Streamlining Photogrammetry-based 3D Modeling of Construction Sites using a Smartphone, Cloud Service and Best-view Guidance

R. Moritani[a], S. Kanai[a], K. Akutsu[a], K. Suda[b], A. Elshafey[c], N. Urushidate[d] and M. Nishikawa[e]

[a]Graduate School of Information Science and Technology, Hokkaido University, Japan
[b]Kankyo Fudo Techno Co. Ltd., Japan
[c]Kani Construction Co. Ltd., Japan
[d]Horiguchi-gumi Co. Ltd., Japan
[e]Toho-Engineering Co. Ltd., Japan
E-mail: r_moritani@sdm.ssi.ist.hokudai.ac.jp, kanai@ssi.ist.hokudai.ac.jp, k-suda@bolero.plala.or.jp

Abstract – Three-dimensional (3D) measurement that captures the state of construction sites is key to promoting ICT-supported construction processes. Structure-from-Motion (SfM) and Multi-View Stereo photogrammetry are the best solutions for small and mid-sized construction companies due to their high portability and low cost. However, efficient creation of high-quality 3D dense models using photogrammetry is difficult for site workers because the model quality relies heavily on manually selected camera poses. To address this issue, we propose a photogrammetry process that improves the quality and efficiency of 3D dense model reconstruction to measure construction sites. The proposed process begins with a small initial photo set. Then, a computer-supported best-view guidance system predicts the geometric quality of the dense model, estimates the best target positions for additional photographs using SfM results, and provides workers with that position information. The effectiveness and efficiency of the process and system were evaluated at a real-world construction site. The evaluation demonstrated that the process and system can prevent capturing unnecessary images, improve the efficiency of the on-site photographic work, and generate a dense model with quality assurance. We also found that a smartphone camera is the most suitable device for implementing the process.

Keywords –
Photogrammetry; 3D Modeling; Next-best View; Cloud Service; Smartphone; i-Construction

1 Introduction

In recent years, the "i-Construction" initiative [1], which is intended to improve productivity at construction sites by utilizing ICT, has been promoted at various places in Japan. However, to apply i-Construction initiatives at small and mid-sized construction companies, both the initial and the operating costs of the supporting technologies must be low.

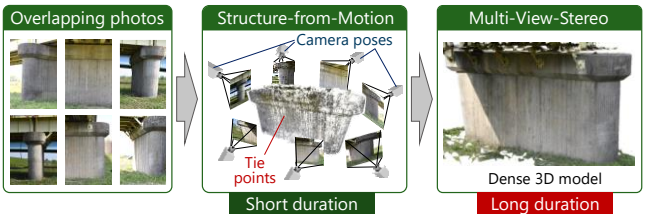


Figure 1. Typical photogrammetry process

Three-dimensional (3D) measurement technology is an essential part of the i-Construction initiative. 3D measurement technology is used to capture the current state of construction sites at various construction stages at high frequency. Currently, terrestrial laser scanners and 3D photogrammetry are used as measurement technologies. Terrestrial laser scanners permit millimeter-accuracy measurements [2]. However, the devices are expensive. In addition, typically, measurement processing is outsourced, which is also expensive. Thus, cost considerations hinder the introduction of laser scanners to small and mid-sized construction companies.

3D photogrammetry [3], shown in Figure 1, is slightly inferior to laser scanners in terms of measurement accuracy. However, it can automatically reconstruct dense 3D models, such as 3D point clouds and textured meshes, from multiple overlapping photos. Photogrammetry can be implemented using UAV-based photography. Consequently, photogrammetry can be

introduced into small and mid-sized construction companies more smoothly than terrestrial laser scanners.

However, introducing photogrammetry and the routine use of high-quality 3D models to monitor construction site activity remains challenging. First, a site worker must take multiple photographs using a heavy, handheld single-lens reflex (SLR) camera. Second, the quality of the dense 3D model reconstructed from the photographs cannot be confirmed on-site because the conventional photogrammetry pipeline requires significant processing time. Moreover, for a given construction site, pre-estimating the camera pose, i.e., the optimal shooting position and orientation, required to capture photographs that can be used to reconstruct a high-quality, dense model is difficult.

To address the various challenges, an innovative photogrammetry process and computer-supported best-view guidance system that can streamline 3D modeling of construction sites is proposed in this paper. The proposed process and the system can be introduced into the everyday activities of small and mid-sized companies by integrating a smartphone, cloud service, and computer-assisted best-view guidance for optimal camera poses. The development of this technology involved both the construction industry and academia, and its effectiveness and efficiency were evaluated experimentally at a real-world construction site.

