EFFICIENCY IMPROVEMENT OF SfM USING IMAGE BLOCKS FOR INFRASTRUCTURE INSPECTION

Masafumi Nakagawa[1], Keiichi Miwa[1], Shoya Nozue[1],
Yasushi Sekiguchi[2], Katsuharu Hirate[3], Yasuaki Noda[4], Masahiro Miyo[4]
[1]Shibaura Institute of Technology, 3-7-5, Toyosu, Koto-ku, Tokyo 135-8548, Japan,
[2]Sumire survey Co., Ltd., 2-7-4, Shin-ohashi, Koto-ku, Tokyo 135-0007, Japan
[3]Marushigeya Corp., 57, Kanshuji Hirata-town, Yamashina-ku, Kyoto-city, Kyoto, Japan
[4]Watanabe Engineering Co., Ltd., 7-4-69, Noda-town, Fukushima-city, Fukushima 960-8055, Japan
Email: mnaka@shibaura-it.ac.jp

KEY WORDS: Structure from Motion, Infrastructure inspection, Image matching, Multiview stereo, Point cloud

ABSTRACT: We focused on the efficiency improvement of Structure from Motion (SfM) for infrastructure inspection using close- and long-range images taken from various viewpoints. We also focused on the camera network optimization using a matching matrix with ordered image blocks. Our methodology consists of three steps: image block preparation, key image-pair estimation using a matching matrix, and point-cloud reconstruction. We have confirmed that our proposed methodology can improve the efficiency of SfM through three preliminary experiments and four experiments in outdoor environments.

1. INTRODUCTION

Infrastructure asset inspection and management are based on a framework for achieving sustainable infrastructure, such as roads, bridges, railways, and water treatment facilities. In general, the inspection and management processes focus on low life-cycle cost during construction, maintenance, rehabilitation, and replacement. Based on this framework, a 3D geometric model must be prepared for building information modeling.

Reliability, completeness, efficiency, and cost are significant indices in monitoring. Thus, a 3D geometric model is often generated from point-cloud data using cameras and LiDAR to achieve reliability, completeness, and efficiency. Moreover, both large scale and high spatial resolution (less than 0.2 mm to recognize cracks) should be satisfied at the same time for 3D modeling in infrastructure inspection. Currently state, although LiDAR can acquire high resolution data, it is not easy to acquire details (less than 0.2 mm spatial resolution) of asset geometry and deterioration. Thus, we focus on cameras uing Structure from Motion (SfM) for ground-based investigation.

SfM is a methodology for reconstructing a scene using multiple cameras simultaneously from all available relative motions through key point detection, feature matching, motion estimation, triangulation, and bundle adjustment. SfM can be used for various 3D modeling and mapping tasks, such as aerial drone surveys, cultural heritage modeling, indoor mapping, and, infrastructure modeling. Although SfM is a useful methodology, large-scale SfM typically requires huge amounts of computational time fo pair-wise image matching and geometric verification to discover connected image components. Thus, many researchers have improved the efficiency of SfM when dealing with large image collections. Conventional research can be classified into three approaches. The first is computational environment improvement, such as General-Purpose computing on Graphics Processing Units (Irschara et al., 2010). The second improves image matching algorithms based on feature descriptors, such as Speeded Up Robust Features (SURF) (Bay et al., 2008) and Features from Accelerated Segment Test (Rosten et al., 2005). The third is camera network optimization such as graph-based image matching (Sweeney et al., 2015).

In this paper, we focused on improving the efficiency of SfM for infrastructure inspection using close- and long-range images taken from various viewpoints. We also focused on reducing the camera network with a matching matrix using ordered image blocks. There have been several related studies, such as graph optimization for camera pose reconstruction (Shen et al., 2016), a learning-based approach for efficient two-view geometry classification (Schönberger, et.al,, 2015) and sparse depth images at each camera to achieve translation averaging and essential matrix filtering (Zhaopeng, et.al, 2015).

Our methodology consists of three steps. First, blocked images are prepared. The blocked images consist of a main block of long-range images and several subblocks of close-range images. Second, feature point pairs between the main block and each subblock are extracted from key image pairs determined using a matching matrix with ordered image blocks. Finally, all camera poses and point cloud data are simultaneously reconstructed. Our experiments verify that our proposed methodology can improve the efficiency of SfM.

2. METHODOLOGY

In the conventional approach, a dense point cloud is generated using SfM with the following four steps, as shown in Figure 1. First, feature points are extracted from input images. Second, feature points are matched using image matching processing to estimate camera poses. Third, sparse point-cloud data are generated using the images and estimated camera poses. Finally, dense point-cloud data are generated using multiview stereo processing.

