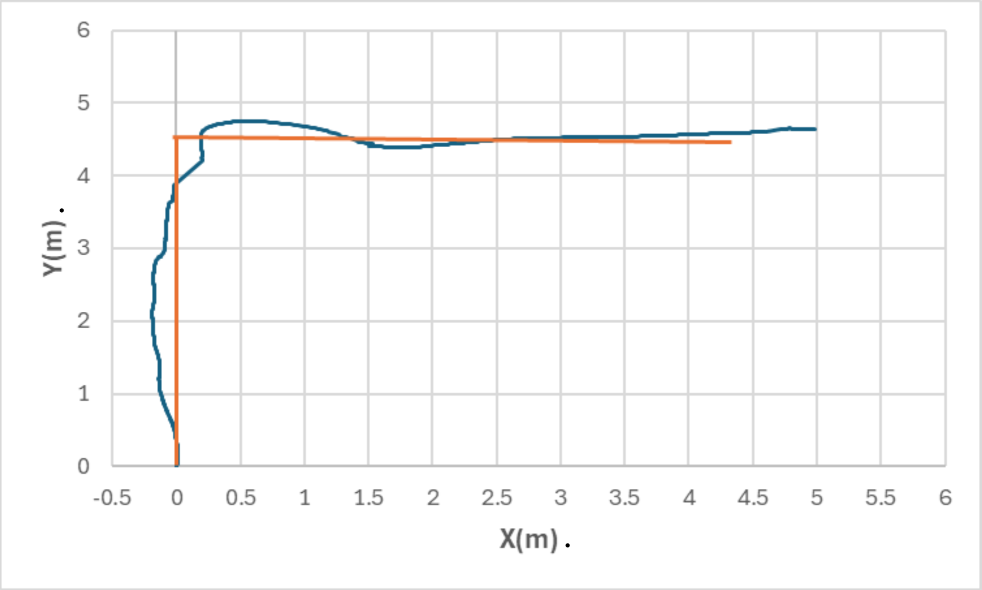


Figure 12: Correction result

Discussion

a. Corner detection from temporal image

As shown in Figure 11, it was confirmed that the proposed method could identify the location of the corner by judging the temporal images to be blurred. In addition to the frames with blurred images, there were other frames where the number of feature points decreased. It is assumed that this is because the measurement environment was in a dark environment.

As shown in Figure 9, the corner position was found to be in the range of 460 to 490. This difference is considered that the detection of the corner position was done by rough estimation based on the proposed corner detection. The reasons for this difference are also considered to be inaccurate feature point detection due to the dark environment and low accuracy of scale factor estimation due to inaccurate calibration.

b. 3D map matching processing

It was confirmed that camera pose estimation is possible for 3D map matching. Figure 13 shows the full point cloud and the segmented point clouds. The processing of the whole point clouds measured by the stereo camera was too high process, so the point cloud was segmented, but this resulted in the loss of useful shape features for registration. As a result, it caused the camera pose and orientation estimation failed.

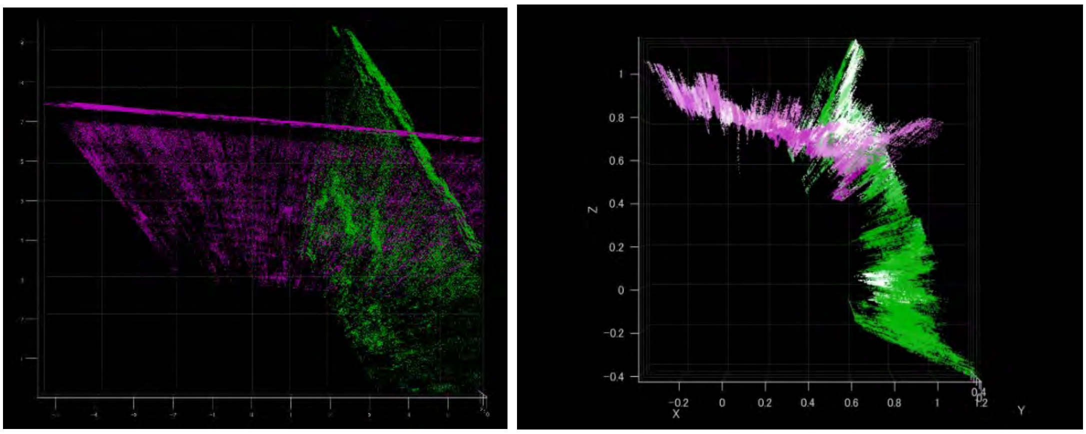


Figure 13: Frame shift before and after a sharp turn
(Left: all point cloud, right: segmented point cloud)

Next, we evaluated the processing time of the point cloud segmentation by the Euclidean distance. The number of frames was set to 725 [frames], and the time required for each 3D map matching process was evaluated. The 3D map matching process of all point cloud results was 212.159358 [sec]. The 3D map matching process of the point cloud segmented by Euclidean distance results was 269.423366 [sec]. This result confirms that point cloud segmentation by Euclidean distance can reduce the processing time. In our study, the developed 3D map matching resulted in a processing time of 212.16 [sec]. This can be attributed to be the large number of input point clouds for the ICP algorithm. In our study, camera pose estimation by 3D map matching and visual odometry are processed in parallel. There seems to be a high dependency on the 3D map matching process because the processing time of 3D map matching is much longer than that of visual odometry. Therefore, as a future improvement method, it is considered that the algorithm construction to apply camera pose estimation by 3D map matching when a corner position is detected during the processing of visual odometry will lead to a reduction in the processing time of 3D map matching.

