Streamlining the Photogrammetry-based 3D Modelling of Construction Sites with a Smartphone, Cloud Service and Best-view Guidance

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

[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-mail: kanai@ssi.ist.hokudai.ac.jp, k-suda@bolero.plala.or.jp

Abstract –

Three-dimensional (3D) measurement that captures the states of construction sites is a key to promote ICT-supported construction processes. Photogrammetry composed of Structure-from-Motion (SfM) and Multi-View Stereo is the best solution for it targeting small and mid-sized construction companies because of its high portability and low cost. However, it is difficult for a site worker to efficiently create high-quality 3D dense models by the photogrammetry since the model quality relies heavily on the manually selected camera poses. To solve the challenge, we proposed a new photogrammetry process that improves the quality and efficiency of 3D dense model reconstruction for measuring construction sites. This process starts with a small number of the 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 shootings using SfM results, and instructs those positions to a worker. The effectiveness and efficiency of the process and system were evaluated at a construction site. As a result, it was found they can prevent excessive image shooting, improve the efficiency of the on-site photographing work, and generate a dense model with quality assurance. We also found that a smartphone is the most suitable shooting device for implementing the process.

Keywords –

Photogrammetry; 3D Modelling; Next-best View; Cloud Service; Smartphone; i-Construction

1 Introduction

In recent years, the "i-Construction" initiative [1] for improving productivity at work of construction sites by utilizing ICT has been promoted at various places in Japan. On the other hand, to develop the i-Construction in the small and mid-sized construction companies, it is also essential that its supporting technologies consist of the ones with low initial and operational costs.

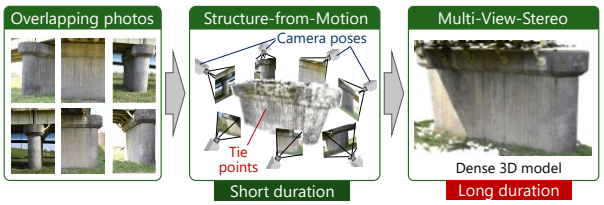


Figure 1. Typical photogrammetry process

As an essential part of the i-Construction initiative, three-dimensional (3D) measurement technology is required that captures the present states of construction sites at various construction stages with high frequency. For this purpose, terrestrial laser scanners and 3D photogrammetry are currently used from a practical perspective. Terrestrial laser scanners permit millimeter-accuracy measurement [2]. However, the scanner devices and outsourcing of the measurement works involve high cost, and they eventually hinder the introduction of the laser scanners to small and mid-sized construction companies.

On the other hand, 3D photogrammetry [3] shown in Figure 1 is slightly inferior to laser scanners in terms of measurement accuracy. Still, it can automatically reconstruct dense 3D models such as 3D point clouds and textured meshes from many overlapping photos. The photogrammetry also has a high affinity for UAV-based aerial photography. For these reasons, the photogrammetry can be introduced into small and mid-sized construction companies more smoothly than terrestrial laser scanners.

However, when trying to reconstruct the high-quality

3D models of construction sites by the photogrammetry as everyday activities and to make use of the models for frequent monitoring of the construction progresses, many challenges remain. First, time-consuming hand-held shooting work with a heavy single-lens reflex camera must be taken by a site worker. Second, the quality of a dense 3D model that will be reconstructed cannot be confirmed on-site by the worker because the conventional photogrammetry pipeline consumes an enormous amount of the processing time. Moreover, finally, it is challenging to pre-estimate the optimal photo-shooting position and orientation so-called "*camera pose,"* where a high-quality, dense model can be reconstructed for a given construction site.

This study develops a new photogrammetry process and computer-supported best-view guidance system that can streamline 3D modeling of construction sites to solve the challenges. The proposed process and the system can be introduced into everyday activities of small and mid-sized companies by integrating a smartphone, cloud service, and the computer-assisted best-view guidance of the optimal camera poses for shootings. The industry and academia developed this technology, and its effectiveness and efficiency were evaluated experimentally at a real construction site.