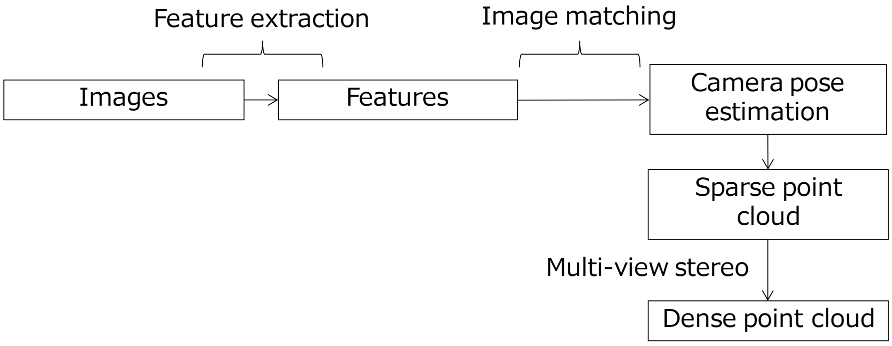

Figure 1. Conventional approach to generating dense point-cloud data from images

When input images consist of low resolution images covering a wide area and high-resolution images covering detailed areas, the images can be grouped into a main block to represent the outline of the measured object and subblocks to represent details of measured objects. When input images consist of the main block and subblocks, a combination of images in image matching can be determined based on a matching matrix, as shown in Figure 2. The row and column indicate input image numbers, and the combinations of row and column show image combinations in image matching. Grayed grids show the existence of actual image combinations. Normally, many actual image combinations exist in the main block and subblocks. On the other hand, other image combinations would be sparse between the main block and each subblock.

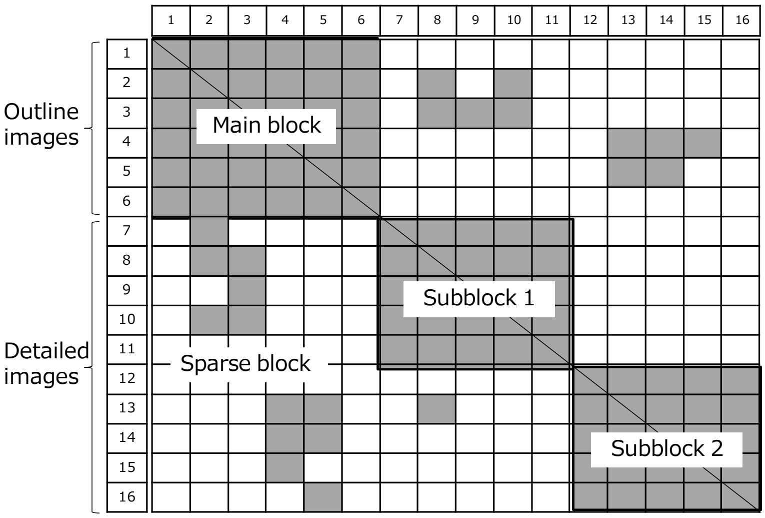

Figure 2. Image combinations in a matching matrix

In conventional SfM processing, corresponding feature points are estimated using all image combinations. Moreover, processing time for corresponding image detection depends on the number of input images. When n images are input, the number of image pair candidates is $n \times (n - 1)/2$. For example, when 16 images are input, there will be 120 image combination candidates. Although feature matching using all combinations of images can estimate precise camera poses, time-consuming processing is required for image matching and SfM processing. However, when the input images consist of a main block and subblocks, there are many useless results in image pair detection because there are several image combinations for image matching in each sparse block. Thus, we propose a methodology to find significant image combinations from the sparse block to improve the efficiency of image matching in SfM.

Although our overall methodology consists of feature extraction, image matching, sparse point-cloud generation, and dense point-cloud generation, as in conventional SfM, we improved the corresponding image detection block, as shown in Figure 3. First, the acquired images are grouped as a main block and subblocks. Second, feature points are extracted from all images. Third, images in each block are matched using extracted feature points. Fourth, key image pairs between the main block and each subblock are extracted using the extracted feature points. Finally, all corresponding feature points are used for camera pose estimation and dense point-cloud generation.

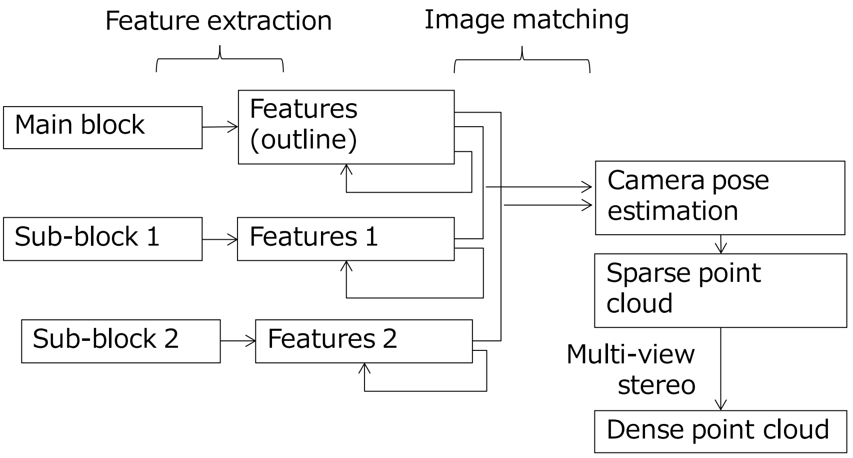

Figure 3. Proposed methodology (overall)

Our image matching procedure for key image-pair detection (see, Figure 4) consists of two parts. In the first part, image combination candidates are estimated in each block. When each block has unsorted images, a round robin approach is applied to estimate image combination candidates.

