

QUALITY PREDICTION OF DENSE POINTS GENERATED BY STRUCTURE FROM MOTION FOR HIGH-QUALITY AND EFFICIENT AS-IS MODEL RECONSTRUCTION

R. Moritani^{1*}, S. Kanai², H. Date², Y. Niina³, R. Honma³

¹ Graduate School of Information Science and Technology, Hokkaido University, Japan - r_moritani@sdm.ssi.ist.hokudaia.ac.jp

² Graduate School of Information Science and Technology, Hokkaido University, Japan - (kanai, hdate)@ssi.ist.hokudaia.ac.jp

³ Asia Air Survey Co., Ltd., - (ysh.niina, ryh.honma)@ajiko.co.jp

Commission VI, WG VI/4

KEY WORDS: Quality prediction, Dense Points, Structure from Motion, Bundle Adjustment, Multi-View Stereo, As-is model

ABSTRACT:

In this paper, we introduce a method for predicting the quality of dense points and selecting low-quality regions on the points generated by the structure from motion (SfM) and multi-view stereo (MVS) pipeline to realize high-quality and efficient as-is model reconstruction, using only results from the former: sparse point clouds and camera poses. The method was shown to estimate the quality of the final dense points as the quality predictor on an approximated model obtained from SfM only, without requiring the time-consuming MVS process. Moreover, the predictors can be used for selection of low-quality regions on the approximated model to estimate the next-best optimum camera poses which could improve quality. Furthermore, the method was applied to the prediction of dense point quality generated from the image sets of a concrete bridge column and construction site, and the prediction was validated in a time much shorter than using MVS. Finally, we discussed the correlation between the predictors and the final dense point quality.

1. INTRODUCTION

Understanding the disparity between the current (as-is) and past states of several aging large-scale infrastructures such as bridges is a priority to administrators in performing their long-term maintenance activities. Nevertheless, current maintenance tasks based on the inspection of documents may not capture deformed regions, leading to poor outcomes.

Structure from motion (SfM) and multi-view stereo (MVS) are probably potential solution methods for addressing the maintenance problem mentioned above. These technologies generate a three-dimensional (3D) as-is model with rich textures of real-world objects from a set of overlapped images. Figure 1 describes the SfM/MVS process. First, the SfM generates a 3D sparse point clouds and camera poses. Second, the MVS generates the dense point clouds or textured mesh model of real-world objects.

However, it is challenging to pre-estimate the optimal camera poses to obtain a high-quality final model in this conventional process. Depending on manually selected camera poses and the number of images, the final dense point clouds or textured mesh model (hereafter called *final dense model*) may include degraded regions (e.g., unwanted holes or irregularities), and they cannot be used for maintenance activities. Moreover, an MVS process consumes an enormous amount of the processing time (e.g., five days by 1,400 photographs).

Therefore, the final goal of this study is to develop the next-best view planning method in SfM/MVS for high-quality and efficient 3D as-is model reconstruction targeted at objects with rich textured surfaces (e.g., concrete surfaces, soils, to mention a few). To achieve this aim, the purpose of this paper is to develop predictors to estimate the geometric quality of a final dense model which would be generated from MVS and to automatically select the low-quality regions where additional images should be

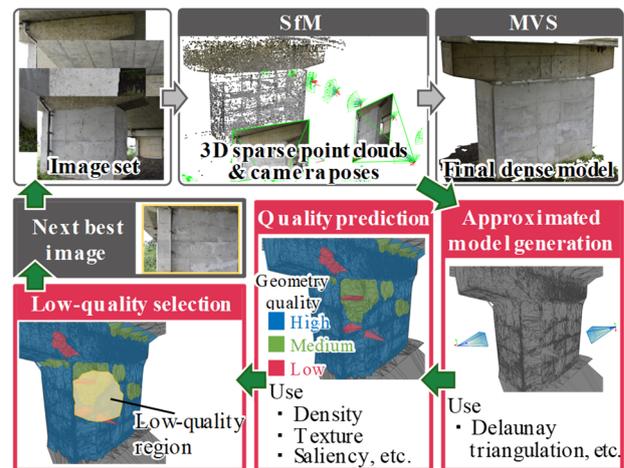


Figure 1. An overview of the proposed quality prediction method.

taken in the next shooting at the earliest possible stage of the SfM/MVS pipeline. As Figure 1 shows, the proposed method can estimate the quality of dense points or textured models without performing MVS. To predict the quality, as shown in Figure 1, we utilize sparse point clouds and camera poses obtained from an SfM whose processing time is much shorter than that of MVS. Moreover, our prediction does not require any *a priori* knowledge on the object to be modeled. Besides, in this paper, we discuss the correlation between the predictors and the final dense point quality.