2 Challenges and Our Approaches

2.1 Challenges in Conventional Photogrammetry Process

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

SfM can complete its processing in a relatively short time. However, MVS must perform tons of stereo matching between every pair of overlapped photos, so it usually spends about 10 to 50 times more processing time than SfM. For example, for 100 photos, SfM requires only 4 min, while MVS spends 120 min. As indicated in this example, most of the 3D model reconstruction process in photogrammetry is characterized by being spent in the MVS step.

As shown in Figure 2, if we try to utilize the photogrammetry to capture construction sites' daily progress, the following problems happen.

- The resolution of a 3D model reconstructed by SfM-MVS depends on the resolution of the captured photo. Therefore, there is a strong tendency for high-resolution photos to be taken by a single-lens reflex camera. However, the single-lens reflex camera is massive, and it has a problem in portability for handheld shooting in construction sites where the scaffolding is weak. Also, with a single-lens reflex camera, the upload function of captured photos via a network is not sufficient as with smartphones. Therefore, we must store all captured photos in the SD card in the camera first. After taking it back to the office, it is necessary to reconstruct the 3D model using photogrammetry software. However, this process erodes efficiency because the photogrammetry processing cannot be performed during shooting or return to the office.
- In order to reconstruct a high-quality, dense model using SfM-MVS, it is necessary for a site worker to carefully consider the photo overlap condition and draw up a shooting plan in which the camera poses to the object are set appropriately. However, it is difficult for the worker to predict from which and how many photos should be taken to reconstruct high-quality models. As a result, defects such as holes or degraded accuracy parts often appear on the model. On the other hand, to avoid these, excessive photo shooting with a high overlap ratio tends to be performed. However, it makes the number of photos enormous, and the MVS processing will take an excessive amount of time.
- To record construction sites using SfM-MVS, it is often necessary to take several hundred to several thousand photos. When the number of phots is vast in this way, MVS processing takes about half a day to a day, so it is not possible to immediately confirm the dense model quality on-site immediately after shooting. For this reason, if an unfortunate quality part due to lack of photos or insufficient shootings is found on the model after MVS, the cost and time loss for reshooting will be massive.

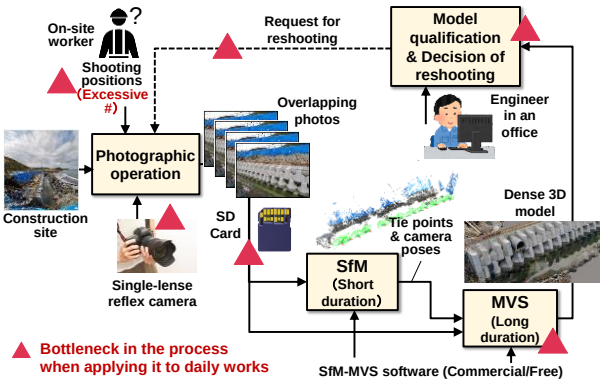


Figure 2. Challenges in conventional photogrammetry process when applying it to daily works

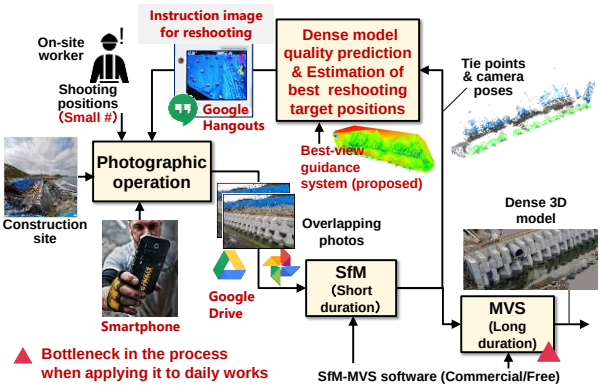


Figure 3. The processing pipeline of the proposed photogrammetry process

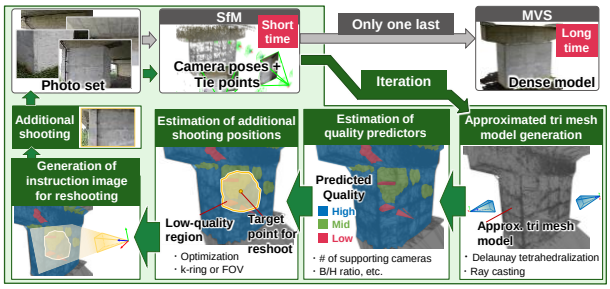


Figure 4. Prediction of the dense model quality and estimation of the best target position for additional shootings

2.2 Our Approaches

Therefore, to solve the above problems, the following new photogrammetry process is introduced in this study. Figure 3 shows the processing pipeline for this new process. The outline is as follows.