2 Challenges and Approaches

2.1 Conventional Photogrammetry Process

As shown in Figure 1, the general photogrammetry pipeline to generate a dense 3D model from a set of photographs comprises two steps: Structure-from-Motion (SfM) and Multi-view Stereo (MVS). SfM derives the camera poses and sparse corresponding points, i.e., the so-called "tie points," on real-world objects, and MVS creates a dense 3D model, such as a 3D point cloud or a textured-mesh model, by stereo matching overlapped photographs [3].

SfM processing can be completed in a relatively short time. However, MVS must process all pairs of overlapped photos; therefore, it requires approximately 10 to 50 times more processing time than SfM. For example, for 100 photos, SfM requires only 4 min, while MVS takes 120 min. This example clearly indicates that the MVS step consumes most of the processing time in the photogrammetric 3D model reconstruction process.

As shown in Figure 2, if we attempt to utilize photogrammetry to capture daily progress at a construction sites, the following problems occur.

- The resolution of a 3D model reconstructed by SfM-MVS depends on the resolution of the captured images. Therefore, high-resolution photographs tend to be taken by an SLR camera. However, SLR cameras tend to be large and heavy; carrying and using an SLR camera in complex situations at a construction site, e.g., scaffolding, can be problematic. In addition, with an SLR camera, uploading captured photos via a network is more complicated and less efficient than with a smartphone. Therefore, initially, all captured photos must be stored on the camera's SD memory card. After taking the camera to an office, the 3D model is reconstructed using photogrammetry software. However, this process drops in efficiency because photogrammetry processing cannot be performed during shooting.
- To reconstruct a high-quality, dense model using SfM-MVS, a site worker must carefully consider the importance of capturing overlapping photographs and develop a shooting plan in which the camera poses relative to the object are set appropriately. However, predicting which and how many photos should be taken to reconstruct high-quality models is difficult for the worker. Consequently, defects, such as holes or reduced accuracy in some parts, often appear in the model. To avoid such defects, shooting plans tend to involve a high overlap ratio. However, increasing the number of photographs also increases the MVS processing time.
- To record construction sites using SfM-MVS, it is often necessary to take several hundred to several thousand photographs. With such a large number of photographs, MVS processing can take approximately half a day or even an entire day. Therefore, the quality of the dense model cannot be confirmed on-site immediately after the images are captured. If the model quality is unsatisfactory due

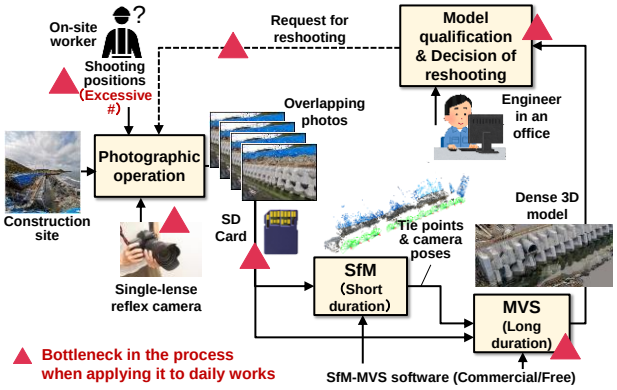


Figure 2. Practical challenges encountered when attempting to apply a conventional photogrammetry process at a construction site

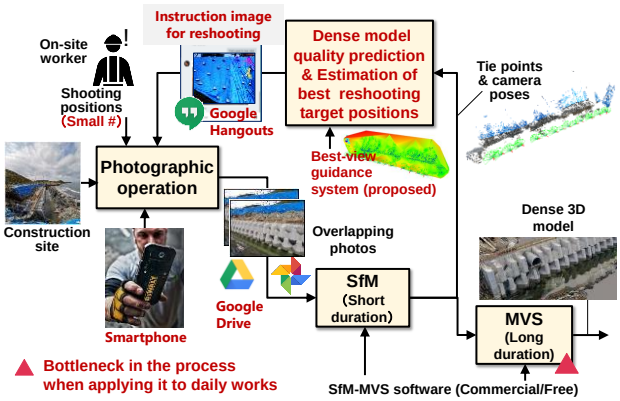


Figure 3. Processing pipeline of the proposed photogrammetry process

to an insufficient number of photographs or unphotographed areas of the site, reshooting will be required, which incurs considerable financial and time costs.

2.2 Proposed Approaches

To address the above problems, the following new photogrammetry process is introduced in this study. The processing pipeline for the proposed process is shown in Figure 3. The process proceeds as follows.

