ACRS 2022
MONGOLIA

EFFICIENT IMAGE MATCHING USING IMU STEREO CAMERA IN SfM/MVS

Masafumi Nakagawa, Yuichiro Yamaguchi
Shibaura Institute of Technology, 3-7-5, Toyosu, Koto-ku, Tokyo 135-8548, Japan,
Email: mnaka@shibaura-it.ac.jp

KEY WORDS: SfM/MVS, Visual odometry, IMU stereo, Geometrical camera network

ABSTRACT: In ground-based structure from motion and multiview stereo (terrestrial SfM/MVS) and indoor mapping, the number of images acquired tends to increase because of several restrictions such as field of view and camera distances. The increase in the number of images drastically affects the processing cost of feature extraction and image matching in SfM/MVS. Conventional solutions for this technical issue are camera parameter estimation with precise global navigation satellite system and inertial measurement unit (GNSS/IMU) data and image acquisition planning with space division based on sparse-dense blocks. However, in indoor environments, GNSS/IMU does not work. Moreover, the space division is complex in sparse-dense block preparation. Thus, we focus on visual odometry processing to improve the efficiency of image acquisition for SfM/MVS. Through experiments on SfM/MVS with several datasets, we confirmed that our proposed methodology can reduce the processing time of image matching.

1. INTRODUCTION

In open sky environments, structure from motion and multiview stereo (SfM/MVS) based on aerial photogrammetry using UAVs is an efficient point cloud acquisition approach. Moreover, when we use images acquired from various viewpoints such as ground and sky together, the SfM/MVS can improve the completeness of point cloud generation. However, in ground-based SfM/MVS (terrestrial SfM/MVS) and indoor mapping, camera distances in terrestrial image acquisition are shorter than those in aerial image acquisition. Therefore, the number of images acquired tends to increase because of several restrictions such as field of view and camera distances. As a technical issue of terrestrial SfM/MVS, the increase in the number of images drastically affects the processing cost of feature extraction and image matching before a camera parameter estimation. Two types of conventional solutions exist for this technical issue. The first solution is a direct camera parameter estimation with precise global navigation satellite systems and inertial measurement unit (GNSS/IMU) data. Direct camera parameter estimation using GNSS/IMU is a popular approach to improve the performance of SfM/MVS. However, in indoor environments, the approach is difficult to be applied to SfM/MVS because GNSS/IMU does not work. The second solution is image acquisition planning with space division based on sparse-dense blocks (Nakagawa et al., 2017). In the approach, high spatial resolution images and point clouds acquired in dense blocks are managed with corresponding points extracted from low spatial resolution images in sparse blocks. However, the space division is complex in sparse-dense block preparation. Thus, in this research, we focus on visual odometry processing (Saito et al., 2022) to improve the efficiency of image acquisition for SfM/MVS. First, we generate a matrix to represent corresponding images with a geometrical camera network. Second, we reduce the processing time of image matching in SfM using the generated matrix. In our experiments, we used a high-resolution camera and IMU stereo camera with SiftGPU, multicore bundle adjustment, and, patch-based MVS. Through experiments on SfM/MVS with several datasets, we confirm that our proposed methodology can reduce the processing time of image matching.

2. METHODOLOGY

Figure 1 shows our proposed methodology. First, high-resolution images and visual odometry data are acquired simultaneously. Second, feature points are extracted from high-resolution images. Third, the image corresponding matrix is generated from a geometrical camera network using acquired visual odometry data. Fourth, corresponding points are extracted from high-resolution images with image corresponding matrix estimated using visual odometry data. Finally, dense point clouds are generated. Although the basic part of the SfM/MVS processing flow is almost the same, it differs from the conventional SfM/MVS processing in the use of the correspondence matrix between images generated from the visual odometry data. The correspondence matrix is a sparse matrix to visualize the geometrical camera network between images (Figure 2) by arranging reference images in rows and corresponding images in columns. The geometrical camera network between images is generated based on a stereo pair model with image overlaps and each calculated camera's line-of-sight vector with the position and orientation data estimated by visual odometry processing. In the geometrical camera network, camera positions are described as nodes, and stereo pairs are described as links.