Several view planning methods for the SfM/MVS have been introduced, including (Hoppe 12), (Hepp, 17), (Jing, 16), (Schmid 12), (Martin 16), and (Roberts, 17); all of which require *a priori* knowledge of rough 3D models, whose generation significantly reduce efficiency as they need to be constructed in advance via the preliminary process of the SfM/MVS, two-dimensional map, or a digital elevation model. Moreover, even if

* Corresponding author

time-consuming MVS process was performed for rough the 3D models, it is not guaranteed that a view planning based on these models estimates the optimal camera pose.

2. QUALITY PREDICTORS EVALUATION

As Figure 1 shows, the proposed method initially performs SfM for the overlapped image sets, and generates 3D sparse point clouds $P = \{i\}$ and camera poses $E = \{e\}$, each of which is expressed as $e = \langle \mathbf{c}, \boldsymbol{\theta} \rangle$, where $\mathbf{c} \in R^3$ denotes a projection center of a camera (i.e., a camera position) and $\boldsymbol{\theta} \in R^3$ is the Euler angles of a camera axis direction. From the set of the 3D sparse points and camera poses, the method then generates an approximated triangular mesh model, and the quality predictors for the final dense model are evaluated on the approximated model using SfM results. Finally, low-quality regions where the corresponding geometry degradation on the final dense model is more likely to be apparent are selected automatically using the approximated model by solving an optimization problem.

2.1 Approximated triangular mesh model generation

To this end, first, we generate an approximated triangular mesh model to test the 3D visibility between the sparse points and the camera positions and to predict quality measures of the 3D geometry of the final dense model. The model generation is based on a method introduced by (Labatut, 2007); however, we simplified their method in order to improve efficiency.

As Figure 2 shows, our method begins with Delaunay triangulation of the sparse point clouds P to generate a set of tetrahedra H , followed by an intersection test between every tetrahedron and a set of rays $V_i = \{v_j^i\}$ ($v_j^i = \mathbf{p}_i - \mathbf{c}_j$) starting from a j -th camera position \mathbf{c}_j to a i -th visible sparse point position \mathbf{p}_i . Here, tetrahedra intersecting with the ray were deleted; thus, we obtain an approximated triangular mesh model M that roughly represents the object to be modeled.

2.2 Quality predictor estimation

Second, we evaluate the quality predictors of the geometry of a final dense model from a set of sparse points P and camera poses

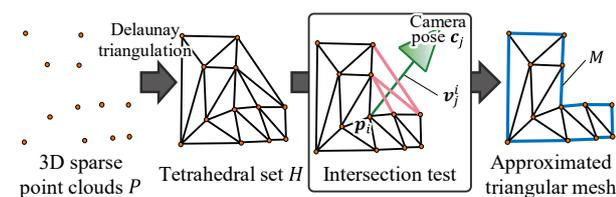


Figure 2. The process of the approximated triangular mesh model generation.

E obtained by the SfM and approximated triangular mesh model M . The predictors estimate how poorly a local region geometry close to a sparse point would be reconstructed in a final dense model by MVS. The basic concept of the predictors was first proposed by (Mauro, 2014); however, we recently introduced two predictors to improve effectiveness and efficiency in quality prediction.

The quality predictors are calculated at each sparse point $i(i \in P)$ on the approximated triangular mesh model M . Before estimating the predictors, we normalize the size of the approximated mesh model to produce an average distance \bar{R} , equal to 1, among the nearest neighbor points on the mesh model. As Figure 3 shows, six quality predictors [(1)–(4) (Mauro, 2014) and (5)–(6) (proposed in this paper)] are evaluated at a sparse point i , and their values are assigned to i :

(1) Density (F_D)

The quality of the final dense model is likely to degrade on the region where the sparse point density is low. Therefore, we evaluated as density $F_D(i)$ (see Eq. (1) below) the number of neighboring sparse points around a sparse point i within a specified radius.

$$F_D(i) = |B_i|; \quad B_i = \{j | d(i, j) \leq r_1, j \in P, i \neq j\}, \quad (1)$$

where r_1 denotes a neighbor radius for the density evaluation of the sparse points specified by the user, and $d(i, j)$ the distance between sparse points i and j .