(1) We introduce a smartphone with a high-resolution camera as the device used to capture on-site photographs. Smartphones are lightweight, which makes them suitable for handheld shooting at construction sites. In addition, by utilizing their internet communication function, images can be automatically uploaded to cloud storage (Google Drive) immediately after shooting.

(2) The quality of the dense model that will be reconstructed from the uploaded images is quickly predicted from the SfM results and a computer-supported best-view guidance system. The prediction is based on our developed quality prediction algorithm [4]. The best target positions for additional photo shoots that would improve the quality are estimated in a few minutes by the guidance system connected to the cloud.

(3) The system automatically generates an instruction image in which the marker symbols of these target positions estimated in (2) are superimposed on the photo of the scene saved in the cloud. Furthermore, the instruction image is immediately transmitted to the worker's smartphone at the construction site using a messenger application (Google Hangouts [5]). Then, the worker takes several additional photos according to the target positions on the instruction image and uploads them to the cloud.

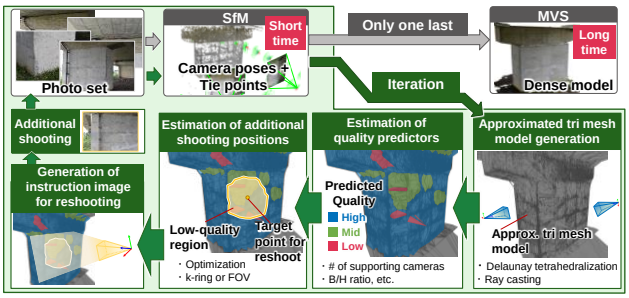


Figure 4. Prediction of the dense model quality and estimation of the best target position for additional shootings

(4) By repeating processes (2) and (3) as often as necessary, a small number of target positions for additional images are estimated only by SfM processing, and additional images are captured according to the target positions. Consequently, the quality of the obtained high-density model is improved successively without requiring multiple MVS processing.

(5) After completing the shooting processes (2)–(4) supported by the best-view guidance system, the time-consuming MVS process is executed only once, and the final dense model is reconstructed. Thus, the MVS process can begin immediately after the required additional images are captured, which improves the overall efficiency of the model reconstruction process.

3 Quality Prediction of Dense Model and Estimation of Additional Shooting Positions

3.1 Approximated Triangular Mesh Model Generation

Figure 4 shows the processing pipeline of the prediction of the dense model quality and estimation of the best target positions to capture additional images. The geometry of the dense model is first approximated by a triangular mesh model generated from the triangulation of tie points created by SfM. The approximation method simplifies a method proposed by Labatut et al. [6] to improve its computational efficiency.

As shown in Figure 5, the triangular mesh generation begins with 3D Delaunay tetrahedralization of 3D tie point set P and creates a set of tetrahedra H. Then, the intersection test is performed between every tetrahedron in P and a set of rays $V_i = \{\boldsymbol{v}_j^i\}\,(\boldsymbol{v}_j^i = \boldsymbol{p}_i - \boldsymbol{c}_j)$ beginning from the projection center of the j-th camera $\boldsymbol{c}_j$ to the i-th visible tie point position $\boldsymbol{p}_i$ $(i \in P)$.

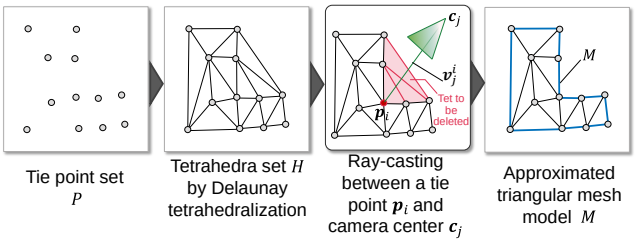


Figure 5. Generation process of the approximated triangular mesh model *M*

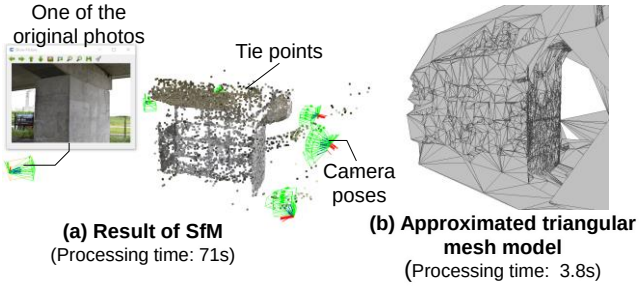


Figure 6. Example of the approximated triangular mesh model *M*

If a tetrahedron intersects with a ray, it is deleted, and the remaining set of tetrahedra is defined as H'. Finally, we obtain the approximated triangular mesh model *M* by taking the surface boundary meshes of H'. The algorithm is described in detail in a previous study [4].