The second part consists of two steps. First, several images are selected from each subblock to generate sampling lines in a matching matrix. The sampling lines are grids that indicate image pairs in a matching matrix. These several images are determined using the number of total corresponding points of each image in a subblock, based on the assumption that a representative image includes many image pairs in a block.

Second, several images are selected from the main block to determine images corresponding with the sampled images in each subblock along the sampling line. These several images are sampled from the main block using the number of matching points in each image, based on the assumption that a representative image in a main block includes many corresponding points in subblock images.

Based on these steps, the number of image combination candidates between the main block and each subblock is reduced in the matching matrix.

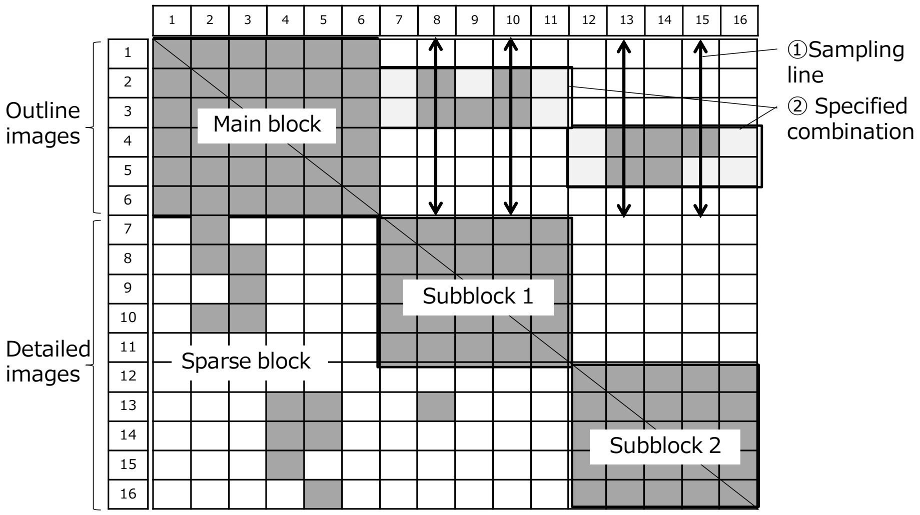

Figure 4. Reduction of image-combination candidates using a matching matrix

3. EXPERIMENTS

3.1 Preliminary simulation experiments

We conducted three preliminary simulation experiments to evaluate our methodology. The first simulation evaluated the effect of the size of the main block on the reduction of image combination candidates. Figure 5 shows a comparison between the ratio of the number of images in a main block to the number of all images when the number of subblocks was 10 and the number of sampling lines was three. The figure also shows that smaller main blocks reduce the number of image combination candidates nonlinearly. For example, when a main block occupies 60% of a matching matrix, the figure shows that the number of image combination candidates can be reduced to about one half.

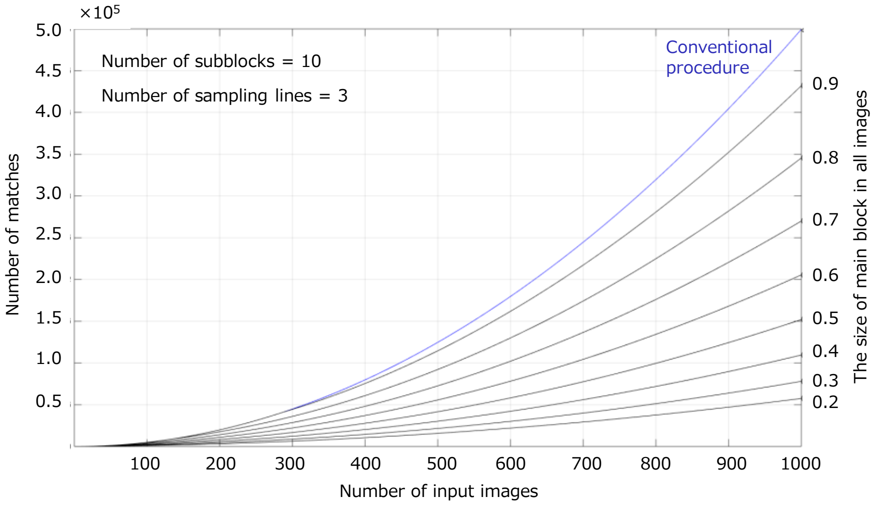

Figure 5. Comparison between the sizes of main blocks in all images

The second simulation evaluated the effect of the number of subblocks on the reduction of image combination candidates. Figure 6 shows a comparison between the number of subblocks and the number of matches when the main block constituted 40% of the total images and the number of sampling lines was one. The figure also shows that an increase in the number of subblocks nonlinearly reduces the number of image combination candidates. For example, with two subblocks exist, the figure shows that the number of image combination candidates can be reduced to about one-third of the initial candidate.

