Sensors and Materials, Vol. 38, No. 7 (2026) 4277–4290
MYU Tokyo

S & M 4564

Terrain-type-specific Evaluation of Korea Multi-purpose Satellite-3 Stereo Imagery Using Hybrid Digital Elevation Model Refinement Method

Jae Myeong Kim,[1] Ji Yoon Kim,[2*] Kwan Young Oh,[3] and Kwang Jae Lee[3]

[1]College of Interdisciplinary Studies, Department of Urban Planning Engineering, Seokyeong University, 124 Seogyeong-ro, Seongbuk-gu, Seoul 02713, Republic of Korea
[2]Department of Technology and Management, Carbon Ace, 484 Tongil-ro, Seodaemun-gu, Seoul 03628, Republic of Korea
[3]Satellite Application Center, Korea Aerospace Research Institute, 169-84 Gwahak-ro, Yuseong-gu, Daejeon 34133, Republic of Korea

(Received November 28, 2025; accepted July 7, 2026)

Keywords: satellite imagery, DEM refinement, U-Net++, low-rank matrix factorization, deep learning, terrain-stratified DEM evaluation

Generating digital elevation models (DEMs) using high-resolution satellite imagery is an efficient method for acquiring topographic information over vast areas. However, the accuracy of satellite-derived DEMs is affected by terrain-specific error mechanisms: steep topography amplifies geometric distortions, dense urban blocks create severe occlusion artifacts, vegetated nonurban land introduces canopy-related overestimation, and low-texture flat surfaces degrade stereo correspondence. Despite these well-known terrain dependences, in most refinement studies, only area-averaged accuracy metrics, which can obscure substantially different residual error structures across landscape types, are reported. To address this deficiency, in this study, we introduce a terrain-adaptive DEM refinement framework that integrates U-Net++ deep learning with low-rank matrix factorization and evaluates its effectiveness across four distinct terrain categories: mountainous, flat, urban, and nonurban areas. The framework operates through initial digital surface model reconstruction via U-Net, morphological filtering for nonground separation, low-rank decomposition for systematic bias elimination, and U-Net++ refinement for boundary and detail restoration. Validation experiments using Korea Multi-purpose Satellite-3 (KOMPSAT-3) stereo imagery over San Francisco, USA, and multiple Canadian sites reveal terrain-dependent improvement patterns. Urban areas exhibited the most dramatic gains, with root mean square error (*RMSE*) reductions exceeding 40–50% relative to three commercial photogrammetric software packages. Mountainous regions showed consistent improvement despite inherently higher baseline errors, while flat and nonurban terrains demonstrated stable sub-3 m *RMSE* convergence. Results of cross-site comparisons between the US and Canadian test areas suggest that the framework exhibits consistent terrain-specific improvement patterns under the conditions examined in this study.

*Corresponding author: e-mail: aliasgis@gmail.com
https://doi.org/10.18494/SAM6193