Figure 6(a) shows an example of the tie point set *P* and camera poses generated from 33 original photos of a bridge pier, and Figure 6(b) shows the corresponding approximated triangular mesh model.

3.2 Quality Predictor Estimation

Next, the quality predictors $F_X(i)$ for a dense model are evaluated at a sparse point i ($\in P'$, P': sparce point set) that constitutes the vertex of the approximated triangular mesh model *M* based on the tie point set *P* and the camera pose set $E = \{\boldsymbol{e}_j = (\boldsymbol{c}_j, \boldsymbol{\theta}_j)\}$, where $\boldsymbol{\theta}_j (\in R^3)$ is a vector of three Euler angles representing the projection orientation of the *j*-th camera. The predictor $F_X(i)$ quantifies how accurately the final dense model can be reconstructed around the sparce point $i (\in P')$. The basic idea of the quality predictor was initially proposed by Mauro et al. [7]. Note that we designed different types of predictors based on that study [7].

In the proposed method, the following four quality predictors are evaluated at each sparce point $i (\in P')$.

- **Reliability** $(F_r(i))$. The local geometric quality of the reconstructed dense model around a sparce point *i* decreases as the n umber of visible cameras $|V_i|$ supporting a point i decreases. Therefore, the *Reliability* predictor of the point i is defined as follows:

$$F_r(i) = |V_i| \tag{1}$$

- **Area** $(F_a(i))$. As the area of a triangle on *M* enlarges, the reconstruction error of the dense model tends to be large. Therefore, the average area of the triangles on *M* adjacent to a point *i* is evaluated as the *Area* predictor defined by as follows:

$$F_a(i) = \frac{1}{|T^i|} \Sigma_{t_j^i \in T^i} area(t_j^i) \tag{2}$$

where T^i denotes a set of triangles adjacent to i.

- **Edge length** $(F_e(i))$. When the object surface to be measured is poorly textured, the edge length of a triangle on *M* tends to be long and the point clouds generated by SfM become sparce. Thus, the average edge length adjacent to a point *i* is evaluated as the *Edge length* predictor expressed as follows:

$$F_e(i) = \frac{1}{|D^i|} \Sigma_{j \in D^i} length(e_j^i) \tag{3}$$

where D^i denotes a set of edges connected to i.

- **Baseline and height ratio** $(F_{bh}(i))$. Based on the principle of stereovision, higher-quality reconstruction by MVS is obtained from a correct ratio between the baseline length and height. The baseline length is the distance between two camera positions $\boldsymbol{c}_j$ and $\boldsymbol{c}_k$, and the baseline height is the distance between the space point position $\boldsymbol{p}_i$ and the midpoint of the baseline $\boldsymbol{c}'_{jk}$. It is well known in photogrammetry that the quality of the dense model is related to this ratio [8]. Therefore, the ratio is evaluated as the *Baseline and height ratio* predictor as follows:

$$F_{bh}(i) = \frac{1}{|J_i|} \Sigma_{(j,k) \in J_i} \left(\frac{\|\boldsymbol{c}_j - \boldsymbol{c}_k\|}{\|\boldsymbol{p}_i - \boldsymbol{c}'_{jk}\|} \right) \tag{4}$$

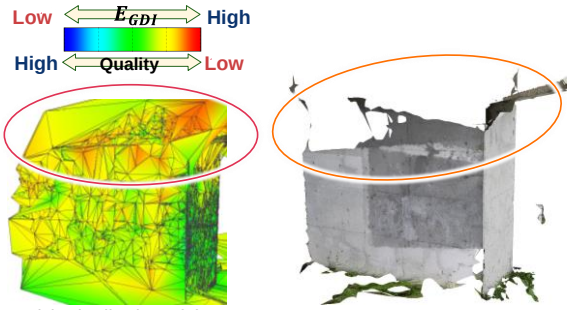
where J_i denotes a set of all possible camera pairs visible from a sparce point i.

The detailed calculation of the indicators is explained in a previous study [4].