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

(2) The quality of the dense model that will be finally reconstructed from the uploaded photos is quickly predicted only from the SfM results by our best-view guidance system using our quality prediction algorithm [4]. Besides, the best target positions for additional shootings that would improve the quality is estimated in a few minutes by the 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 back again on the cloud.

(4) By repeating the above processes (2) and (3) as often as necessary, the small number of target positions for additional shootings are estimated only by SfM processing, and the shootings are conducted according to the target positions. Simultaneously, the quality of the high-density model obtained is improved successively without executing any MVS processing.

(5) After completing the shooting process of (2)-(4) supported by the best-view guidance system, the MVS process that requires much time is executed only once, and the final dense model is reconstructed. Thus, the MVS process can be started right after the additional shootings at the construction sites are completed, which improves the overall efficiency of the model reconstruction.

The principle of the quality prediction of the dense model and the estimation of the optimum target positions for additional shootings will be described below. We also introduce the evaluations of the effectiveness and efficiency of the proposed photogrammetry process at a real construction site.

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 for additional shootings. 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 the one formerly proposed by Labatut et al. [6] to improve its computational efficiency.

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

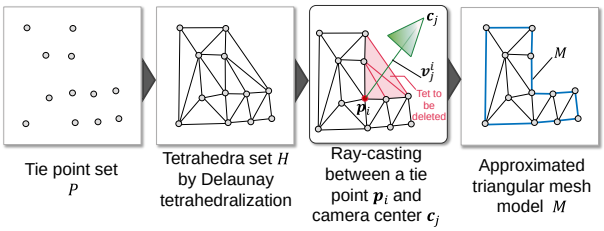


Figure 5. Generation process of the approximated triangular mesh model M

intersects with the ray, it is deleted, and the remaining set of tetrahedra is defined as H'. Finally, we obtained the approximated triangular mesh model M by taking the surface boundary meshes of H'. For detail algorith m, refer to the reference [4].

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

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 is initially proposed by Mauro et al. [7], but we designed different kinds of predictors from [7].

The following four quality predictors are evaluated respectively at each sparce point $i (\in P')$ in our method.

- **Reliability** ($F_r(i)$): The local geometric quality of the reconstructed dense model around a sparce point i decreases as the number of visible cameras $|V_i|$ supporting a point i decreases. Therefore, the *reliability* predictor of the point i is defined by Equation (1).

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

- **Area** ($F_a(i)$): When the area of a triangle on M is larger, the reconstruction error of the dense model tends to be larger. So, the average area of the triangles on M adjacent to a point i is evaluated as the *area* predictor defined by Equation (2).

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

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. So, the average edge length adjacent to a point i is evaluated as the *Edge length* predictor defined by Equation (3).

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

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

- **Baseline and height ratio** ($F_{bh}(i)$): Based on the principle of the stereovision, higher-quality reconstruction by MVS is obtained from a correct ratio between the baseline length and height. The baseline length is a distance between two camera positions $\boldsymbol{c}_j$ and $\boldsymbol{c}_k$, while the baseline height is a 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 secures if this ratio is in the right range [8]. Therefore, the ratio is evaluated as the *Baseline and height ratio* predictor defined by Equation (4).

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

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

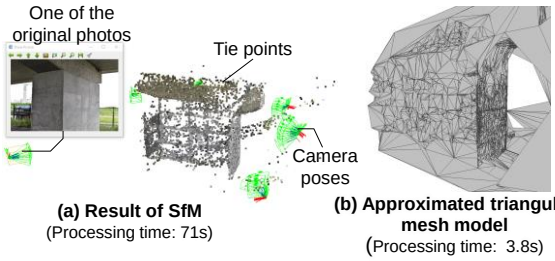


Figure 6. An example of the approximated triangular mesh model M

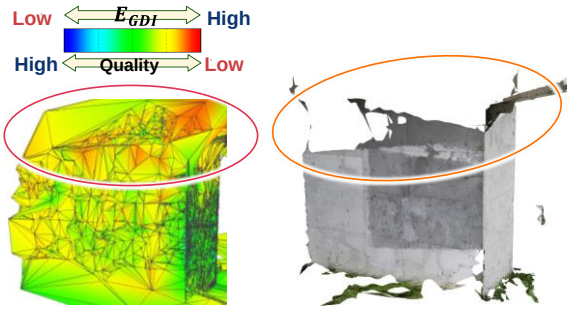
The detail calculation of the indicators, see reference [4].