The maximum error in the X-axis was approximately 1.0[m], and the maximum error in the Y-axis was 0.1[m]. Therefore, we considered that this was caused by the light/dark environment. Thus, we can focus on an improvement method for environments where it is difficult to acquire feature points in light/dark environments.

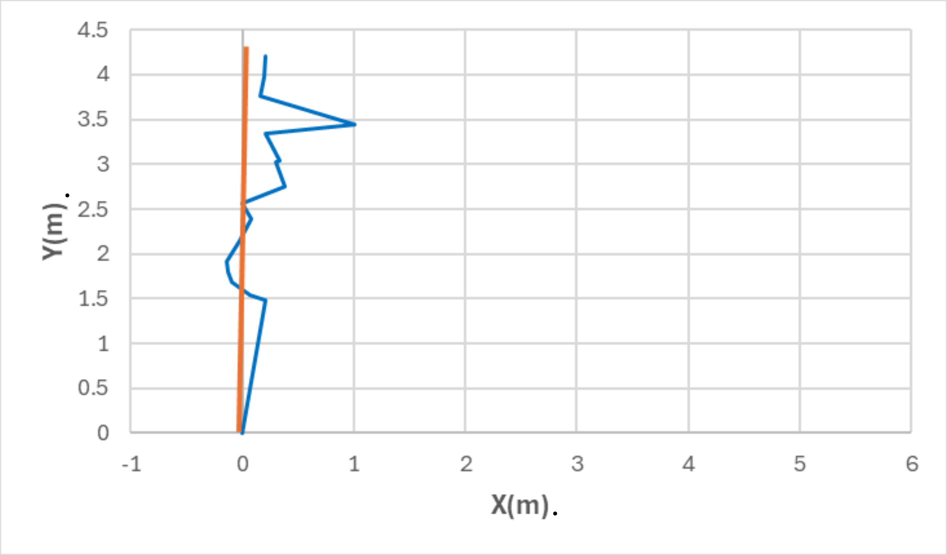


Figure 14: Self-position estimation by 3D map matching up to the position of a sharp turn

Conclusion

In our study, we focused on 3D map matching in the case of sudden turns in visual odometry. We proposed a method to compensate for errors in visual odometry. Through our experiments, we confirmed that our methodology can estimate and correct the camera pose with visual odometry and 3D map matching. We confirmed that our methodology with point cloud segmentation based on Euclidean distance can reduce the processing time. In addition, we found that this system is highly dependent on the processing speed of 3D map matching. As our future work, our methodology will be improved to achieve higher processing speed and accuracy in 3D map matching, because the current 3D map matching requires too long processing time for real-time processing, and the processing time of 3D map matching is longer than that of visual odometry. We focus on parallel processing using GPU of camera pose estimation and 3D map matching to improve the processing speed. As an improvement plan, we focus on applying 3D map matching only at the corner. We also focus on other technical issues such as robustness to the lighting environment and motion blur at corners.

Reference

Z. Zhang, (2000). A flexible new technique for camera calibration, IEEE, pp.1330-1334.

Lucas R, Nuno M Ricardo, Maria I. Pereira, Antoine Hiolle, Andry M. Pinto, (2022). A Practical Survey on Visual Odometry for Autonomous Driving in Challenging Scenarios and Conditions, IEEE, pp.72182-72205.

A Handa,T Ehelan, J McDonald,A, (2014). Benchmark for RGB-D Visual odometry, 3Dreconstruction and SLAM, IEEE, pp.1050-4729.

IlmirZ.Ibragimov, IlyaM.Afanasyev, (2017). Comparison of ROS-based visual SLAM methods in homogeneous indoor environment, 14th Workshop on Positioning, Navigation and Communications(WPNC).

Zachary Ted, Lahav Lipson, Jia Deng, (2023). DEEP Patch Visual Odometry, NeurlPS.

Zachary Teed, Jia Deng, (2021). DROID-SLAM: Deep Visual SLAM for Monocular, Stereo, and RGB-D Cameras, Zachary Teed, Jia Deng, NeurlPS.

Chaolei, W., Tianmiao, W., Jianhong, L., Yang, C., Yicheng, Z., Cong, (2012). Monocular visual SLAM for small UAVs in GPS-denied environments, IEEE.

Kazuha Saito, Gai Ozaki, Kenta Okudaira, Masafumi Nakagawa, (2022). Performance Verification Of Visual Odometry with IMU-Stereo Camera in Indoor UAV, The 16th South East Asian Technical University Consortium Symposium, pp.27-30.

AlexanderKern, MarkusBobbe, Yogesh Khedar, (2020). Real-time Mapping for Unmanned Aerial Vehicles, AlexanderKern, MarkusBobbe, Yogesh Khedar, Ulf Bestmann International Conference on Unmanned Aircraft Systems(ICUAS).

Ke, S, Kartik, M, Bernd, P, Michael, W, Sikang, L, and Yash, M, (2018). Robust Stereo Visual Intertial Odometry for Fast Autonomous Flight, IEEE, pp.965-972.

Faragher, R, M., Harle, R, K, Nashville, (2013). SmartSLAM - An Efficient Smartphone Indoor Positioning System Exploiting Machine Learning and Opportunistic Sensing, pp.1006-1019.

Kazuha Saito, Masafumi Nakagawa, Yusuke Kawasaki, Masaaki Takebayashi, Masafumi Miwa, (2024). Verification of Indoor-outdoor Seamless Positioning and Camera Positioning Rectification Systems Mounted on Infrastructure Inspection UAVs, The 18th South East Asian Technical University Consortium Symposium, pp.81-84.

Zachary Teed, JiaDeng, (2020). Video To Depth With Differentiable Structure From Motion, ICLR, DEEPV2D, Computer Science Center for Statistics & Machine Learning Princeton Language and Intelligence (PLI).