Proceedings of the 21st ICSMGE, Vienna, Austria, 14 – 19 June 2026. Pistrol, Adam & Schweiger (eds.)
Published by: ÖGG, Austrian Society for Geomechanics, Salzburg, Austria, ISBN 978-3-9503898-4-5
https://doi.org/10.53243/ICSMGE2026-1711

A Comparative Study of Deep Learning and Photogrammetry for 3D Rock Slope Reconstruction

Yurok Choi, Young Seok Na, Eugene Ee, Tae Sup Yun
School of Civil and Environmental Engineering, Yonsei University, Seoul, South Korea, taesup@yonsei.ac.kr

ABSTRACT: Three-dimensional (3D) models of rock slopes are key inputs to the evaluation of rock-slope stability. This study compares the performance of a deep learning–based multi-view reconstruction network, MUSt3R (Multi-view Network for Stereo 3D Reconstruction), and a photogrammetric pipeline, SfM–MVS (Structure-from-Motion and Multi-View Stereo), on the rock slope of Inwangsan, Seoul. The image dataset comprised a low-density set of 7 frames extracted from a fixed-angle video with 80% overlap, and a high-density set of 61 photos captured from diverse viewpoints using a smartphone. Applying both methods yielded four models; we then conducted visual assessments and quantitative comparisons of accuracy, number of reconstructed points, and reconstruction time. Under the high-density condition, both methods reproduced overall geometry, joints, and surface undulations without distortion; however, MUSt3R better reconstructed soil cover, vegetation, and fine textures, whereas SfM–MVS exhibited localized blur artifacts. Under the low-density, fixed-angle condition, SfM–MVS reconstructed the global geometry more stably, whereas MUSt3R produced realistic soil/vegetation but showed biases in overall geometry and diminished fine-scale surface detail. Quantitatively, using the high-density SfM–MVS model as the reference for Iterative Closest Point (ICP) registration, the high-density MUSt3R model achieved an RMSE of 0.0204 m, approximately 30% lower than the 0.0288 m of the low-density SfM–MVS model, whereas the low-density MUSt3R model failed to converge. In terms of efficiency, MUSt3R outperformed SfM–MVS: with refinement iterations fixed at 10, MUSt3R required 0.25–0.5 times the reconstruction time and generated 4–10 times more points. In conclusion, when sufficient viewpoint diversity and image density are ensured, MUSt3R maintains overall accuracy while offering superior detail reproduction and computational efficiency; when viewpoint diversity is limited, the reconstruction stability of SfM–MVS is comparatively advantageous.

KEYWORDS: Rock slope, 3D reconstruction, Photogrammetry, Deep learning, Point cloud

1 INTRODUCTION

Rock-slope failures can cause substantial economic losses and casualties, posing significant threats to public safety and infrastructure integrity, and are therefore a critical engineering concern (Froude and Petley, 2018). The stability of a rock slope is governed by the strength of the rock mass, the development of discontinuities (e.g., joints and faults), the geologic framework, and prevailing hydrogeological conditions. In particular, the geometry of discontinuities is a primary determinant of failure mode (Hoek and Bray, 1981). Accordingly, rigorous evaluation of rock-slope stability requires acquisition and analysis of structural data for the slope.

Traditional field measurements along clinometer-defined scanlines and direct-contact readings of discontinuities have inherent limitations: (i) portions of the slope may be inaccessible, (ii) the procedure poses safety risks, and (iii) operator-dependent steps introduce human error (Slob, 2010). Consequently, the resulting discontinuity dataset may not be statistically representative of the entire slope.

Laser scanning has been widely adopted to address these issues. Early studies using airborne laser scanning (ALS) acquired slope-scale topography and documented cases of rock-slope instability (Haugerud et al., 2003). With the subsequent adoption of terrestrial laser scanning (TLS), numerous studies reported high-precision, high-density 3D models and quantitative discontinuity analyses (Lato and Vöge, 2012; Abellán et al., 2014). Nevertheless, TLS campaigns entail substantial equipment and operational costs, are time consuming, and are vulnerable to vegetation-induced occlusions and mixed returns. Moreover, because observations are made from a limited number of fixed stations, complex outcrops frequently exhibit line-of-sight blockages that create shadow zones, making complete and reliable data acquisition difficult (Soma et al., 2020).