To consolidate the four quality predictors into a single indicator representing the degradation of the dense model, first, we converted each of the predictors given by Equations (1–4) to a normalized energy $\in [0,1]$ using the logistic function $L(\,)$ proposed by Mauro et al. [7] and quadratic function $K(\,)$ as follows:

$$E_X(i) = \begin{cases} L(F_X(i) - \mu_X, \sigma_X), & X \in \{a, e\}; \\ 1 - L(F_X(i) - \mu_X, \sigma_X), & X \in \{r\}; \\ 1 - K(F_X(i), \sigma_X), & X \in \{bh\}, \end{cases} \tag{5}$$

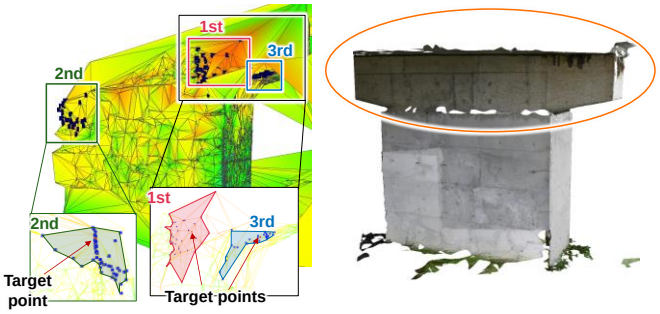
where μ_X denotes the average of F_X, σ_X is the standard deviation of F_X, $L(x - \mu, \sigma) = 1/\left(1 + \exp\left(-\frac{2(x-\mu)}{\sigma}\right)\right)$, and $K(x, \sigma) = 1/(1 + (x - 0.5/\sigma)^2)$. In Equation (5), higher energy means that the geometry of the final



(a) Distribution of the geometry degradation indicators (E_{GDI}) on M

(b) Dense model generated by MVS

Figure 7. Correlation between the distribution of the geometry degradation indicators and the dense model by MVS for original photos



(a) Derived three low-quality regions and the target points ($s = 3$)

(b) Dense model generated from the 33 original photos and 3 additional photos set by MVS

Figure 8. Three target points for additional image capture and the improved dense model obtained used MVS after additional images have been captured

dense model is more degraded.

Finally, the four energy values $E_X(i)$ are aggregated by taking an average to denote a *geometry degradation indicator* at a sparse point i as $E_{GDI}(i)$ as follows.

$$E_{GDI}(i) = \big(E_r(i) + E_a(i) + E_e(i) + E_{bh}(i)\big)/4 \qquad (6)$$

A region with high indicator value $E_{GDI}(i)$ on the approximated triangular mesh model $M^{'}$ indicates that the local region around the sparse point i on the dense model has a more significant possibility of degrading the geometry. It also implies that valid photos are lacking in the region with a high indicator value and that additional image s should be preferentially captured to improve the dense model quality of the region around the point i.

Figure 7(a) shows the distributions of the indicator values $E_{GDI}(i)$ on the approximated mesh model M of the pier shown in Figure 6. The predicted quality of the upper part of the pier is low (red), which suggests that the number of images captured capturing in this area was insufficient. Figure 7(b) shows a dense model generated by MVS from the original 33 photos. In Figure 7(a), the upper parts of the pier shape with high indicator values were not fully reconstructed in the dense model. Thus, it is evident that the quality prediction based on the geometry degradation indicator $E_{GDI}(i)$ is functioning.

3.3 Estimation of Additional Shooting Positions

Low-quality areas on a dense model should be improved by capturing additional images as efficiently as possible. To this end, it is preferable to identify target positions of as many low-quality areas as possible for an additional photo shoot. Therefore, based on the geometry degradation indicator, the target points for the additional photo shoots are selected by an optimization.

First, for every sparce point $i(\in P')$ on the approximated model M, the geometry degradation indicator $E_{GDI}(i)$ value is added as an attribute value w_i. Then, the degree of degradation in the peripheral region of i is estimated both from a target point $j(\in P')$ and from the indicator values of the sparce points i' included in the region near the target point j. A photo shoot to capture additional images should be oriented to cover the areas with the most considerable geometry degradation. Finally, $s$target points for additional photo shoots are derived from the sparce point set P' using the combinatorial optimization and a greedy algorithm. Details of the optimization process are presented in a previous study [4].

Figure 8(a) shows the three low-quality areas and the target points for additional image capture. The areas were derived from the distribution of the geometry degradation indicators in Figure 7(a) with $s = 3$. The indicator values in the areas around the target points are higher than in other areas, and the target points can be placed at the low-quality areas appropriately.

Figure 8(b) shows a dense model reconstructed by MVS from 36 images, including the three additional photos corresponding to the three target points. Compared to the model generated from the 33 initial images in Figure 7(b), the quality of the reconstructed area at the top of the pier increased significantly despite adding only three images. Therefore, the effectiveness of the target point selection algorithm can be confirmed.