ISSN 0914-4935

1. Introduction

Accurate digital elevation models (DEMs) derived from high-resolution optical satellite imagery are fundamental to topographic mapping, infrastructure planning, and environmental analysis. Although satellite stereo imagery provides a cost-effective pathway for constructing DEMs over extensive areas, the resulting elevation products exhibit error characteristics that are strongly dependent on the underlying terrain type. In urban environments, tall structures create severe occlusion and shadow effects, leading to matching failures and spurious elevation spikes. In nonurban and forested regions, vegetation canopy introduces systematic overestimation of true ground elevation. Mountainous terrain amplifies error sources owing to steep slopes that distort the base-to-height ratio, introducing both random noise and systematic biases, whereas flat terrain often lacks the texture required for reliable feature correspondence. Consequently, DEM errors are not uniform but manifest distinctly across heterogeneous land-surface conditions. Conventional satellite DEM generation generally relies on photogrammetric processing based on rational function or on rational polynomial coefficient (RPC) sensor models, block adjustment, dense stereo matching, and subsequent terrain filtering.[1–3] In the case of Korea Multi-purpose Satellite-3 (KOMPSAT-3) imagery, previous studies have shown that DEM accuracy is strongly affected by the RPC adjustment strategy, control information, and post-correction procedure.[4,5] However, these general-purpose workflows often struggle to simultaneously preserve broad terrain structure and suppress localized artifacts when a single processing strategy is applied to heterogeneous landscapes. In recent years, deep learning has substantially improved dense prediction and surface reconstruction in remote sensing. Encoder–decoder architectures such as U-Net[6] and deep residual networks[7] have shown strong performance in image-to-image estimation tasks, including monocular height estimation.[8] U-Net++ further extends U-Net by introducing nested skip pathways and deep supervision, thereby enabling more effective multiscale feature aggregation.[9] More recently, learning-based frameworks have also been applied to multiview stereo satellite reconstruction, demonstrating the potential of deep neural networks in optical satellite-based 3D reconstruction.[10] However, local reconstruction alone does not necessarily resolve scene-wide systematic bias or sparse residual outliers in satellite-derived DEMs. For this reason, low-rank-plus-sparse decomposition provides a useful complementary strategy for separating global systematic structures from localized anomalies.[11] Another limitation of many DEM refinement studies is that performance is commonly summarized using integrated statistics over the entire study area, such as root mean square error (*RMSE*). Although such metrics are useful for overall comparison, DEM assessment studies have shown that outliers and non-normal error distributions can strongly affect the interpretation of elevation accuracy.[12] Recent review studies have also emphasized that DEM accuracy assessment practices remain inconsistent and often depend on study-specific criteria rather than a unified evaluation framework.[13] Therefore, a terrain-aware evaluation framework is needed to clarify where DEM refinement is effective and where significant errors remain. Accordingly, in this study, we investigate a hybrid DEM refinement workflow composed of U-Net-based digital surface model (DSM) reconstruction, simple morphological filter (SMRF)-based nonground removal, low-rank decomposition-based bias correction, and U-Net++-based local refinement using KOMPSAT-3 stereo imagery. Rather than relying solely on integrated accuracy measures, the refined DEMs are evaluated separately for four terrain categories: urban, nonurban, flat, and mountainous areas. Through this terrain-stratified design, we aim to identify class-dependent error reduction patterns, residual bias behavior, and the relative contribution of each processing stage under different terrain conditions.

2. Theoretical Background

The proposed procedure combines morphological filtering, deep learning, and matrix decomposition to address the distinct error sources that are dominant in different terrain environments. In this section, we describe each algorithmic component and explain how its contribution varies across terrain contexts.

2.1 U-Net++ deep learning model

U-Net++ is an encoder–decoder network derived from U-Net.(6–9) By introducing nested skip pathways between the encoder and decoder, it reduces the semantic gap between multiscale features and improves feature fusion. This structure is suitable for DEM refinement because terrain surfaces contain both broad geomorphic patterns and fine local discontinuities. In this study, U-Net is used in the initial DSM reconstruction stage to reduce local gaps and discontinuities, whereas U-Net++ is applied in the final refinement stage after nonground removal and bias correction. The final U-Net++ stage functions as a dense elevation-regression model that restores local terrain continuity, boundary sharpness, and fine topographic detail. Related deep architectures have also shown effectiveness in residual learning, height estimation, and satellite-based 3D reconstruction tasks.(7–10)

2.2 Low-rank matrix factorization

Even after deep-learning-based reconstruction and morphological filtering, DEMs may retain systematic errors such as global offsets, broad tilts, stripe-like patterns, and localized outliers. To reduce these residual errors, a low-rank-plus-sparse decomposition is applied as a post-processing step.(11) In this framework, the error matrix is separated into a low-rank component representing scene-wide systematic structure and a sparse component representing localized anomalies. This approach is suitable for satellite-derived DEM refinement because residual errors often coexist at multiple spatial scales. Wide-area distortion may remain after stereo reconstruction, while isolated building remnants, canopy artifacts, or matching failures may survive local filtering. The decomposition is therefore used to suppress both systematic and localized errors, with terrain-dependent effects in mountainous, flat, and urban regions.

2.3 SMRF

SMRF is a morphology-based ground filtering method developed for airborne light detection and ranging (LiDAR) terrain classification.(14) It uses a linearly increasing window and slope-based thresholding to separate terrain from elevated objects. Although originally developed for point-based LiDAR data, the same principle can be applied to rasterized DSMs in which buildings, trees, and other nonground objects appear as locally elevated structures. SMRF is conceptually related to the earlier progressive morphological filter,(15) which removes nonground measurements through repeated morphological openings with increasing window sizes and elevation-difference thresholds. In this study, SMRF is applied to the reconstructed DSM to generate a preliminary ground surface and nonground mask before low-rank correction and final U-Net++ refinement. This step is especially important for suppressing building and vegetation artifacts, although steep natural slopes in mountainous terrain may still increase the risk of misclassification.