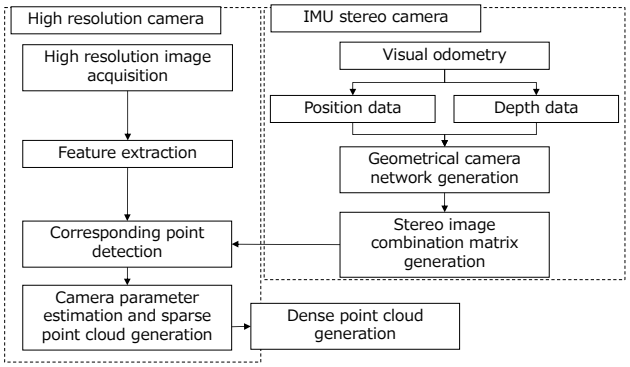


Figure 1. Proposed methodology

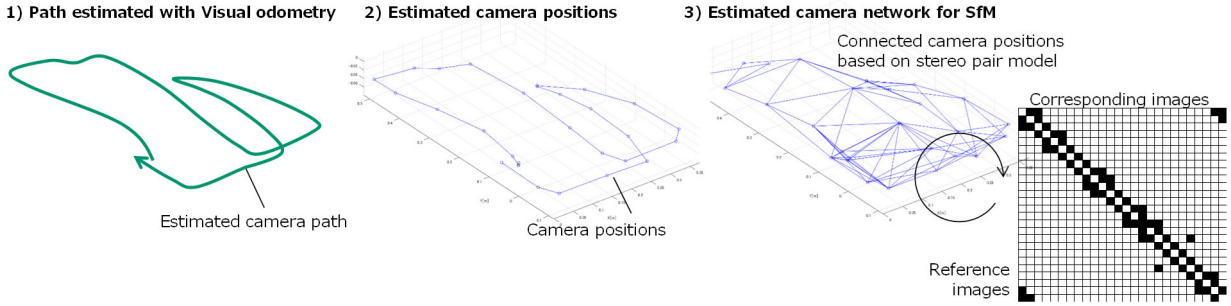


Figure 2. Stereo image combination and geometrical camera network generation

Visual odometry is an image-based position and rotation estimation methodology with cameras Although visual odometry with a single camera is a basic approach, various visual odometry methodologies have been developed, such as an approach using a stereo camera and an approach using a single camera with an IMU, to improve the accuracy of position and orientation estimation. In this study, we focus on a combination of stereo camera and IMU (IMU stereo camera) (Taragay et al., 2007) to estimate more stable position and rotation data. In visual odometry processing using an IMU stereo camera, we apply a methodology to improve the redundancy of camera position and orientation estimation with accumulated error adjustment with loop closure detection and image-based relocalization processing (Saito et al. 2022). As a result, geometrical network distortion caused by accumulated errors in visual odometry processing can be almost ignored, and precise information on stereo image combinations can be estimated using the geometrical camera network.

When an IMU stereo camera is used for our proposed methodology, depth images generated by stereo image matching can be integrated with camera position and rotation data estimated by visual odometry to improve the precision of image combination for SfM/MVS. However, in terrestrial SfM/MVS, distances from the camera to measured objects are almost the same because of closed-range image acquisition. Thus, depth images are not effective data for our proposed methodology in terrestrial SfM/MVS when we preset a fixed parameter as a distance value from a camera to measured objects. Therefore, only position and rotation data estimated with visual odometry are used for our methodology to simplify the correspondence network generation between images with overlapped image estimation, as shown in Figure 3. First, the positions and directions of all cameras are estimated by visual odometry. Next, stereo combinations (i.e. overlapped images) are estimated using distances and line-of-sight angles among cameras. Then, estimated stereo combinations are filtered with a length of the stereo baseline. Although a short baseline can provide precise image matching, 3D measurement accuracy is low. Alternatively, although a long baseline can achieve higher 3D measurement accuracy, the stability of image matching is low. Based on these ideas, stereo combinations are estimated.