4 Case Study

4.1 Evaluation of Reconstruction Qualities

A case study was conducted at a seawall construction site shown in Figure 9 (51 m × 2.4 m) on

Figure 9. Scene of the construction site of wave-dissipating block installations at Toyoura coast

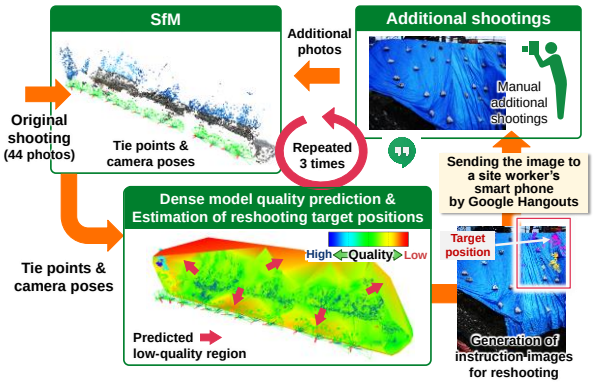


Figure 10. Process flow of the case study

the Toyoura coast in Tomamae-Cho, Hokkaido. The proposed photogrammetry process was performed. The process involved an original photo shoot, dense model quality prediction, on-site best-view guidance, and additional photo shoots to record the installation of wave-dissipating blocks.

To decrease the time required to upload images from a camera to the cloud system, a smartphone (HUAWEI Mate20-Pro) with a built-in high-resolution camera was used. The image resolution was set to 9.7 million pixels to reduce the transfer time, and a wide-angle lens was used. By using a smartphone, captured images could be automatically uploaded to Google Drive immediately. In addition, with this image resolution setting, the upload time could be significantly reduced to within a few seconds per image. Since the SfM and best-view guidance server does not necessarily need to be installed near the construction site, we installed it at the Sapporo campus of Hokkaido University. A high-speed internet connection is available between the university and the construction site on the Toyoura coast.

Figure 10 outlines the process flow of this case study. Forty-four original photos were taken from sparse positions using the smartphone camera by an on-site worker while walking on the top of the wave-dissipating blocks. Next, the next-best target positions to capture additional images were estimated. Then, an instruction image was generated on the server-side and sent to the worker's smart phone. Finally, according to the instruction image, the worker took five to 10 additional photos once and repeated the process of transmitting the images to the server three times. The time required by the process is summarized in Table 1.

Table 1. Processing time in the proposed photogrammetry process

	Total photo # (Additional photo #)	Time for SfM processing	Time for estimating the best target positions	Time for MVS processing
Original	44	**1 min 30 sec**	**3.96 sec**	15 min
1st Addition	49 (5)	**1 min 20 sec**	**4.17 sec**	-
2nd Addition	58 (9)	**1 min 20 sec**	**5.08 sec**	-
3rd addition	68 (10)	**1 min 40 sec**	-	**20 min**

Red Bold： Processing time actually required for the dense model generation using the proposed process

Figure 11 shows the dense model reconstructed by MVS from only the 44 original photos, the estimated best target positions, the corresponding instruction images, and an example of the photos added by the worker. The dense model geometries generated by MVS with those additional images added at each stage are also shown in Figure 11.

As can be seen from Figure 11, it is possible to visually confirm that the 3D model can be generated with relatively good quality even with images captured by the built-in smartphone camera. In addition, using the model quality prediction and estimation of the best target positions to capture additional images, the defects and holes between blocks generated in the model reconstructed from the original images disappeared in the model generated after images were added, and the correct block geometry could be reproduced. The area near the drainage pipes on the upper left of the slope was greatly expanded. As shown in Table 1, estimating the best shooting target position once could be completed in approximately 1.5 min.

From the above results, it is evident that, in a construction site, using a smartphone camera to capture images and as communication device is suitable for 3D photogrammetry measurement in which the model geometry is successively improved. Although the improvement in model quality depends on the number of shots, the result suggested that the proposed process might be able to complete the reconstruction of the dense 3D model on the day that the one-site images were captured.

On the other hand, some areas around the blocks still require additional photo shoots and setting the criteria for terminating these repetitive image capture processes was left as an open problem.