(2) Uncertainty (F_U)

The depth accuracy of a dense point obtained by the stereo-matching in the MVS process decrease when the angle between two different rays v_j^i and v_k^i , directed from a camera position \mathbf{c}_j and \mathbf{c}_k toward a sparse point i , becomes too small. Therefore, the maximum angle among all pairs of rays is evaluated by the following equation (Eq. (2)) below as Uncertainty $F_U(i)$:

$$F_U(i) = \max_{v_j^i, v_k^i \in V_i} \angle(v_j^i, v_k^i). \quad (2)$$

(3) 2D-saliency (F_{2D})

If a local region around the pixel on an image j to which a sparse point i is re-projected is poorly textured, the stereo-matching in the MVS tends to fail. Therefore, we evaluate the average image gradient at the re-projected pixel as 2D-saliency $F_{2D}(i)$ by the following equation (Eq. (3)) below:

$$F_{2D}(i) = \frac{1}{|N_{S_i}|} \frac{1}{|S_i^j|} \sum_{j \in N_{S_i}} \sum_{k \in S_i^j} \|\nabla(k)\|, \quad (3)$$

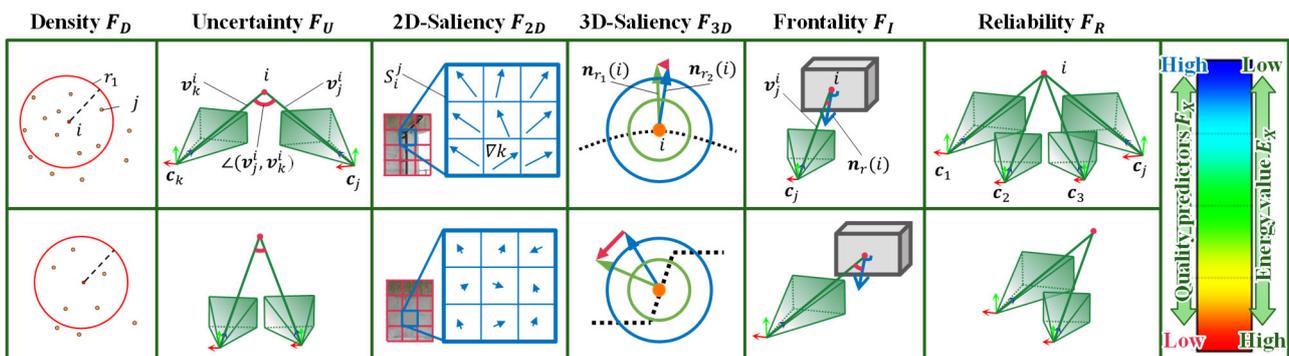


Figure 3. Six quality predictors and the relation between F_X and E_X .

where S_i^j denotes the set of pixels in a squared region centered at the pixel on the image j where a sparse point i is re-projected onto j , N_{S_i} a set of images from which the sparse point i can be seen, and $\nabla(k)$ the image intensity gradient centered at a pixel k .

(4) **3D-saliency (F_{3D})**

If the local surface of an object to be modeled has a more complex geometry, more images are needed to reconstruct it. Hence, we estimate the geometric complexity of the final dense model from the Difference of Normals (DoN) (Ioannou, 2012) as a 3D-saliency $F_{3D}(i)$ at a sparse point. The DoN at i is defined by Eq. (4) as the difference of the normal $\mathbf{n}_r(i)$ for different neighbor radii:

$$F_{3D}(i) = \|\mathbf{n}_{r_1}(i) - \mathbf{n}_{r_2}(i)\|, \quad (4)$$

where $\mathbf{n}_r(i)$ is the unit normal vector at a sparse point i which is calculated by the principal component analysis for the local-neighbor sparse point set included in a sphere of a radius r centered at i .

(5) **Frontality (F_f)**

The geometry reproducibility of the final dense model tends to decrease as further apart the images are taken from the right in front of the object surface. We evaluate the geometry as an averaged incident angle between a surface normal vector $\mathbf{n}_r(i)$ at a sparse point on M , and a ray vector \mathbf{v}_j^i from a camera pose j to i and express it as a frontality $F_f(i)$ defined by the Eq. (5) below:

$$F_f(i) = \frac{1}{|V_i|} \sum_{v_j^i \in V_i} \angle(\mathbf{v}_j^i, \mathbf{n}_r(i)). \quad (5)$$

(6) **Reliability (F_R)**

The reliability of the geometry estimation diminishes as the number of visible cameras $|V_i|$ supporting a sparse point i decreases. Therefore, it is evaluated by Eq. (6) below as reliability $F_R(i)$:

$$F_R(i) = |V_i|. \quad (6)$$

2.3 Calculation of Geometry degradation value E_{deg}

To calculate the geometry degradation value, first, we convert each of the six quality predictors given by Eqs. (1)–(6) to normalized energy $\in [0,1]$ by using the following logistic function:

$$E_X(i) = \begin{cases} L(F_X - \mu_X, \sigma_X), & X \in \{3D, I\}; \\ 1 - L(F_X - \mu_X, \sigma_X), & X \in \{D, U, 2D, R\}, \end{cases} \quad (7)$$

where, μ_X denotes the average of F_X , σ_X the standard deviation of F_X , and $L(x - \mu, \sigma) = 1 / (1 + \exp(-\frac{2(x-\mu)}{\sigma}))$. Then, the six energy values E_X are aggregated by taking an average to denote a *geometry degradation indicator* at a sparse point i as $E_{deg}(i)$ using Eq. (8) below:

$$E_{deg}(i) = \frac{1}{6} (E_D(i) + E_U(i) + E_{2D}(i) + E_{3D}(i) + E_I(i) + E_R(i)). \quad (8)$$