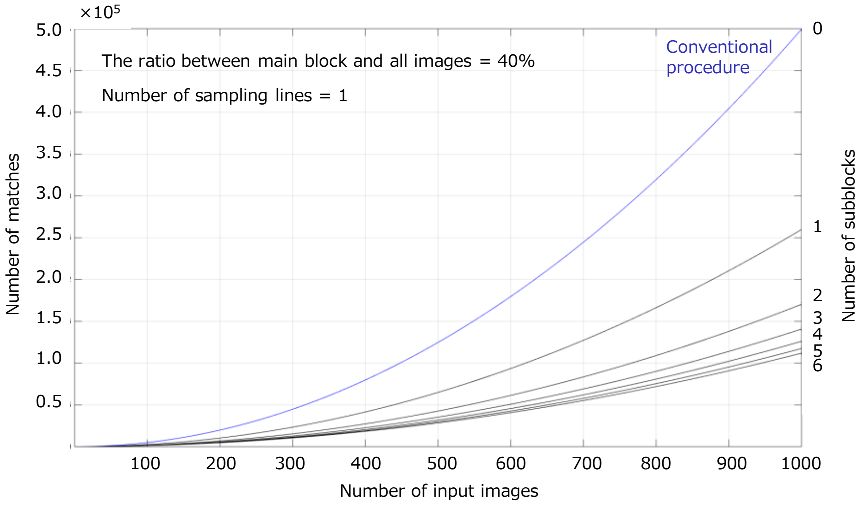

Figure 6. Comparison between the numbers of subblocks

The third simulation evaluated the effect of the number of sampling lines in each subblock on the reduction in image combination candidates. Figure 7 shows a comparison between the number of sampling lines and the number of matches when the main block constituted 40% of the total images and the number of detailed subblocks was 10. The figure shows that fewer sampling lines reduces the number of image combination candidates.

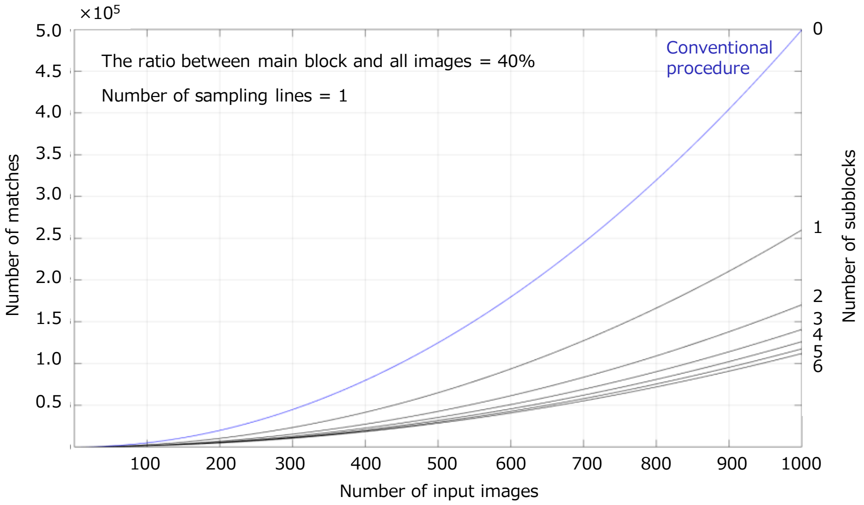

Figure 7. Comparison between the numbers of sampling lines

3.2 Experiments in outdoor environments

A handheld digital camera (DMC-LF1, Panasonic) was used for all experiments. We applied SURF to extract feature points and matching points from images. We also used Bundler (Snavely, 2010) and Patch-based Multiview Stereo (Furukawa et al., 2010) libraries with a notebook PC (Intel Core i7-6567U, 3.30 GHz) to reconstruct camera poses and point-cloud data. We selected four objects: a revetment, a concrete bench, and two retaining walls showing some deterioration, as shown in Figure 8. Four datasets for each object were prepared and assumed to be 3D measurement in infrastructure inspections. Acquired images in each dataset were manually grouped into a main block and several subblocks, as shown in Table 1.