4.2 Estimation of Processing Efficiency

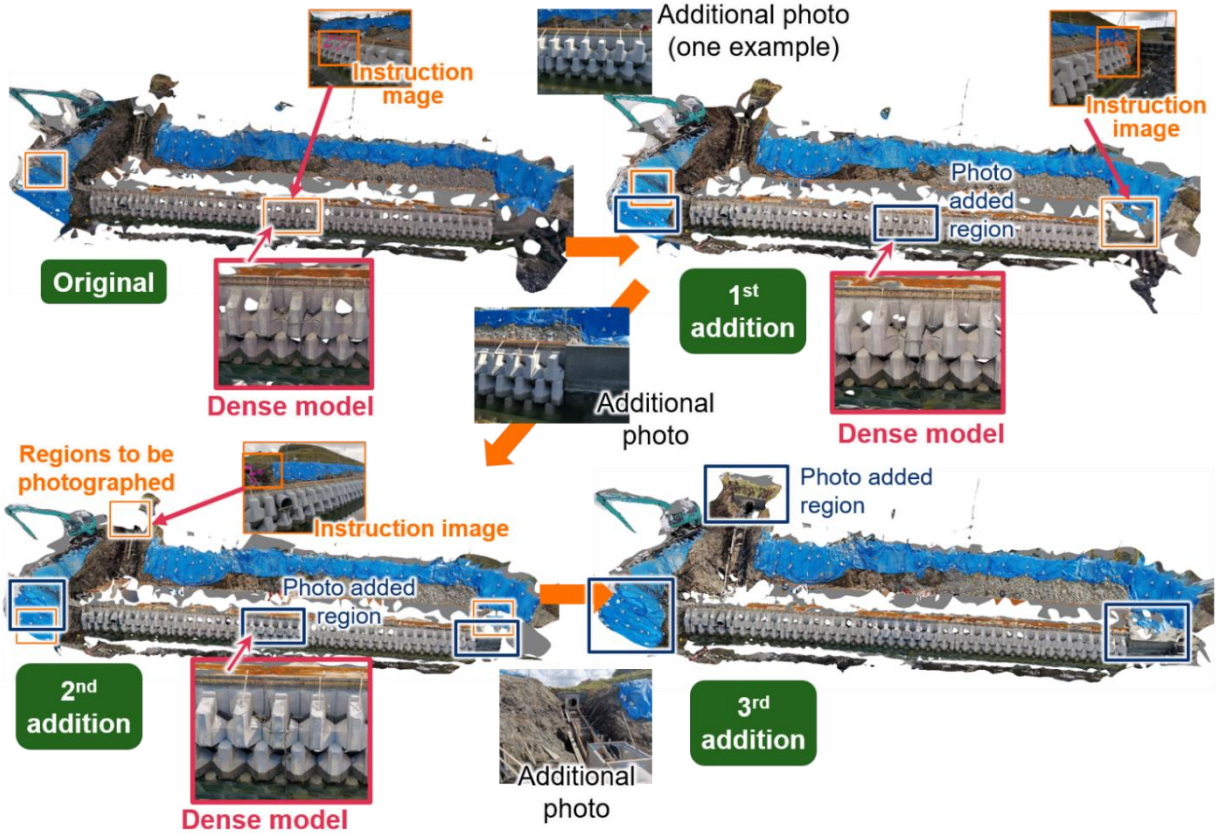


Figure 11. Changes in the dense models by 1st, 2nd, and 3rd addition of images

To quantify how efficient dense model reconstruction using the proposed photogrammetry process is compared to two conventional processes. We compared processing times for the following three approaches.

(1) Conventional photogrammetry with single photo shoot and an excessive number of images. Here, a dense model was reconstructed by performing the SfM-MVS together for approximately 400 images captured at the site.
(2) Conventional photogrammetry with a single additional photo shoot. First, 44 original photos were obtained, and a dense model was generated once by SfM-MVS. Then, the low-quality portions of the dense model were identified manually, and 24 target points for an additional photo shoot were determined. Finally, the SfM-MVS process was re-executed with 68 images to reconstruct a final dense model.
(3) The proposed process. Here, the process started with 44 original images. Then, SfM was performed, target positions to obtain additional images were identified using a computer, and five, nine, and ten photographs were added to the original images. Finally, the dense model was reconstructed by performing MVS only once for the 68 images acquired.

Note, for the processing efficiency comparison, the Toyoura coast construction site was taken as an example (Section 4.1).

The bar chart in Figure 12 shows the comparison results. In estimating the processing time, referring to the values obtained from the construction site of section 4.1, the required shooting time per photograph was estimated to be 15 s. The SfM and MVS processing time per photo was 0.03 and 0.3 min, respectively. Moreover, the time required to estimate the shooting target positions was assumed to be constant at 6 s. Note that the image upload time was included in the shooting time because it was only a few seconds per photo.

As can be seen from the comparison in Fig. 12, process (1) required approximately 3.5 h to reconstruct the dense model from the excessively captured photos. With the proposed process (3), the dense model could be reconstructed in 40 min, which is approximately one-fifth of the time required by process (1).