Advances in computer vision and the widespread use of unmanned aerial vehicles (UAVs) have made multi-view, image-based photogrammetry increasingly prevalent (Salvini et al., 2017). Structure-from-Motion (SfM) estimates camera poses and the 3D structure of a scene from images captured from different viewpoints, yielding a sparse point cloud. Multi-View Stereo (MVS) then exploits inter-view parallax, together with the recovered camera parameters, to produce a dense point cloud. However, photogrammetry is sensitive to texture and illumination variations; it requires sufficient viewpoint diversity and high image overlap; and feature detection and matching often fail in low-contrast or low-texture regions, on specular or transparent surfaces, and in areas with vegetation. In addition, as the number of input images increases, the computational cost of pairwise image matching and bundle adjustment (BA) grows nonlinearly, leading to substantial memory use and processing time (Heinly et al., 2015).

To overcome these limitations, deep-learning-based 3D reconstruction has gained significant attention as a promising alternative capable of addressing key weaknesses in traditional methods**.** Unlike photogrammetry, approaches that exploit learned geometric and photometric priors can infer structure from sparse views and remain robust to low texture or contrast, illumination changes, and partial occlusions (e.g., vegetation). By integrating information across views at a global level, these methods encourage view-consistent recovery of surfaces and depth, while learned regularization suppresses noise and matching artifacts. Operating in a learned feature space also mitigates the computational burden of largescale pairwise matching and bundle adjustment during inference, improving efficiency. Taken together, deep-learning-based methods enable more faithful and stable reconstruction of outcrop surfaces, even in complex scenes with weak texture and non-uniform lighting (Yao et al., 2018).

This study presents a controlled comparison between a deep learning–based 3D reconstruction method and classical photogrammetry (SfM–MVS) using imagery of the same rock slope. We constructed two datasets: a high-density, multi-angle set and a low-density, fixed-angle set. Performance was evaluated by visual inspection and by quantitative metrics, including root-mean-square error (RMSE) from Iterative Closest Point (ICP; Besl and McKay, 1992), point cloud density, and computational cost. The results clarify the

strengths and limitations of each approach, identify their effective domains of use, and offer practical guidance for rock-slope reconstruction.

2 STUDY SITE AND DATASETS

2.1 *Study site*

The study site was a rock slope on Inwangsan in Seoul (Figure 1). Within this slope, the area of interest (AOI) measured approximately 3 m in height and 7 m in width. A relatively low rock face was selected to enable sufficient image capture using a smartphone. Two datasets were collected to compare reconstruction methods under different acquisition conditions: a low-density, fixed-angle dataset and a high-density, multi-angle dataset.

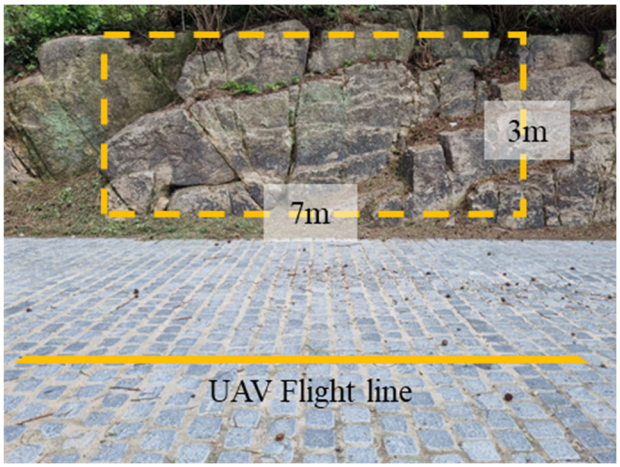

Figure 1. Study site showing AOI dimensions and UAV flight line

2.2 *Datasets*

2.2.1 *Low-density, fixed-angle video frame set*

As illustrated in Figure 2, a video-based frame set was generated to evaluate the methods under low-density, fixed-angle conditions. A UAV recorded video from a constant altitude and fixed viewing angle oriented toward the slope face. All frames were processed to extract Scale-Invariant Feature Transform (SIFT) features (Lowe, 1999); feature matching and homography estimation were then performed using the Random Sample Consensus (RANSAC) algorithm (Fischler and Bolles, 1981). Frames meeting the target overlap were retained, and the selection procedure was iterated to obtain seven images (3840 × 2160 pixels) with approximately 80% overlap.