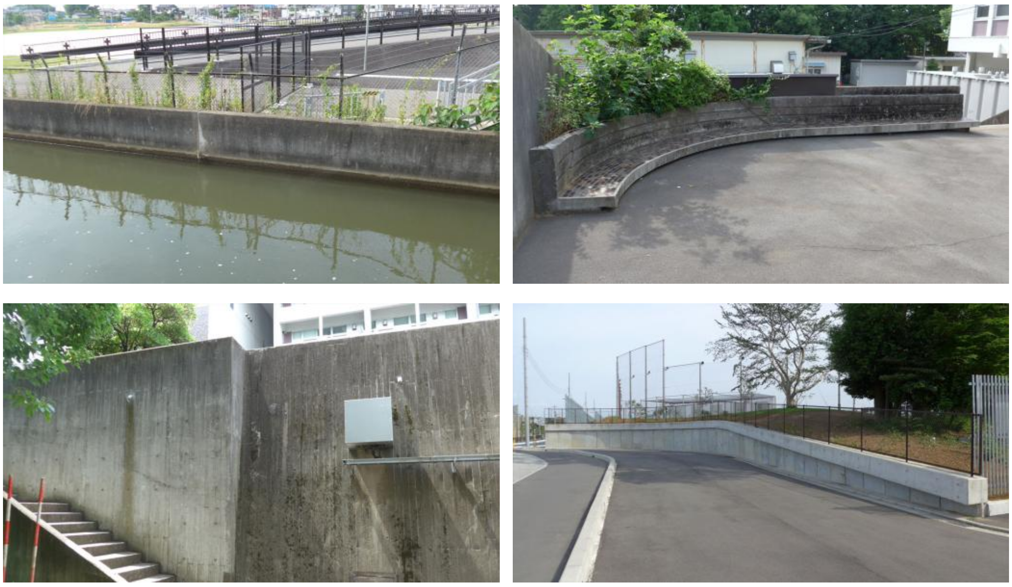

Figure 8. Measured objects (upper left: revetment; upper right: concrete bench; lower left: retaining wall (1); lower right: retaining wall (2))

Table 1. Input images

Dataset	Measured object	The number of input images	The number of main block	The number of sub-blocks
1	Revetment	104	1	1
2	Concrete bench	78	1	2
3	Retaining wall (1)	126	1	3
4	Retaining wall (2)	268	1	10

4. RESULTS

Figure 9 shows estimated matching matrices and matching results using dataset 1. The figure shows an estimated matching matrix for conventional processing (upper left in the figure), a matching result from conventional processing (upper right in the figure), an estimated matching matrix for the proposed processing (lower left in the figure), and the matching result from the proposed methodology (lower right in the figure).

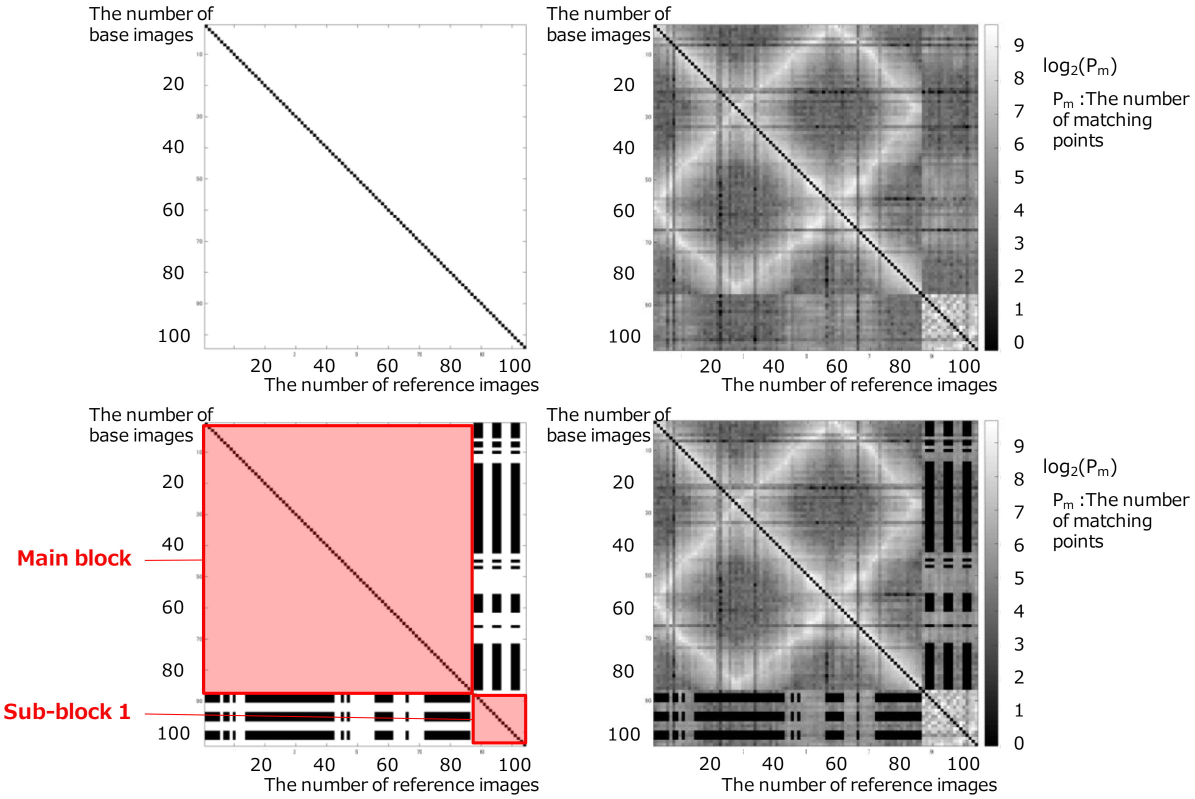

Figure 9. Estimated matching matrices and matching results for dataset 1

In the left matrices, white and red areas indicate image combinations for matching candidates, and black areas indicate omitted combinations in image matching. The lower left matrix shows that the acquired images consisted of a main block with several subblocks, which are indicated as red areas. In the lower right matrices, the number of matching points in each image is indicated by logarithmic gray scale values. The figure also shows that image-matching candidates in each matching matrix were reduced from the upper matrices to the lower matrices when the number of sampling lines was nine. Results using datasets 2, 3, and 4 are also shown in Figures 10, 11, and 12, respectively.