To consolidate the four quality predictors to one 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 [7] and quadratic function $K(\,)$ as Equation (5).

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

where μ_X denotes the average of F_X, σ_X 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 dense model degrades more.

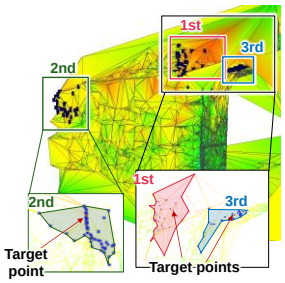
Finally, the four energy values $E_X(i)$ are aggregated



(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. The three target points for additional shootings and the improved dense model after these additional shootings by MVS

by taking an average to denote a *geometry degradation indicator* at a sparse point i as $E_{GDI}(i)$ as Equation (6).

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

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 photos should be preferentially shot 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 in Figure 6. The predicted quality of the upper part of the pier is low (red), which suggests that the photo capturing in this area were 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. It can be seen 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 additional shootings as efficiently as possible. To this end, it is preferable to determine the target positions of the shooting that includes as many low-quality areas as possible in an additional shooting photo. Therefore, based on the geometry degradation indicator, the target points for the additional shootings are selected by the 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. The additional shooting should be oriented to cover the areas with most considerable geometry degradation. Finally, s target points for additional shootings are derived from the sparce point set P' using the combinatorial optimization and greedy method. For details of the optimization, refer to reference [4].

Figure 8(a) shows the three low-quality areas and the target points for additional shooting, which 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.

Furthermore, 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 33 initial images in Figure 7(b), the reconstructed area at the top of the pier increased significantly despite only three added 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 (51m × 2.4m) on the Toyoura coast in Tomamae-Cho, Hokkaido. The proposed photogrammetry process was performed where the original photographing, quality prediction of the dense model, on-site best-view guidance, and additional photo

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

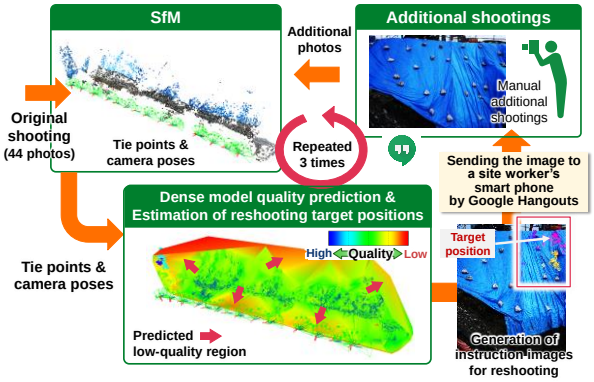


Figure 10. Processing flow of the case study

shooting were repeated to record the work of the installation of wave-dissipating blocks.

In order to decrease the image uploading time 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 suppress the transfer time, and the wide-angle lens was used for the shooting. By using a smartphone, images could be automatically uploaded to Google Drive immediately after shooting. Also, with this image resolution setting, the upload time can be significantly reduced to within a few seconds per image. Since the server computer for SfM and best-view guidance does not necessarily have to be installed near the construction site, we installed it in the Sapporo campus of Hokkaido University. The high-speed internet connection is available between the university building and the construction site of the Toyoura coast.

Figure 10 outlines the processing flow of this case study. As the original photos, 44 images were taken from sparse positions with the smartphone camera by an on-site worker while walking on the top of the wave-dissipating blocks. Next, the next best target positions for the additional shooting were estimated. Then, an instruction image was generated on the server-side and sent to the smartphone of the worker. Finally, according to the instruction image, the worker photographed 5 to 10 additional photos once and repeated the process of transmitting to the server-side three times. The time taken for 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 44 original photos, the estimated best shooting target positions, the corresponding instruction images, and the 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 for the images captured by the built-in camera of a smartphone. Besides, using the model quality prediction and estimation of the best target position for additional shootings, the defects and holes between blocks generated in the model reconstructed from the original images finally disappeared in the model after the image addition, and the correct block geometry could be reproduced. The area near the drainage pipes on the upper left of the slope was greatly expanded. Also, as shown in Table 1, estimating the best shooting target position once could be completed in about 1.5 minutes.