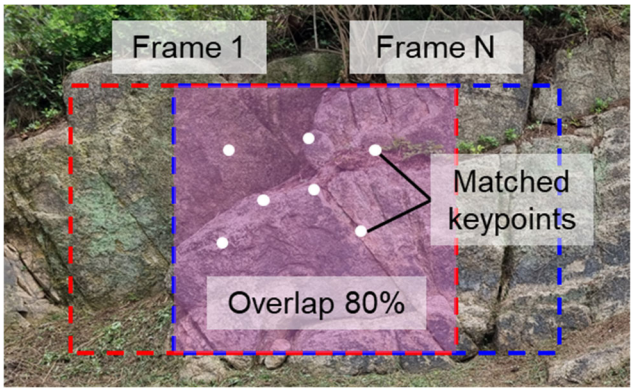

Figure 2. Video-based frame selection using SIFT and RANSAC

2.2.2 *High-density, multi-angle smartphone image set*

A smartphone-based multi-view image set was acquired to evaluate the methods under dense, multi-angle conditions. The target rock face was imaged from diverse viewpoints to secure wide-baseline variation, with the intended overlap between adjacent views set to at least 80%. Data were captured with a Samsung Galaxy S21+ smartphone, yielding 61 images at 4032 × 3024 pixels.

3 METHODS

3.1 *Reconstruction pipelines*

Photogrammetric 3D reconstruction used the open-source Meshroom (AliceVision; Griwodz et al., 2021), which implements an SfM–MVS pipeline to automatically estimate camera poses, generate dense point clouds, and reconstruct meshes from the input images. For the deep-learning pipeline, we adopted the Multi-view Network for Stereo 3D Reconstruction (MUSt3R; Cabon et al., 2025). a state-of-the-art model selected for its demonstrated ability to process large, unstructured image sets while maintaining global consistency. MUSt3R was implemented in PyTorch 2.7.0 with CUDA 12.6, with the number of refinement iterations fixed at 10 for all runs to ensure a consistent basis for comparison. All experiments were executed on Windows 10 with an Intel Core i9-11900 CPU and an NVIDIA GeForce RTX 3060 GPU.

3.2 *Evaluation metrics*

The performance of the four generated models was evaluated through both visual inspection and quantitative metrics. For quantitative accuracy, 50,000 points were randomly sampled from each reconstructed point cloud. Each sample was registered to the high-density SfM–MVS model, which served as the reference dataset for this relative comparison using point-to-point Iterative Closest Point (ICP; Besl and McKay, 1992), after which a nearest-neighbor root-mean-square error (RMSE) was computed. Computational efficiency was defined as the number of reconstructed points per second of reconstruction time (points/s).

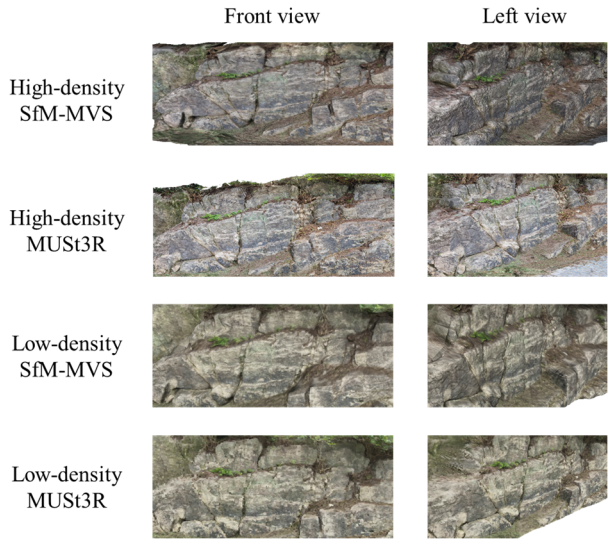

Figure 3. Reconstruction results of rock slope using SfM–MVS and MUSt3R

4 RESULTS

4.1 *Visual comparison*

As shown in Figure 3, in the high-density dataset, both methods produced complete models that clearly represented joints and undulations on the rock surface; however, unlike MUSt3R, SfM–MVS failed to accurately reconstruct soil and vegetation, and foliation on the rock surface appeared blurred. In the low-density dataset, SfM–MVS preserved the global geometry at a level comparable to the high-density results, but local undulations and joints were either exaggerated or blurred, and the depiction of soil and vegetation was inaccurate. By contrast, MUSt3R yielded a realistic appearance for soil and vegetation, yet the global geometry was distorted and fine-scale microtopography was not faithfully captured.