Therefore, a region with high-energy $E_{deg}(i)$ on the approximated triangular mesh model M indicates that there is

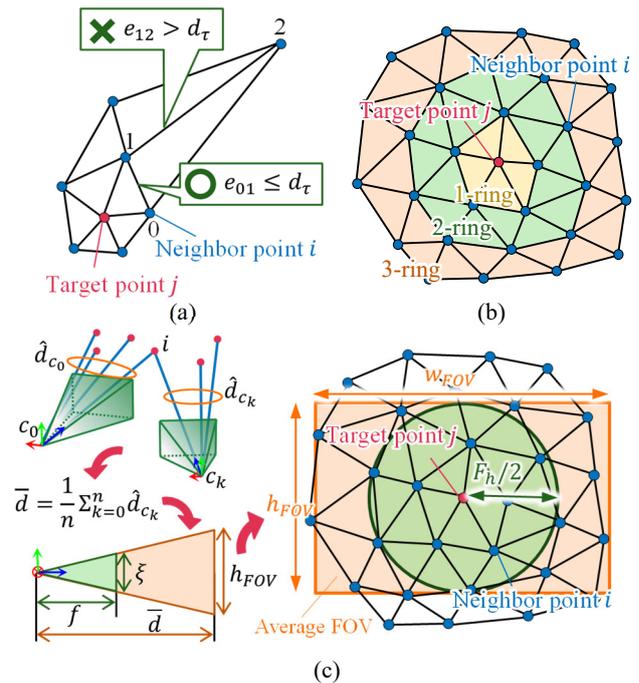


Figure 4. Selection conditions of the neighbor point i ; (a) the edge length (condition 1), (b) selection using the k -ring region of i (condition 2a), and (c) selection using the FOV height of an existing camera (condition 2b).

more possibility of degrading the geometry around a local region close to the sparse point i on the final dense model and that additional images are required to adequately reconstruct the 3D geometries around the region.

3. LOW-QUALITY REGION SELECTION

After finding the quality predictors and geometry degradation indicator, there is a need to improve the estimated low-quality regions by taking additional images efficiently. In the additional image capture, it is desirable that an image should be taken to include as large low-quality regions as possible. To completely determine the best pose of additional image capture, both of camera position and shooting target point which is an intersection between the optical axis of the camera and the surface of the approximated triangular mesh model M must be specified. Between the two, the *shooting target point* affects more the best new image capture pose. Therefore, in this research, we develop the following algorithm to automatically select the best shooting target point for the next additional image acquisition by solving an optimization problem based on the estimated geometry degradation indicator.

First, we assign a sparse point $i (\in P)$ a weight w_i equal to its geometry degradation indicator value $E_{deg}(i)$. Then, from the $E_{agg}(i)$ of a target point $j (\in P)$ and the other sparse points i included in a local region adjacent to j on M , we predict the degree of geometric degradation of the region adjacent to j on the final dense model, and we select s images which preferentially capture those regions with a higher degree of geometric degradation.

In the algorithm, these s shooting target points are selected from sparse point clouds P as a solution of the following optimization problem defined by Eqs. (9)–(12) below:

$$\begin{aligned} & \underset{\{x_j\}, \{z_i\}}{\text{maximize}} && \sum_{j \in P} \left(\frac{(\sum_{i \in P} w_i z_i) + w_j x_j}{(\sum_{i \in P} z_i) + x_j} \right) && (9) \\ & \text{subject to} && z_i \leq \sum_{j \in P} a_{ij} x_j && (10) \\ & && \sum_{j \in P} z_j \geq 2 && (11) \\ & && \sum_{j \in P} x_j = s && (12) \\ & && x_j \in \{0, 1\} \ (\forall j \in P) \\ & && z_i \in \{0, 1\} \ (\forall i \in P) \end{aligned}$$

where, the indicator variables x_j and z_i represent

$$x_j = \begin{cases} 1 & \text{if a sparse point } j \text{ is adapted as} \\ & \text{the target point,} \\ 0 & \text{if } j \text{ is not adapted as a target point.} \end{cases} \quad (13)$$

$$z_i = \begin{cases} 1 & \text{if a sparse point } i \text{ is covered as} \\ & \text{neighbor points of a target point } j, \\ 0 & \text{if } i \text{ is not covered as neighbor points of } j. \end{cases} \quad (14)$$