From the above results, using the built-in camera of the smartphone as a shooting and communication device is suitable for 3D measurement of construction sites by photogrammetry process in which the model geometry is successively improved. Although it depends on the number of shots, it was suggested that the proposed process might be able to complete the reconstruction of the dense 3D model on the day of the on-site shot.

On the other hand, some areas around the blocks still require the additional shootings, and it was left as an open problem for setting the criteria for terminating these repetitive shooting processes.

4.2 Estimation of Processing Efficiency

Finally, a trial calculation was performed to quantify how efficient the dense model reconstruction by the proposed photogrammetry process is compared to the conventional process. In the calculation, we compared the processing times when the following processes (1) to (3) were carried out, taking the example of generating a dense model of the construction site in the Toyoura coast in Section 4.1.

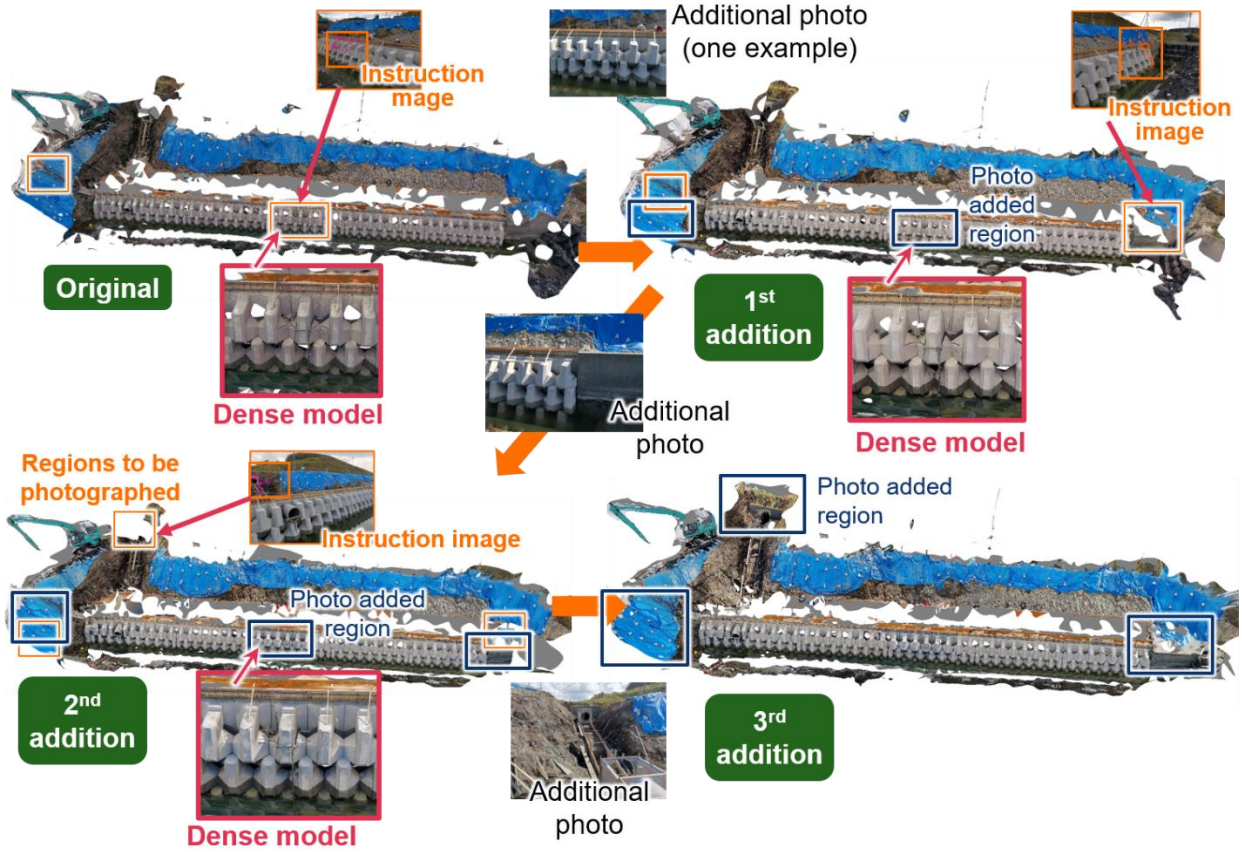


Figure 11. Change in the dense models by 1st, 2nd and 3rd additional shootings

(1) A conventional process where a dense model was reconstructed by performing the SfM-MVS in a lump for around 400 photos excessively captured in the site.

(2) First, 44 original photos were acquired, and a dense model was generated once by SfM-MVS. Next, the user visually confirms the low-quality portions of the dense model, manually determines the 24 target points for additional shootings. Finally, the user re-executes SfM-MVS for a total of 68 images to reconstruct a final dense model.

(3) Using the proposed process, starting from the shooting of 44 original images, performing the SfM, deriving the shooting target positions with a computer, adding five, nine and ten photos, respectively. And finally, the dense model was reconstructed by performing MVS only once for the 68 images acquired.

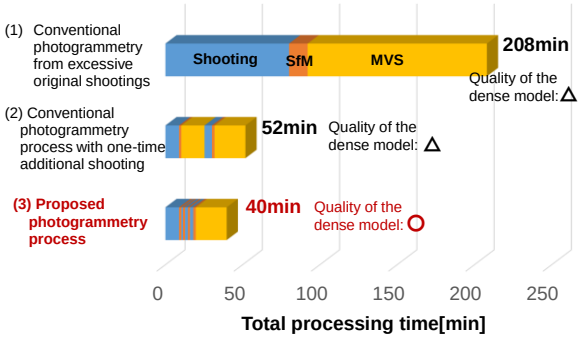


Figure 12. Comparison of the total processing time among (1) conventional photogrammetry process with excessive shootings, (2) conventional photogrammetry process with one-time additional shooting and (3) the proposed photogrammetry process

The bar chart in Figure 12 shows the comparison results. In estimating the processing time, referring to the values obtained from the example in Section 4.1, the required shooting time per photo was estimated to be 15 seconds. The SfM and MVS processing time per photo was 0.03 and 0.3 min, respectively. Moreover, the estimation time of the shooting target positions was assumed to be constant at 6 seconds. Also, the photo 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. 17, the process (1) requires about 3.5h to reconstruct the dense