2.4 Performance evaluation metrics

DEM accuracy is evaluated using *RMSE*, mean absolute error (*MAE*), and Bias for each terrain category. *RMSE* reflects the overall magnitude of error and is sensitive to large deviations, *MAE* indicates average prediction accuracy with less influence from outliers, and Bias represents systematic overestimation or underestimation. Because the influence of outliers and study-dependent evaluation criteria have been emphasized in DEM assessment studies,(12,13) the above three metrics are analyzed separately by terrain type rather than only over the full study area.

3. Methodology

In this study, we employ a four-stage DEM refinement workflow that combines deep learning, morphological filtering, and low-rank matrix decomposition to improve the quality of satellite-derived DEMs. As shown in Fig. 1, the overall workflow consists of four sequential steps: U-Net-based DSM generation, SMRF-based nonground removal and preliminary DEM generation, low-rank matrix decomposition-based outlier correction, and U-Net++-based DEM refinement. Each stage is designed to mitigate local noise and scene-wide distortion, ultimately producing a refined DEM with a spatial resolution of 5 m.

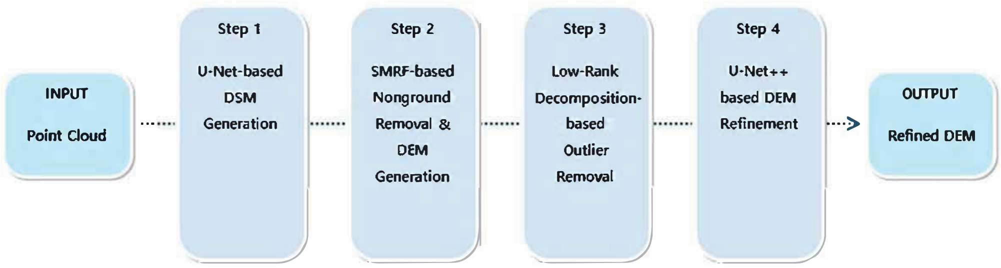


Fig. 1. (Color online) Deep-learning-based procedure.

3.1 Study areas and terrain classification

Study areas were selected in San Francisco, USA, and in Canada to examine terrain-specific performance across two geographically distinct test regions. Both regions encompass diverse landscapes within their satellite imagery, including densely developed urban areas, suburban areas, flat coastal zones, and mountainous inland areas. For topographical evaluation, each DEM pixel was classified into four categories on the basis of topographic characteristics and land cover: Mountainous Terrain, Flat Terrain, Urban Area, Nonurban Area. Mountainous Terrain is characterized by steep slopes and significant elevation differences, typically featuring dense vegetation on the slopes. Flat Terrain is characterized by gentle slopes and minimal elevation changes, such as coastal plains and flat agricultural lands. Urban Area comprises areas with high building density, featuring a mix of residential, commercial, and industrial facilities of various heights. Nonurban Area is characterized by gentle to moderate terrain, where buildings are sparse and natural vegetation, agricultural land, or open spaces constitute the majority of the land cover. Through this four-category classification, error mechanisms specific to each topographic environment can be analyzed, and the impact of topographic complexity and land cover density on DEM quality can be distinguished.

3.2 Input data configuration

Initial DSMs were generated from KOMPSAT-3 stereo satellite imagery, which provides ground sampling distances of 0.8 m (panchromatic) and 3.2 m (multispectral). For each study region, two experimental configurations were prepared to assess performance under different acquisition conditions. The ‘Multi’ configuration combines stereo pairs acquired on different dates within a short temporal baseline, enabling multiview fusion. The ‘Single’ configuration uses same-date stereo pairs to isolate single-acquisition performance. Tables 1 and 2 summarize the detailed specifications for the USA and Canada datasets, respectively.

3.3 U-Net-based DSM generation

The first stage is the generation of a homogeneous DSM using a U-Net-based deep learning model, taking an initial point cloud or a DSM output from commercial software as input. In this study, a 1-m-resolution DSM was used as input to correct imbalances arising from elevation changes, boundaries, and shadow areas within the image, thereby reconstructing the terrain surface into a homogeneous DSM. The generated DSM served as the base data for nonground point removal and DEM generation in the subsequent steps. In urban zones, the primary role of this stage is filling occlusion gaps caused by tall buildings. In mountainous and vegetated terrain, it smooths canopy-induced elevation spikes while preserving the underlying slope structure.