The binary parameter a_{ij} of Eq. (10) shows whether a sparse point i must be included in the neighbor points of a target point j or not when j is selected as a target point, and is defined as follows:

$$a_{ij} = \begin{cases} 1 & \text{if sparse points } i \text{ and } j \text{ satisfy the following} \\ & \text{neighborhood conditions 1) and 2) below,} \\ 0 & \text{otherwise.} \end{cases} \quad (15)$$

The neighborhood selection conditions 1) and 2) are respectively defined as the following:

- Condition 1): As Figure 4(a) shows, the condition holds if a sparse point i is adjacent to j on M and the edge length between points i and j is less than or equal to the threshold τ_d . This condition implies that the geometry degradation indicators of the adjacent points from the adjacent point candidates should be removed when the distances from them to j are significantly longer. Here, we derive the threshold τ_d as $\tau_d = \mu + 3\sigma$, where μ denotes the average and σ the standard deviation of all edge lengths on M .

On the other hand, the condition 2) consists either of the condition 2a) that uses the topological connectivity of the approximated triangular mesh model, or the condition 2b) that employs a field of view (FOV) similar to that of the existing camera.

- Condition 2a): As Figure 4(b) shows, this condition holds if a sparse point i exists inside the k -ring of a target point j on M . The experimental result in the next section describes the appropriate selection of k .
- Condition 2b): As Figure 4(c) shows, this condition holds if the distance from a target point j to the neighbor point i is shorter than or equal to the half of the average FOV height (h_{FOV}) of the existing cameras. For this selection, we first calculate the average distance \bar{d}_{c_k} from a camera position \mathbf{c}_k to its visible sparse points. Then, the average distance \bar{d} for all \bar{d}_{c_k} of all the cameras is evaluated, and the height h_{FOV} of the average FOV is calculated from a focal length f and a sensor size ξ by $h_{FOV} = \bar{d}\xi/f$. The condition 2b) holds if the point i exists within the sphere centered at the target point j with the radius $h_{FOV}/2$, and the point i and j are connected topologically on M .

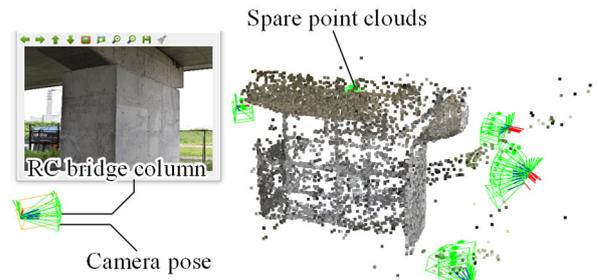


Figure 5. Sparse point clouds and camera poses of the RC bridge column.

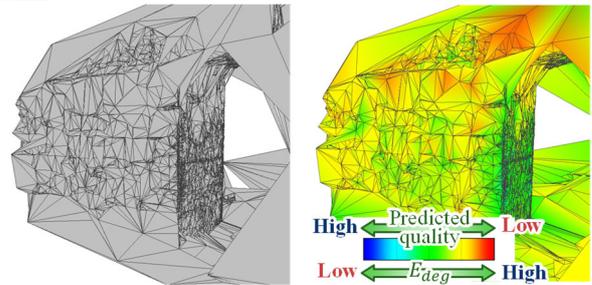


Figure 6. The approximated triangular mesh model.

Figure 7. The estimated geometry degradation indicator value.

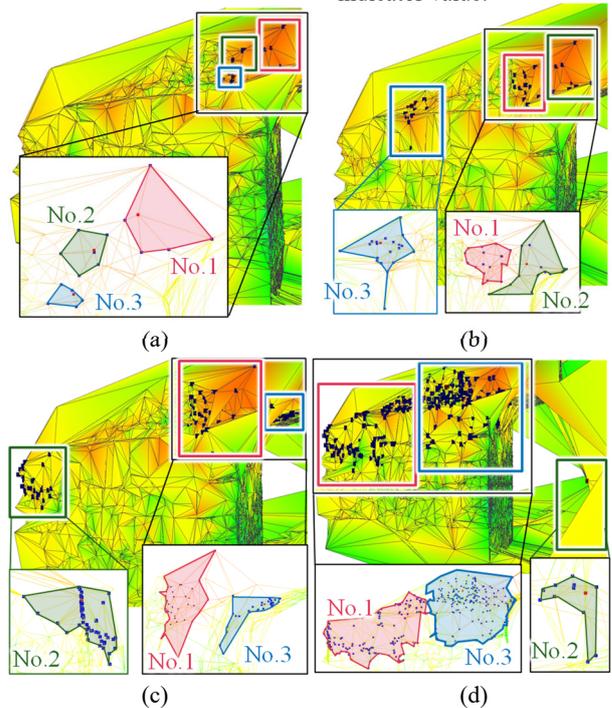


Figure 8. The selected low-quality regions; (a) $k = 1$ by condition 2a, (b) $k = 2$ by condition 2a, (c) $k = 3$ by condition 2a, and (d) using the average FOV height by condition 2b. A red dot indicates a shooting target point.