4.2 *Quantitative comparison*

4.2.1 *Geometric accuracy (RMSE)*

The high-density MUSt3R model yielded the lowest RMSE of 0.0204 m, while the low-density SfM–MVS model recorded 0.0288 m. This corresponds to an approximately 29.2% reduction in mean error for high-density MUSt3R relative to low-density SfM–MVS. By contrast, the low-density MUSt3R case was classified as non-convergent because the iterative refinement failed to meet the convergence criterion (Table 1).

Table 1. ICP-derived RMSE relative to the high-density SfM–MVS reference

3D Model	High-density MUSt3R	Low-density SfM–MVS	Low-density MUSt3R
RMSE (m)	0.0204	0.0288	Not converged

4.2.2 *Computational efficiency*

Under the high-density setting, MUSt3R produced 7.4 times more points while requiring 0.55 times the reconstruction time (a 45% reduction) of SfM–MVS, yielding approximately 13.4 times higher efficiency. Under the low-density setting, it produced 4.2 times more points and required 0.24 times the reconstruction time (a 76% reduction), resulting in approximately 17.4 times higher efficiency (Table 2).

Table 2. Point counts and reconstruction times for SfM–MVS and MUSt3R

3D Model	High-density		Low-density	
	SfM–MVS	MUSt3R	SfM–MVS	MUSt3R
Number of points	780,000	5,770,000	76,000	320,000
Reconstruction time (s)	1287	710	87	21

5 DISCUSSION

The visual and quantitative results reveal a clear trade-off between the two methods contingent on data acquisition conditions. The observed differences can be attributed to their distinct methodological characteristics. MUSt3R, as a deep learning–based approach, directly predicts 3D point maps and camera poses. When observations are sparse, it relies on pre-trained statistical priors to plausibly regularize textures and vegetation. This explains its realistic appearance for soil and vegetation even in the low-density case, but also makes it susceptible to distortions in global geometry when geometric constraints are weak. In contrast, SfM–MVS enforces strong global consistency through feature matching, bundle adjustment, and triangulation. This preserves the global shape comparatively well even at low densities, but its reliance on local feature matching tends to produce blurring and omissions in regions with poor texture or sparse viewpoints.

The quantitative results support these observations. The superior RMSE of the high-density MUSt3R model indicates that, when provided with sufficient multi-view data, deep learning methods can preserve fine-scale surface attributes (e.g., surficial materials, vegetation, and micro-relief) while maintaining global shape consistency with the reference model. However, the non-convergence of the low-density MUSt3R case highlights its instability when viewpoint diversity is limited. The larger point counts generated by MUSt3R in less time further suggest its potential to capture fine-scale elements that SfM–MVS often struggles to represent, demonstrating a significant efficiency advantage.

Practically, these findings suggest that securing wide baselines and at least 80% image overlap is pivotal for 3D rock slope modeling. When such coverage is achievable, MUSt3R is recommended for its superior depiction of micro-topography and surficial materials while maintaining consistent global geometry and high computational efficiency. However, under constrained viewpoints, the more robust global shape preservation of SfM–MVS makes it the preferable and lower-risk option.

This study has several limitations. The comparison was conducted at a single site and AOI, and used the high-density SfM–MVS model as the reference, i.e., without an absolute ground truth from a higher-accuracy sensor like a terrestrial laser scanner. Experiments were also conducted with a single deep learning model and fixed hyperparameters (10 refinement iterations).