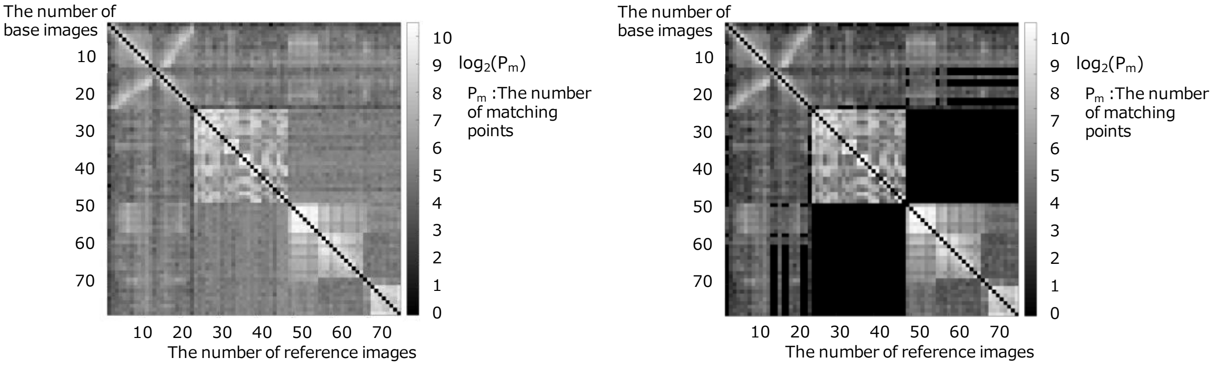

Figure 10. Estimated matching matrices and matching results for dataset 2 (left: matching result from conventional processing, right: matching result from proposed methodology)

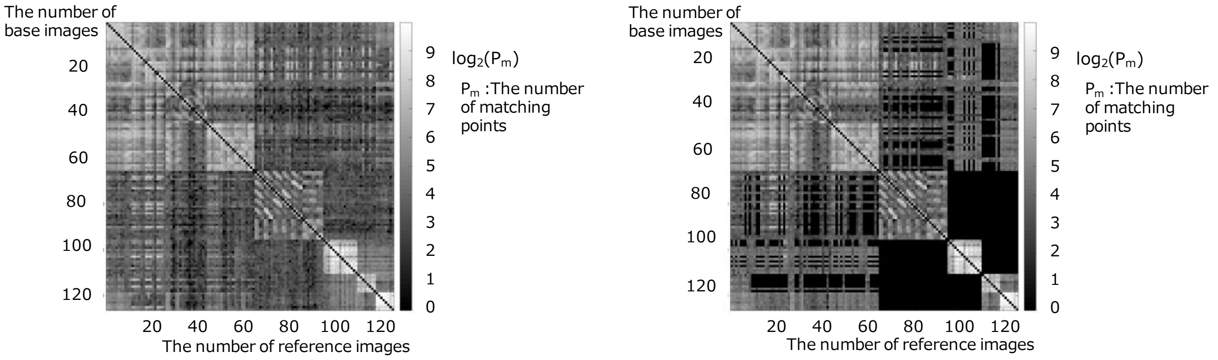

Figure 11. Estimated matching matrices and matching results for dataset 3 (left: matching result from conventional processing, right: matching result from proposed methodology)

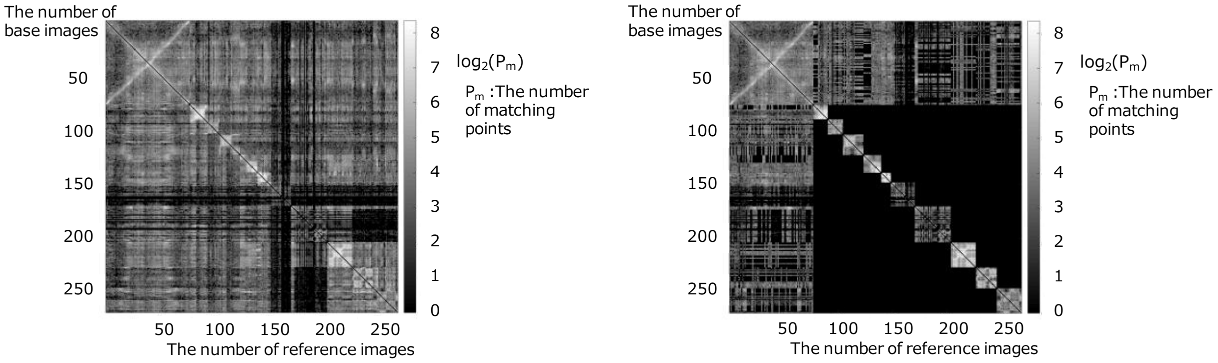

Figure 12. Estimated matching matrices and matching results for dataset 4 (left: matching result from conventional processing, right: matching result from proposed methodology)

In addition, matched images in the main block in datasets 1, 3 and 4 were sparser than those in the subblocks, because images for a main block were acquired along a line with reciprocating translations. Although our methodology depends on an image acquisition rule to prepare ordered images, the number of image-matching candidates in each matching matrix may be reduced further.