The objective function of Eq. (9) describes the average value of the geometry degradation indicators at the target point j and its selected neighbor points i that satisfies condition 1) and 2). The constraint condition of Eq. (10) implies that there must be at least one shooting target point j which can cover a sparse point i so that the point i is covered as the neighbor point of j . Furthermore, Eq. (11) describes the constraint condition that a target point j must cover at least two neighbor points. Moreover Eq. (12)

indicates that s images are added at a time in the next additional shooting.

In this paper, the approximately optimum solution of the optimization problem (9)–(12) is obtained by the greedy algorithm, and s near-optimum target points for the next additional image shooting are determined.

4. EXPERIMENTS

We applied the proposed method to SfM/MVS reconstruction from the image set of a concrete bridge column and construction site. A few tens of initial input images were manually selected from the original image set of the target object which includes a few hundred images. From this initial input images, the proposed method automatically recommended a couple of shooting target points for the next additional images. Then, based on the target points, we manually selected an additional image from the original image set so that it includes as many as possible of the recommended shooting target points and used them for MVS process in addition to the initial input images. Finally, we evaluated how the additional images improved the geometric quality of the final dense mesh model generated from MVS.

4.1 Reinforced concrete bridge column

We employed the proposed quality prediction and low-quality region selection method on the image set obtained from reinforced concrete (RC) bridge column ($7.0\text{ m} \times 1.9\text{ m} \times 5.0\text{ m}$). The original image set included 110 photos. First, we manually selected 33 initial images from the original set and used a commercial SfM/MVS software (ContextCapture, Bentley Co. Ltd.) to generate sparse point clouds and camera poses. Furthermore, using the software, we predicted the quality of the final textured mesh model, selected a few low-quality regions on the approximated triangular mesh model, and obtained first, second, and third-best shooting target points corresponding to the low-quality regions. Next, we selected an image from the initial image set so that it includes as many as possible of the shooting target points and added it to the original image set. Finally, we compared the geometric quality between the final textured mesh model generated from the initial image set and that from the image set consisting of the initial and additional images.

Figure 5 shows the camera poses and sparse point clouds generated by SfM. The sparse point clouds included 12,070 points. Figure 6 shows the approximated triangular mesh model generated from the sparse point clouds, and Figure 7 shows the distribution of the geometry degradation indicator E_{agg} constructed from the 33 initial images in the case that the neighbor radius $r_1 = 10\bar{R}$ and $r_2 = 20\bar{R}$. As Figure 7 shows, the degradation indicator E_{agg} significantly decreased on the region bounded in the rectangles, and it suggested that the local geometries of the final textured mesh model corresponding to these regions would be low-quality in the approximated triangular mesh.

We specified the number of the additional shooting target points (additional images) at a time as $s = 3$. Figure 8(a)–(d) show the automatically-selected three low-quality regions using condition 2a); the k -ring setting for $k = 1, 2$, and 3, and condition 2b); the average FOV height, respectively. It was found that the geometry degradation indicator around the selected target points is higher compared to those obtained in other regions; hence, the optimization algorithm correctly selected the low-quality regions. Moreover, as Figure 9(a) shows, the final textured mesh model



Figure 9. The result of the final textured mesh model generated from MVS; (a) before adding the image and (b) after adding the image.

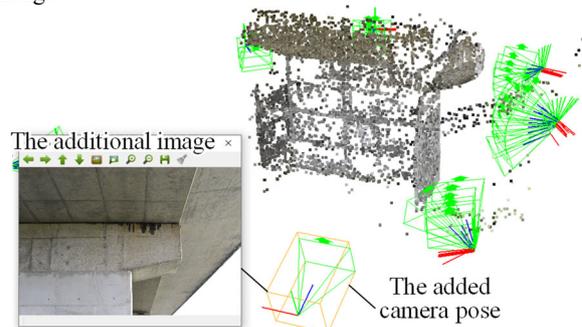


Figure 10. The additional image and the added camera pose.

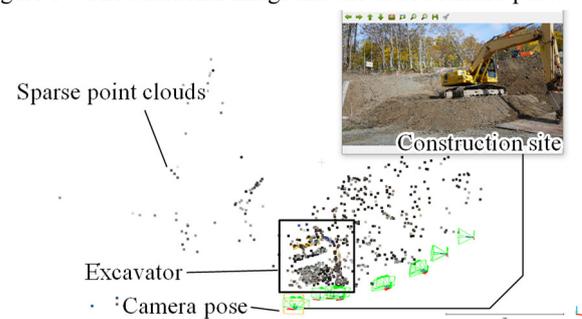


Figure 11. Sparse point clouds and camera poses of the construction site.