6 CONCLUSION

This study performed a controlled comparison of a deep learning-based method (MUSt3R) and a conventional photogrammetric pipeline (SfM–MVS) for 3D rock slope reconstruction under high-density and low-density imaging conditions. The results demonstrate that when sufficient viewpoint diversity and image density are ensured, MUSt3R maintains global accuracy while offering superior detail reproduction and computational efficiency. Conversely, when image data is sparse and viewpoint diversity is limited, SfM–MVS provides more stable and reliable reconstruction of the global geometry. Future work should incorporate absolute ground truth references (e.g., terrestrial LiDAR) and explore a wider range of deep learning models and parameter settings to broaden the generality of these findings.

7 ACKNOWLEDGEMENTS

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korean government (MSIT) (Nos. RS-2021-NR060085 and RS-2023-NR076991).

8 REFERENCES

Froude, M.J., and Petley, D.N. 2018. *Global fatal landslide occurrence from 2004 to 2016*. Natural Hazards and Earth System Sciences 18(8), 2161–2181.

Hoek, E., and Bray, J.D. 1981. *Rock slope engineering*. CRC Press.

Slob, S. 2010. *Automated rock mass characterisation using 3-D terrestrial laser scanning*. PhD thesis, Delft University of Technology, Zutphen: Wohrmann Print Services.

Haugerud, R.A., Harding, D.J., Johnson, S.Y., Harless, J.L., Weaver, C.S., and Sherrod, B.L. 2003. *High-resolution lidar topography of the Puget Lowland, Washington*. GSA Today 13(6), 4–10.

Lato, M.J., and Vöge, M. 2012. *Automated mapping of rock discontinuities in 3D lidar and photogrammetry models.*

International Journal of Rock Mechanics and Mining Sciences 54, 150–158.

Abellán, A., Oppikofer, T., Jaboyedoff, M., Rosser, N.J., Lim, M., and Lato, M.J. 2014. *Terrestrial laser scanning of rock slope instabilities*. Earth Surface Processes and Landforms 39(1), 80–97.

Soma, M., Pimont, F., Allard, D., Fournier, R., and Dupuy, J.L. 2020. *Mitigating occlusion effects in leaf area density estimates from terrestrial LiDAR through a specific kriging method*. Remote Sensing of Environment 245, 111836.

Salvini, R., Mastrorocco, G., Seddaiu, M., Rossi, D., and Vanneschi, C. 2017. *The use of an unmanned aerial vehicle for fracture mapping within a marble quarry (Carrara, Italy): photogrammetry and discrete fracture network modelling*. Geomatics, Natural Hazards and Risk 8(1), 34–52.

Heinly, J., Schönberger, J.L., Dunn, E., and Frahm, J.M. 2015. *Reconstructing the world in six days.* In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2015), Boston, United States, 7–12 June 2015, pp. 3287–3295. IEEE Computer Society.

Yao, Y., Luo, Z., Li, S., Fang, T., and Quan, L. 2018. *MVSNet: depth inference for unstructured multi-view stereo.* In: Proceedings of the European Conference on Computer Vision (ECCV 2018), pp. 767–783.

Besl, P.J., and McKay, N.D. 1992. *Method for registration of 3-D shapes*. In: Proceedings of Sensor Fusion IV: Control Paradigms and Data Structures, 21–24 April 1992, vol. 1611, pp. 586–606. SPIE.

Lowe, D.G. 1999. *Object Recognition from Local Scale-Invariant Features*. In: Proceedings of the 7th IEEE International Conference on Computer Vision (ICCV 1999), Kerkyra (Corfu), 20–27 September 1999. Vol. 2, pp. 1150–1157.

Fischler and Bolles, R.C., M.A. 1981. *A RANSAC-Based Approach to Model Fitting and Its Application to Finding Cylinders in Range Data*. In: Proceedings of the 7th International Joint Conference on Artificial Intelligence (IJCAI-81), Vancouver, BC, Canada, August 1981. Vol. 2, pp. 637–643.

Griwodz, C., Gasparini, S., Calvet, L., Gurdjos, P., Castan, F., Maujean, B., et al. 2021. *AliceVision Meshroom: an open-source 3D reconstruction pipeline*. In: Proceedings of the 12th ACM Multimedia Systems Conference, 7–11 June 2021, pp. 241–247.

Cabon, Y., Stoffl, L., Antsfeld, L., Csurka, G., Chidlovskii, B., Revaud, J., and Leroy, V. 2025. *Must3r: multi-view network for stereo 3D reconstruction*. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2025), pp. 1050–1060.