model from the excessively captured photos. On the other hand, in the proposed process (3), the dense model can be reconstructed in a short processing time of 40 min which only took one-fifth of the process (1).

Also, since the process (1) requires MVS processing for many photos, the quality of the dense model cannot be confirmed until 3.5 hour after the shooting. On the other hand, in the proposed process (3), although it is an estimation, the shooting target positions that reflect the prediction of the dense model quality for the currently captured photos can be fed back to the on-site worker in a waiting time of about 2-3 minutes right after the photo capturing. Thus, it is possible to prevent forgetting to take photos or insufficient shooting at the site and to realize efficient shooting work.

Furthermore, in processes (1) and (2), the selection of the shooting positions is left to the user, so there is no guarantee that the additional photos will improve the reconstruction model's quality. On the other hand, in the process (3), since the computer selects the best positions at which additional photos should be shot based on the model quality estimation, unlike in processes (1) and (2), it is highly possible that additional photos will effectively contribute to the quality improvement of the reconstructed dense model.

5 Conclusion

We proposed a new photogrammetry process that improves the quality and efficiency of dense model reconstruction of construction sites. This process starts with a small number of the 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 shootings only using SfM results, and feedback those positions to a site worker. The feedback could complete only in a few minutes. The effectiveness of the proposed process and the system was evaluated at a construction site. As a result, it was found they could prevent excessive image shooting, improve the efficiency of the on-site photographing work, and generate the dense model with a certain degree of quality assurance. We also found that a smartphone, which is easy to send and receive images to and from the construction site, was the most suitable shooting device for implementing the process.

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

Acknowledgements

This research was financially 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.