Table 1
Specifications of USA satellite imagery.

	USA_Multi	USA_Single
Satellite	KOMPSAT-3	KOMPSAT-3
Number of scenes	2 scenes	2 scenes
Total scene area	19.71 × 15.99 km^2 19.90 × 16.01 km^2	18.31 × 16.00 km^2 18.31 × 16.01 km^2
Acquisition date	2020/02/24–2020/03/11	2019/10/27–2019/10/27
GSD (m)	PAN 0.8 m, MS 3.2 m	PAN 0.8 m, MS 3.2 m

Table 2
Specifications of Canada satellite imagery.

	Canada_Multi	Canada_Single
Satellite	KOMPSAT-3	KOMPSAT-3
Number of Scenes	2 scenes	2 scenes
Total Scene Area	18.65 × 16.01 km^2 17.95 × 16.01 km^2	19.85 × 16.00 km^2 20.07 × 16.01 km^2
Acquisition Date	2021/09/21–2021/09/28	2019/10/20–2019/10/20
GSD (m)	PAN 0.8 m, MS 3.2 m	PAN 0.8 m, MS 3.2 m

3.4 SMRF-based nonground removal and DEM generation

The reconstructed DSM is processed through SMRF to generate a ground/nonground binary classification. Iterative morphological operations with progressive window expansion separate terrain surface from above-ground objects. The output consists of a preliminary DEM containing only classified ground points and a binary mask identifying nonground regions. In urban areas, the bulk of building-induced elevation artifacts are removed, and in vegetated areas, canopy height contributions are suppressed.

3.5 Low-rank decomposition-based outlier removal

In the third stage, low-rank matrix factorization is applied to the preliminary DEM to suppress systematic elevation patterns and residual outliers. The elevation difference matrix is decomposed into low-rank and sparse components. In mountainous terrain, in this stage, sensor-geometry-related tilts amplified by the large elevation range are primarily corrected; in urban areas, residual building artifacts that survived morphological filtering are isolated; and in flat terrain, small systematic offsets that would otherwise degrade absolute elevation accuracy are removed.

3.6 U-Net++-based DEM refinement

In the fourth stage, the U-Net++ model was used to restore local terrain detail and resolve boundary discontinuities in the corrected DEM. The secondary DEM and the initial DSM were jointly used as inputs to recover terrain boundaries, remove residual noise, and preserve natural slope transitions. This process produced a tertiary refined DEM with improved local consistency and elevation registration. The final DEM was then converted to a 5 m resolution and exported as the refined output product. Owing to its nested skip connections, U-Net++ integrates multiscale features effectively, improving boundary sharpness at terrain transitions, such as ridge lines in mountainous areas and building-terrain interfaces in urban areas, while maintaining smooth slope continuity.

3.7 Comparison subjects and evaluation methods

To evaluate the performance of the proposed method, we generated DEMs of the same region and compared them with those generated using three commercial software packages: PCI Geomatica, Hexagon ERDAS Image, and SimActive Correlator3D. These tools generate DSMs using traditional satellite image stereo matching and triangulation algorithms, which are widely used in global satellite DEM production. The evaluation index was calculated separately for each terrain type to obtain *RMSE*, *MAE*, and Bias values for mountain, urban, nonurban, and flat areas. This allowed us to analyze how the accuracy of each method varies depending on terrain complexity and land cover density.

4. Experiments and Results

For comparison, tables 3 and 4 present terrain-specific visuals of between PCI-generated DEMs and the DEMs produced by the proposed procedure for the USA and Canada study areas. In mountainous areas, commercial DEMs show systematically high surfaces due to unresolved vegetation canopies, and procedure outputs accurately represent the underlying ground relief from which the canopy's effects have been removed. In urban areas, commercial products maintain noticeable building footprints as highly elevated artifacts, and our proposed procedure effectively suppress them while preserving the surrounding terrain structure. Flat terrain shows subtle but significant improvements in noise reduction and deflection elimination, while in nonurban areas, scattered buildings and vegetation are more clearly separated from the ground surface.