Figures 13 and 14 show qualitative evaluation results of camera pose estimation and dense point-cloud generation, respectively, using our datasets. Images in each row show the results using each dataset. Figure 13 shows reconstructed camera poses and sparse point-cloud data. The images indicate that camera poses were estimated successfully. Figure 14 shows reconstructed dense point-cloud data. The figure shows that our proposed approach can reconstruct point cloud data successfully, even if part of image matching was omitted.

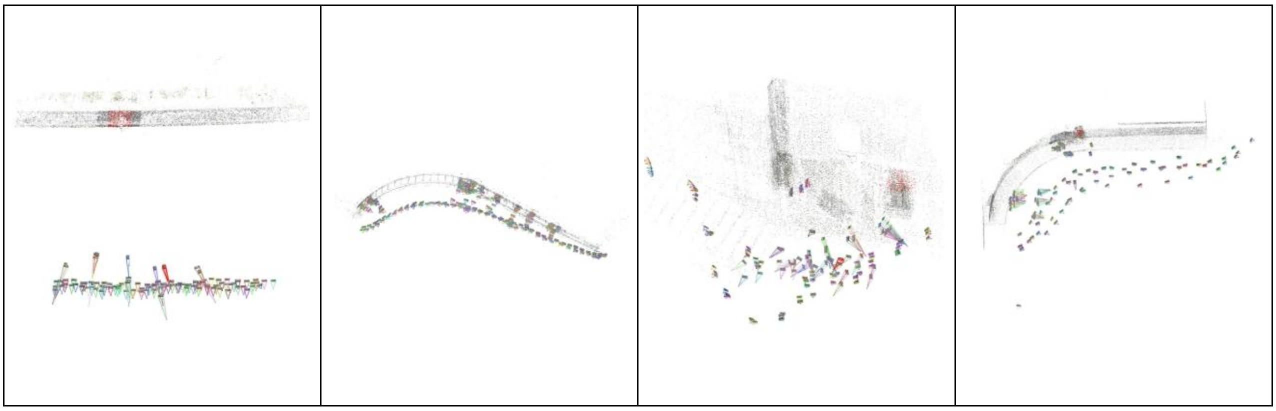

Figure 13. Reconstructed camera poses (from left: dataset 1 (revetment); dataset 2 (concrete bench); dataset 3 (retaining wall 1); dataset 4 (retaining wall 2))

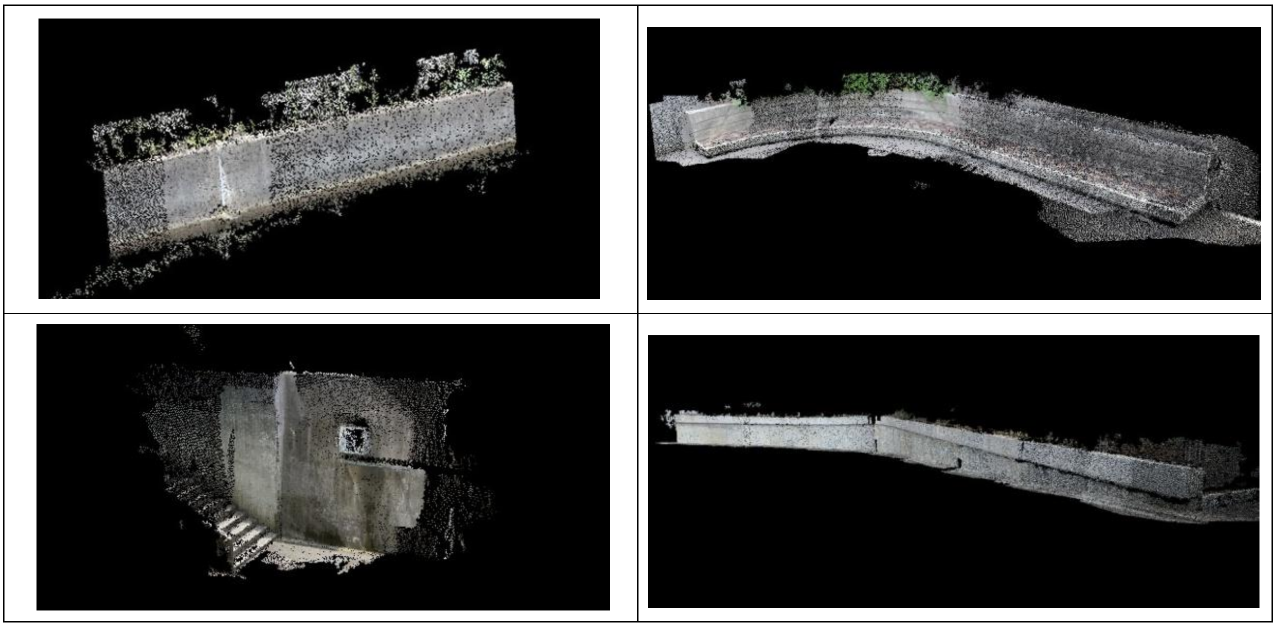

Figure 14. Reconstructed pointcloud data (upper left: dataset 1 (revetment); upper right: dataset 2 (concrete bench); lower left: dataset 3 (retaining wall 1); lower right: dataset 4 (retaining wall 2))