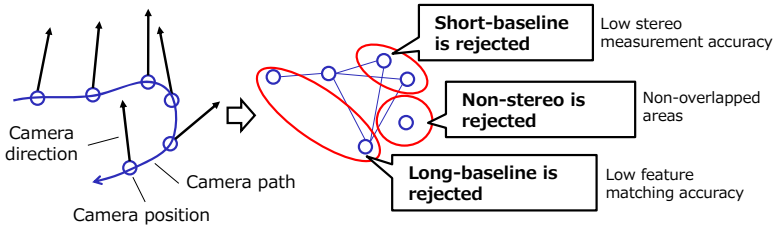


Figure 3. Detection of stereo combinations

3. EXPERIMENTS

We measured six scenes and objects, such as concrete walls, roads, stairs, and trees, assumed as terrestrial SfM/MVS and low-altitude aerial UAV surveying for infrastructure management, indoor mapping, and cultural heritage recording. Moreover, we acquired images using handheld cameras with three types of camera paths, shuttling, gazing, and free-moving paths. Then, the processing times for image matching and point cloud generation were compared to evaluate our proposed methodology. The imaging system consisted of a high-resolution camera (DSC-RX0M2, SONY) mainly for point cloud generation and an IMU stereo camera (RealSense T265, Intel) for visual odometry processing (Figure 4). All cameras were synchronized with PC time. Although two high-resolution cameras were used to increase the number of images in the experiment, only the left camera was used for data processing. Interval image acquisition (1 Hz) was applied to the high-resolution cameras. In SfM/MVS processing, all images of the left camera were used including images containing blur caused by rapid moving and rotation. Moreover, all images were resized into 2400 x 1600 pixels as input data for the SfM/MVS. We applied a generation of stereo image combination matrix (MATLAB/Python), SiftGPU, multicore bundle adjustment, and patch-based multi-view stereo (PMVS).

Figure 4. Data acquisition system

4. RESULTS

Table 1 shows the six sets of experimental results. Blue lines in camera paths show three types of image acquisition paths, such as shuttling, gazing, and free-moving paths. The black areas in matching matrices indicate the matching results between the estimated images.

Table 1. Processing results in point cloud generation and image matching

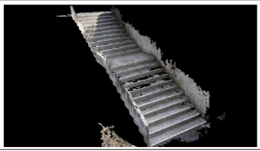
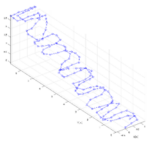
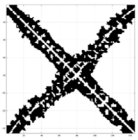
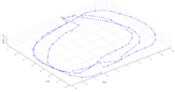
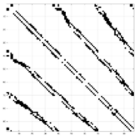
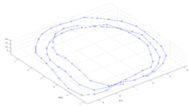
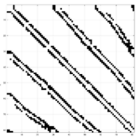
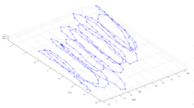
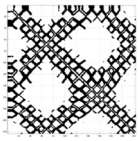
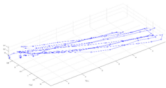
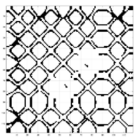
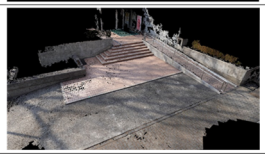
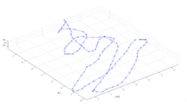
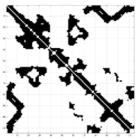
ID	Generated point clouds	Camera path	Matching matrix
1			
2			
3			
4			
5			
6			

Table 2 shows an overview of data processing and the processing time (processing environment: Core i7-11370H, 3.3 GHz). Processing time for image input and lens distortion rectification is summarized as "others" in each result. In addition, the processing time of the stereo image combination and geometrical camera network generation was excluded from the table because it was within 1 second for each. Table 2 also shows that the number of output point clouds was almost the same between the conventional methodology and the proposed methodology. This result indicates that our methodology kept the quality of point clouds. Moreover, Table 2 shows the proposed methodology reduced the number of pairs in the SfM processing. Thus, the processing time for sparse point cloud generation improved drastically, as shown in "image matching."