Tables 5 and 6 summarize the *RMSE* values by terrain type for the United States and Canada sites. For each terrain category, *RMSE* values are reported for DEMs generated from both single-image and multi-image datasets. Overall, the proposed deep learning refinement method generally achieved lower or competitive *RMSE* values relative to those of PCI and ERDAS and was comparable to or better than Correlator3D depending on terrain type and image configuration. The improvement was particularly notable in urban areas, where the proposed method substantially reduced building-related elevation artifacts and achieved clear gains over the commercial software. In mountainous areas, *RMSE* values remained higher than in other terrain classes because of steep relief and complex viewing geometry; nevertheless, the proposed method still showed substantial reduction relative to PCI and ERDAS, although Correlator3D retained lower *RMSE* in some cases. In flat and nonurban areas, the proposed method also demonstrated stable improvement, although the final *RMSE* values varied by study region, terrain class, and single-/multi-image configuration. Multi-image processing generally improved accuracy across most methods, and the proposed framework remained effective under both single-image and multi-image conditions.

Table 3
(Color online) Detailed PCI and deep learning results: DEM visualization (USA, Single).

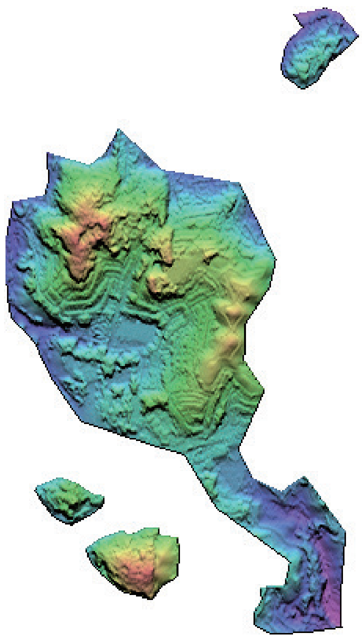
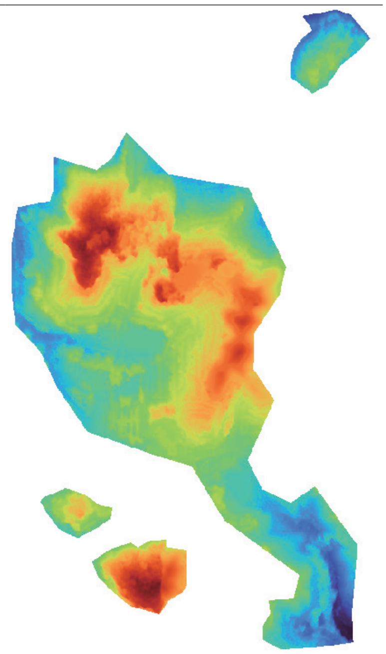
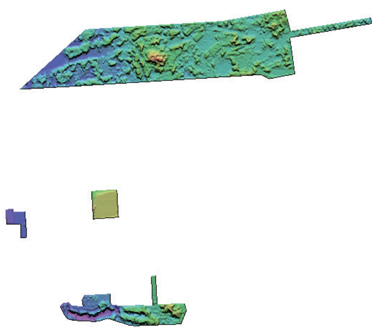
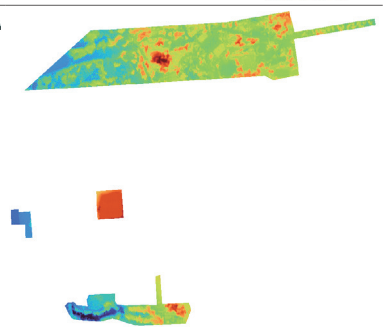
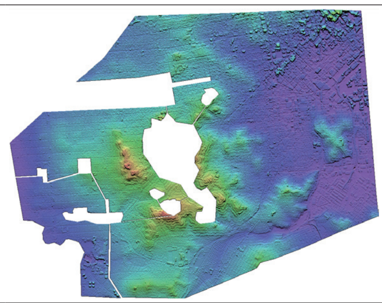
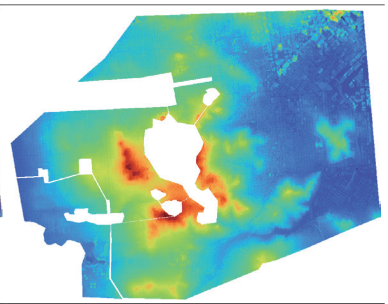
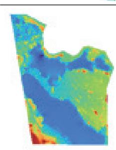
	PCI	DEM
Mountainous terrain		
Flat terrain		
Urban area		
Nonurban area		

Table 4
(Color online) Detailed PCI and deep learning results: DEM visualization (Canada, multi).