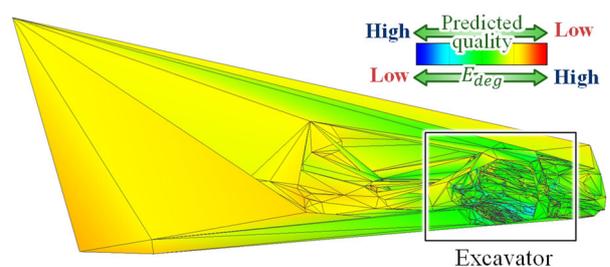


Figure 12. The estimated geometry degradation indicator value.

generated from MVS lacked a major portion corresponding to the estimated low-quality regions.

Based on the recommended shooting target points shown in Figure 8(c), we manually selected an additional image from the initial image set so that it includes as many as possible of the first, second, and third-best shooting target points. Figures 10 and 9(b) show the selected additional image and the generated final textured mesh model respectively. It was clearly shown that, even when only one image was added to the image set, the geometry of the final textured mesh model was considerably improved. Therefore, based on the SfM results, we confirmed that the proposed quality prediction and low-quality region selection method is effective for estimating and improving the geometric qualities of a final dense model.

For 33 images in the initial image set, the processing time of the approximated triangular mesh model generation took 3.6 s, quality predictor estimation 14.1 s, the selection of the three low-quality regions 0.7 s (condition 2a), and 2.4 min (condition 2b), SfM 71 s, and MVS 22 min. Therefore, the proposed method allows for much shorter quality estimation time of the final dense model than the MVS process.

4.2 Construction site

The proposed method was also applied to the image set of a construction site ($50 \times 30m$) consisting of about 600 images. Similar to the previous experiment, we manually selected 15 initial images from the original image set. As Figure 11 shows, the sparse point clouds and camera poses were generated from SfM. Figure 12 shows the approximated triangular mesh model M and the estimated geometry degradation indicator value. Figures 13(a) and (b) show the low-quality region selection results when using the k -ring ($k = 3$) of the condition 2a) and the average FOV height of the condition 2b), respectively. It was found that regardless of the conditions, the geometry degradation indicator value increases in the selected region on M . Indeed, as Figure 14(a) shows, by using the initial image set only, the final textured mesh model was imperfectly generated and included many holes on the regions corresponding to those with high geometry degradation indicator value. Furthermore, as Figure 15 shows, we selected and added an image which includes the second-best shooting target point (No. 2) among the three estimated as low-quality regions in Figure 13(a). Figure 14(b) shows the final textured mesh model obtained by MVS. By adding only one image, the holes that originally appeared in the mesh model decreased and the reconstructed region significantly increased.

In applying 15 images in the initial image set, the processing time of the approximate triangular mesh model generation took 0.17 s, quality predictor estimation 3.5 s, the selection of the three low-quality region 0.07s (condition 2a) and 0.4 s (condition 2b), SfM 33 s, and MVS 10 min. Therefore, the proposed method could verify the quality of the final textured mesh model much shorter than the MVS process.

5. CONCLUSIONS

In this paper, a method for predicting the quality of the final dense model generated by SfM and MVS process was proposed to realize a high-quality and efficient as-is model reconstruction. Our algorithm was applied to an image set of a concrete bridge column under six quality predictors, and we found that these estimate the geometric quality of the final dense point clouds or textured mesh models. We also proposed a method to automatically select the low-quality regions using an optimization problem based on the quality predictors. Experimental results revealed that the region geometry which had been estimated as low-quality one is eventually improved after the image enclosing the low-quality region was added. Besides, the estimation processing period via SfM was significantly shorter than through MVS.

Regarding future research, we will extend the proposed method to automatically select the next-best camera position where the selected low-quality regions can be efficiently covered while respecting the surface sampling distance constraints. Application of a similar technique to image selection will also be considered. Preliminary image segmentation based on machine learning to