Our numerical evaluation results are as follows. First, Table 2 shows processing results for image matching and point-cloud generation. Although the number of stereo pairs in image matching was reduced in each dataset compared with the conventional methodology, the number of point clouds after the dense image matching changed very little. The result shows that our methodology can improve the efficiency of image matching. We have also confirmed that our methodology performed well when many subblocks occupied a dataset. The reduction rate of image pair candidates was 79% (4258 [pairs] / 5356 [pairs]) in dataset 1 including a subblock, as shown in Table 3. On the other hand, the number of candidates was reduced by 32% (11510 [pairs] / 35778 [pairs]) in dataset 4 including 10 subblocks. The main difference between datasets was the number of subblocks. Thus, we confirmed that our proposed methodology improves image matching procedure for SfM.

Table 2. Processing results in image matching and point cloud generation

	Dataset	1	2	3	4
Conventional methodology	Candidates in image matching [pairs]	5356	3003	7875	35778
	Generated point cloud [pts]	82202	198581	280028	607522
Proposed methodology	Candidates in image matching [pairs]	4258	1993	4663	11510
	Generated point cloud [pts]	82493	192975	314331	642779

Table 3. Reduction rate of image pair candidates

Dataset	1	2	3	4
The number of sub-blocks	1	2	3	10
Reduction rate of image pair candidates	79% (4258/5356)	66% (1993/3003)	59% (4663/7875)	32% (11510/35778)

Second, Table 4 shows processing time for the conventional methodology and our proposed approach. The processing consists of feature detection, image matching, SfM, and dense point reconstruction. The processing time for the feature detection with our proposed methodology was similar to that for the conventional methodology, because he same images were used in both cases. There were few differences between the conventional methodology and the proposed methodology in the SfM and dense point reconstruction. However, there were differences between the conventional methodology and the proposed methodology in the image matching. With dataset 1, the processing time was shortened to 89% (387 [s]/433 [s]). Moreover, the processing time was shortened to 54% (280 [s]/518 [s]) for dataset 4.

Table 4. Processing time

	Dataset	1	2	3	4
Conventional methodology	Feature Detection [s]	44	40	44	89
	Image matching [s]	433	453	230	518
	Structure-From-Motion [s]	65	51	77	190
	Dense reconstruction [s]	171	212	515	640
Proposed methodology	Feature Detection [s]	44	40	44	89
	Image matching [s]	387	331	181	280
	Structure-From-Motion [s]	58	44	74	146
	Dense reconstruction [s]	214	217	283	686

5. SUMMARY

In this paper, we have focused on improving the efficiency of SfM for infrastructure inspection using close- and long-range images taken from various viewpoints. We proposed the camera network reduction using a matching matrix with ordered image blocks. We have verified that our proposed methodology can reduce the number of image-matching candidates to improve processing speed in SfM. Our experiments using four datasets have confirmed that our proposed methodology can improve the efficiency of SfM.

ACKNOWLEDGEMENT

This work was supported by Japan Construction Information Center (JACIC).

REFERENCES

Irschara, A., Kaufmann, V., Klopschitz, M., Bischof, H., Leberl, F., 2010, Towards fully automatic photogrammetric reconstruction using digital images taken from UAVs, *International Society for Photogrammetry and Remote Sensing Symposium, 100 Years ISPRS*, pp.65-70.

Bay, H., Ess, A., Tuytelaars, T., Van Gool, L., 2008, SURF: Speeded Up Robust Features, *Computer Vision and Image Understanding (CVIU)*, Vol. 110, No. 3, pp.346-359.

Rosten, E., Drummond, T., 2005, Fusing Points and Lines for High Performance Tracking, *Proceedings of the IEEE International Conference on Computer Vision*, Vol. 2, pp.1508-1511.

Sweeney, C., Sattler, T., Hollerer, T., Turk, M., Pollefeys, M., 2015, Optimizing the Viewing Graph for Structure-from-Motion, *ICCV '15 Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV)*, pp. 801-809.

Shen, T., Zhu, S., Fang, T., Zhang, R., Quan, L., 2016, Graph-Based Consistent Matching for Structure-from-Motion, *European Conference on Computer Vision (ECCV) 2016*, pp.139-155.

Schönberger, J., L., Berg, A., C., Frahm, J., 2015, Efficient Two-View Geometry Classification, *German Conference on Pattern Recognition (GCPR)*, 12 pages.

Zhaopeng, C., Tan, Ping., 2015, Global Structure-from-Motion by Similarity Averaging, *2015 IEEE International Conference on Computer Vision (ICCV)*, pp864-872.

Snavely, N., 2010, Bundler: Structure from motion (SFM) for unordered image collections.
(from http://www.cs.cornell.edu/~snavely/bundler/).

Furukawa, Y., Ponce, J., 2010, Accurate, Dense, and Robust Multi-View Stereopsis, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 32, Issue 8, pp.1362-1376.