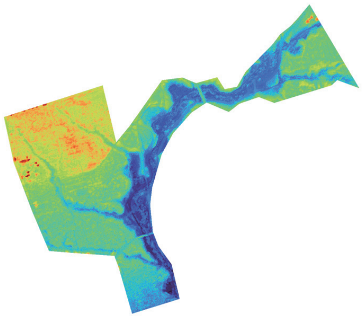
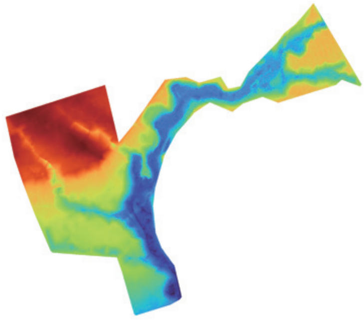
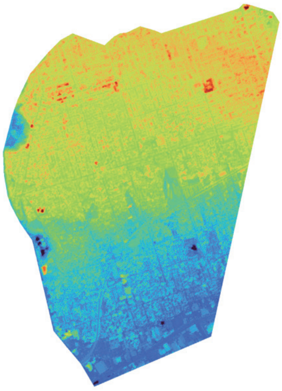
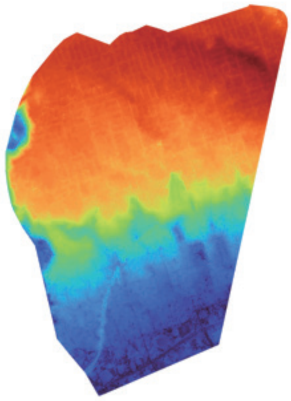
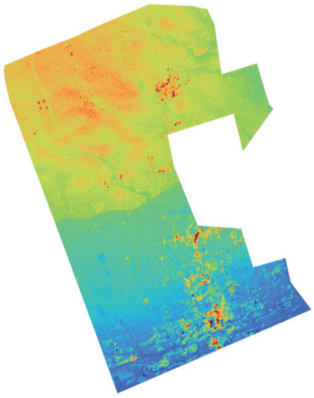
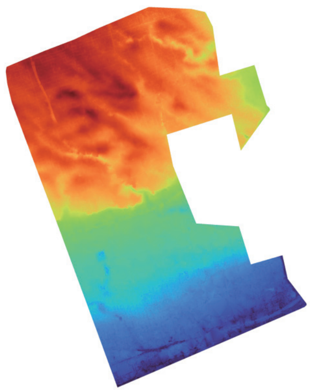
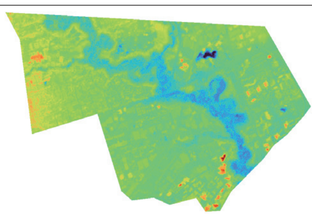
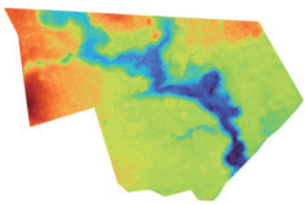
	PCI	DEM
Mountainous terrain		
Flat terrain		
Urban area		
Nonurban area		

Table 5
RMSE by terrain type for USA sites (unit: m).

Classification	Commercial software	SW generation results (single/multi)	Deep learning generation results (single/multi)
Mountainous terrain	PCI	14.403/14.400	4.546/4.417
	ERDAS	13.063/14.907	4.630/4.042
	Correlator3D	8.501/7.593	4.085/3.942
Flat terrain	PCI	13.478/13.114	3.955/3.622
	ERDAS	10.595/10.071	3.916/3.517
	Correlator3D	3.682/3.531	3.782/3.415
Urban area	PCI	10.234/7.797	3.175/2.916
	ERDAS	8.957/7.396	2.906/2.658
	Correlator3D	3.784/2.684	2.947/2.790
Nonurban area	PCI	6.577/8.368	2.852/3.297
	ERDAS	5.890/7.248	2.612/3.241
	Correlator3D	28.727/7.442	3.632/3.812

Table 6
RMSE by terrain type for Canada sites (unit: m).