In addition, since process (1) requires MVS processing of a significant number of images, which takes considerable amount of time, the quality of the dense model cannot be confirmed until the processing is

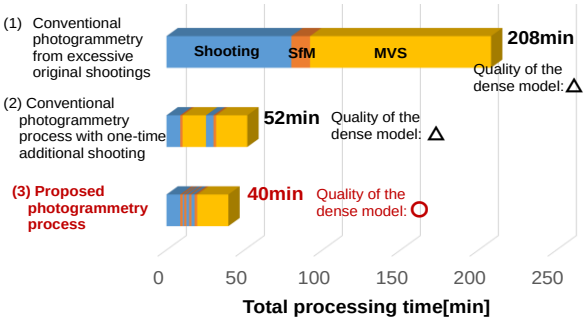


Figure 12. Comparison of the total processing times among (1) conventional photogrammetry process with excessive image capture, (2) conventional photogrammetry process with a single shoot to capture additional images and (3) the proposed photogrammetry process

complete, i.e., approximately 3.5 h after the photo shoot. On the other hand, with the proposed process (3), we estimate that the shooting target positions that reflect the prediction of the dense model quality for the currently captured photos can be fed back to a site worker in approximately 2–3 min after capturing the photo. Thus, it is possible to ensure that all required images are captured and to realize an efficient photo shoot.

Furthermore, in processes (1) and (2), the selection of shooting positions is left to the user; thus, there is no guarantee that the additional photos will improve the reconstructed model's quality. In contrast, with the proposed process (3), since the computer selects the best positions at which additional photos should be taken based on the model quality estimation, it is highly likely that additional images will effectively contribute to the quality improvement of the reconstructed dense model.

5 Conclusion

In this paper, we have proposed a new photogrammetry process that improves the quality and efficiency of dense model reconstruction of construction sites. The proposed process begins with a small original photo set. Then, the computer-supported best-view guidance system predicts the geometric quality of the dense model, estimates the best target positions for additional photo shoots using only SfM results, and feeds back those positions to a site worker. Depending on the number of target positions, the feedback process could complete in 1.5 min. The effectiveness of the proposed process and the system was evaluated at a real-world construction site. As a result, it was found the process and the system could prevent excessive image capture, improve the efficiency of the on-site photo shoots, and generate the dense model with a certain degree of quality assurance. We also found that a smartphone, which can send and receive images to and from the construction site, was the most suitable shooting device for implementing the process.

However, currently, some server-side operations still require manual processing. In the future, we would like to implement fully automated processes that include SfM and best-view guidance on a cloud server.

Acknowledgments

This research was supported by the following grant-in-aid from the Ministry of Land, Infrastructure, Transport, and Tourism in FY2018-19: “Project on the introduction and utilization of innovative technology for dramatically improving the productivity of construction sites: Utilizing data to improve the productivity in construction labor in civil engineering.”

References

[1] Ministry of Land, Infrastructure, Transport, and Tourism. i-Construction (in Japanese). On-line: https://www.mlit.go.jp/tec/i-construction/index.html, Accessed: 10/06/2020.

[2] Riveiro B. and Lindenbergh R. *Laser Scanning: An emerging technology in structural engineering*, CRC Press/Belkema, Leiden, 2019.

[3] Luhmann T., Robson S., Kyle S. and Boehm J. *Close-range Photogrammetry and 3D Imaging (3rd Edition)*, De Gruyter, Berlin/Boston, 2019.

[4] Moritani R., Kanai S., Date H., Niina Y. and Honma R. Quality prediction of dense points generated by structure from motion for high-quality and efficient as-is model reconstruction. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, XLII-2/W13: 95–101, 2019.

[5] Google Hangouts, On-line: https://hangouts.google.com/, Accessed: 10/06/2020.

[6] Labatut P., Pons J.-P. and Keriven R. Efficient multi-view reconstruction of large-scale scenes using interest points, Delaunay triangulation and graph cuts. In *IEEE 11th International Conference on Computer Vision*, pages 1–8, Rio de Janeiro, Brazil, 2007.

[7] Mauro M., Riemenschneider H., Signoroni A., Leonardi R. and Van Gool L.J. A unified framework for content-aware view selection and planning through view importance, In *British Machine Vision Conference*, pages 1–11, Nottingham, U.K, 2014.

[8] Yan L., Fei L., Chen C., Ye Z. and Zhu, R. A multi-view dense image matching method for high-resolution aerial imagery based on a graph network. *Remote Sens.*, 8(10):799, 2016.