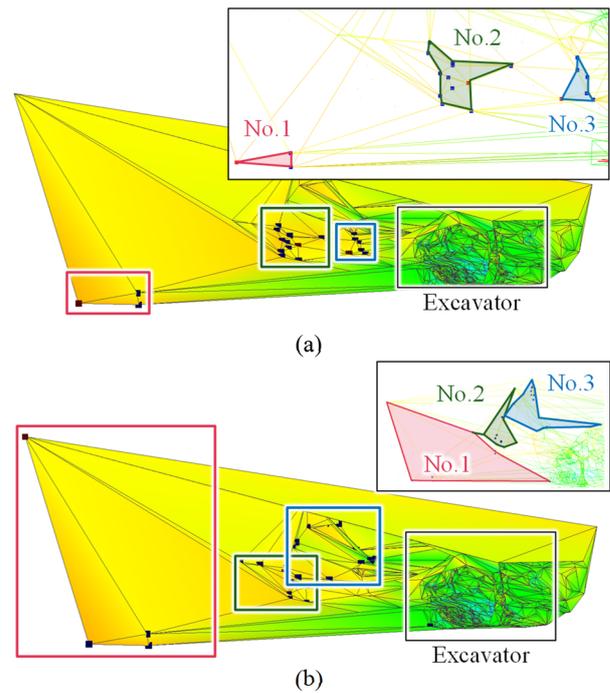


Figure 13. The selected low-quality regions; (a) using the k -ring (condition 2a) and (b) using the average FOV height (condition 2b). A red dot indicates a shooting target point.



Figure 14. The result of the final textured mesh model; (a) before adding an image and (b) after adding an image.

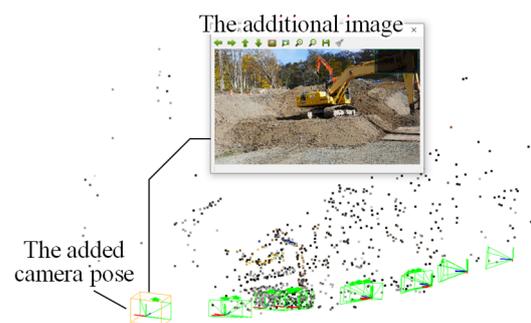


Figure 15. The additional image and the added camera pose.

eliminate background portion in the original image set will also be considered to make the quality prediction more precise.

ACKNOWLEDGEMENTS

Part of this study is financially supported by the Ministry of Land, Infrastructure, Transport and Tourism under the project titled "Introduction and utilization of innovative technology to dramatically improve the productivity of construction sites" (Test work name: general national highway route.239 (representative: Horiguchi-gumi Co., Ltd.)). The image set of the construction site was provided by Kankyo-Fudo Techno Co., Ltd.

REFERENCES

- Bentley Systems, ContextCapture.
<https://www.acute3d.com/applications/>
- Hepp, B., Nießner, M., Hilliges, O., 2017. Plan3D: Viewpoint and Trajectory Optimization for Aerial Multi-View Stereo Reconstruction. *ACM Transactions on Graphics* 38, 4. doi:10.1145/3233794
- Hoppe, C., Wendel, A., Zollmann, S., Pirker, K., Irschara, A., Bischof, H., Kluckner, S., 2012. Photogrammetric Camera Network Design for Micro Aerial Vehicles. *Computer Vision Winter Workshop (CVWW)* 8, 1–3.
- Ioannou, Y., Taati, B., Harrap, R., Greenspan, M., 2012. Difference of Normals as a Multi-scale Operator in Unorganized Point Clouds, in: *2012 Second International Conference on 3D Imaging, Modeling, Processing, Visualization & Transmission*, 501–508. doi:10.1109/3DIMPVT.2012.12
- Jing, W., Polden, J., Tao, P.Y., Lin, W., Shimada, K., 2016. View planning for 3D shape reconstruction of buildings with unmanned aerial vehicles, in: *2016 14th International Conference on Control, Automation, Robotics and Vision (ICARCV)*, 1–6. doi:10.1109/ICARCV.2016.7838774
- Labatut, P., Pons, J.-P., Keriven, R., 2007. Efficient Multi-View Reconstruction of Large-Scale Scenes using Interest Points, Delaunay Triangulation and Graph Cuts. *2007 IEEE 11th International Conference on Computer Vision* 1–8. doi:10.1109/ICCV.2007.4408892
- Martin, A.R., Rojas, I., Franke, K., Hedengren, D.J., 2016. Evolutionary View Planning for Optimized UAV Terrain Modeling in a Simulated Environment. *Remote Sensing*, 26. doi:10.3390/rs8010026
- Mauro, M., Riemenschneider, H., Signoroni, A., Leonardi, R., Van Gool, L.J., 2014. A Unified Framework for Content-Aware View Selection and Planning Through View Importance, in: *British Machine Vision Conference (BMVC)*, 1–11. doi:10.5244/C.28.69
- Roberts, M., Shah, S., Dey, D., Truong, A., Sinha, S., Kapoor, A., Hanrahan, P., Joshi, N., 2017. Submodular Trajectory Optimization for Aerial 3D Scanning, in: *Proceedings of the IEEE International Conference on Computer Vision*, 5334–5343. doi:10.1109/ICCV.2017.569
- Schmid, K., Hirschmüller, H., Dömel, A., Grixia, I., Suppa, M., Hirzinger, G., 2012. View Planning for Multi-View Stereo 3D Reconstruction Using an Autonomous Multicopter. *Journal of Intelligent & Robotic Systems*, 309–323. doi:10.1007/s10846-011-9576-2