Table 2. Processing result overview and processing time

ID: 1	Results (conventional)	Results (proposed)
The number of input images	145 [images]	
The number of searched pairs	10440 [pairs]	2864 [pairs]
Processing time (Total)	5488 [sec]	4419 [sec]
Feature detection	26 [sec]	25 [sec]
Image matching	1344 [sec]	318 [sec]
Dense pc generation	3954 [sec]	3943 [sec]
Others	164 [sec]	133 [sec]
The number of point clouds	6,044,488 [pts]	6,033,970 [pts]

ID: 2	Results (conventional)	Results (proposed)
The number of input images	97 [images]	
The number of searched pairs	4656 [pairs]	369 [pairs]
Processing time (Total)	2574 [sec]	2275 [sec]
Feature detection	18 [sec]	17 [sec]
Image matching	559 [sec]	42 [sec]
Dense pc generation	1905 [sec]	2131 [sec]
Others	92 [sec]	85 [sec]
The number of point clouds	3,135,355 [pts]	3,427,755 [pts]

ID: 3	Results (conventional)	Results (proposed)
The number of input images	81 [images]	
The number of searched pairs	3240 [pairs]	374 [pairs]
Processing time (Total)	2536 [sec]	2048 [sec]
Feature detection	16 [sec]	15 [sec]
Image matching	357 [sec]	47 [sec]
Dense pc generation	2101 [sec]	1930 [sec]
Others	62 [sec]	56 [sec]
The number of point clouds	2,329,981 [pts]	2,119,187 [pts]

ID: 4	Results (conventional)	Results (proposed)
The number of input images	223 [images]	
The number of searched pairs	24753 [pairs]	7512 [pairs]
Processing time (Total)	8236 [sec]	6423 [sec]
Feature detection	34 [sec]	41 [sec]
Image matching	2469 [sec]	775 [sec]
Dense pc generation	5481 [sec]	5381 [sec]
Others	252 [sec]	226 [sec]
The number of point clouds	10,104,306 [pts]	9,766,795 [pts]

ID: 5	Results (conventional)	Results (proposed)
The number of input images	118 [images]	
The number of searched pairs	6903 [pairs]	1378 [pairs]
Processing time (Total)	2915 [sec]	3208 [sec]
Feature detection	20 [sec]	20 [sec]
Image matching	716 [sec]	161 [sec]
Dense pc generation	2080 [sec]	2937 [sec]
Others	99 [sec]	90 [sec]
The number of point clouds	1,990,045 [pts]	2,982,274 [pts]

ID: 6	Results (conventional)	Results (proposed)
The number of input images	114 [images]	
The number of searched pairs	6441 [pairs]	1440 [pairs]
Processing time (Total)	4567 [sec]	3374 [sec]
Feature detection	20 [sec]	20 [sec]
Image matching	696 [sec]	167 [sec]
Dense pc generation	3726 [sec]	3374 [sec]
Others	125 [sec]	124 [sec]
The number of point clouds	6,928,877 [pts]	6,231,082 [pts]

5. SUMMARY

We proposed a methodology to improve the efficiency of image matching for terrestrial SfM/MVS with visual odometry processing using an IMU stereo camera. Through experiments with several datasets on SfM/MVS using a high-resolution camera and an IMU stereo camera, we confirmed that our proposed methodology can reduce the processing time of image matching in SfM. Although the processing performance of SfM was improved drastically using a geometrical camera network, the processing time of MVS did not improve, because it mainly depends on the dense point cloud generation. However, the proposed methodology can improve the total processing time of SfM/MVS. Moreover, the proposed methodology can improve the efficiency of data acquisition with on-site confirmation of data acquisition for SfM/MVS.

REFERENCES

Nakagawa, M., Miwa, K., Nozue, S., Sekiguchi, Y., Hirate, K., Noda, Y., Miyo, M., 2017. Efficiency Improvement of SfM using Image Blocks for Infrastructure Inspection. The 38th Asian Conference on Remote Sensing 2017, 9 pages.

Saito, K., Ozaki, G., Okudaira, K., Nakagawa, M., 2022. Performance Verification of Visual Odometry with IMU-Stereo Camera in Indoor UAV. South East Asian Technical University Consortium Symposium 2022, pp.27-30.

Taragay, O., Zhiwei, Z., Supun, S., Rakesh, K. 2007. Visual Odometry System Using Multiple Stereo Cameras and Inertial Measurement Unit. IEEE Conference on Computer Vision and Pattern Recognition, 8 pages.