Classification	Commercial software	SW generation results (single/multi)	Deep learning generation results (single/multi)
Mountainous terrain	PCI	10.922/11.713	3.230/2.955
	ERDAS	10.020/10.734	3.223/3.786
	Correlator3D	3.119/3.757	3.303/3.890
Flat terrain	PCI	6.818/7.686	1.632/1.570
	ERDAS	5.265/6.380	1.490/1.857
	Correlator3D	2.582/6.560	1.932/1.878
Urban area	PCI	17.923/16.991	1.713/1.708
	ERDAS	12.700/12.979	1.827/2.212
	Correlator3D	4.897/14.133	2.198/2.507
Nonurban area	PCI	10.333/10.475	2.384/2.138
	ERDAS	8.652/8.651	2.773/2.817
	Correlator3D	4.897/7.292	2.198/4.156

Tables 7 and 8 summarize the *MAE* values by terrain type. The *MAE* results exhibited a trend broadly similar to that of *RMSE*, with the proposed method showing substantial error reduction in many terrain classes, although it did not yield the lowest *MAE* in every case. In urban areas, the proposed method produced considerably lower *MAE* values than did PCI and ERDAS, indicating clear improvement in average prediction accuracy. In mountainous terrain, the proposed method also reduced *MAE* substantially relative to those of PCI and ERDAS, although that of Correlator3D remained lower in some single-image cases. In flat and nonurban areas, the proposed method maintained competitive *MAE* values and generally reduced average error relative to the major commercial baselines. Multi-image datasets also showed an overall tendency toward lower *MAE*. These results indicate that the proposed framework improves average elevation accuracy across diverse terrain conditions, while the degree of improvement depends on terrain class and input configuration.

Table 7
MAE by terrain type for USA sites (unit: m).

Classification	Commercial software	SW generation results (single/multi)	Deep learning generation results (single/multi)
Mountainous terrain	PCI	9.281/9.049	3.202/3.543
	ERDAS	6.895/5.987	3.061/3.358
	Correlator3D	2.32/2.198	2.95/2.424
Flat terrain	PCI	5.538/3.751	2.672/2.158
	ERDAS	4.464/3.448	2.579/2.079
	Correlator3D	4.968/5.653	3.151/3.364
Urban area	PCI	5.564/5.68	2.434/2.676
	ERDAS	5.46/5.434	2.178/2.393
	Correlator3D	1.714/2.479	2.324/3.359
Nonurban area	PCI	9.855/9.844	3.57/3.768
	ERDAS	9.627/8.543	3.206/3.633
	Correlator3D	5.144/15.151	3.22/2.791

Table 8
MAE by terrain type for Canada sites (unit: m).

Classification	Commercial software	SW generation results (single/multi)	Deep learning generation results (single/multi)
Mountainous terrain	PCI	8.277/9.581	2.124/2.099
	ERDAS	7.833/9.002	2.085/2.621
	Correlator3D	1.994/2.840	2.049/2.782
Flat terrain	PCI	4.998/6.105	1.311/1.414
	ERDAS	3.906/5.160	1.292/1.630
	Correlator3D	1.711/3.008	1.444/1.388
Urban area	PCI	9.983/10.475	1.427/1.456
	ERDAS	8.252/9.033	1.550/1.808
	Correlator3D	2.420/7.218	1.660/2.071
Nonurban area	PCI	6.313/7.029	1.475/1.479
	ERDAS	6.068/6.698	1.851/2.075
	Correlator3D	2.420/3.807	1.660/2.766

Finally, the results of comparison for Bias are presented in Tables 9 and 10. Bias indicates the extent to which each method systematically overestimates or underestimates elevation, and the ideal value is zero. The commercial software showed substantial terrain- and method-dependent variation, with Bias values ranging from strongly negative to strongly positive values. In contrast, the proposed method substantially reduced systematic bias across most test cases, although the magnitude of improvement varied by site and terrain class. For single-image processing, systematic bias was effectively removed in the USA test areas and was generally much smaller in the Canadian test areas. For multi-image datasets, commercial methods showed partial improvement, but the proposed method generally maintained lower Bias than the commercial software, although residual bias remained in several Canadian terrain classes. These results indicate that the low-rank matrix decomposition step effectively improved elevation registration at the scene scale, even when complete bias elimination was not achieved in every terrain class.

Table 9
Bias by terrain type for USA sites (unit: m).

Classification	Commercial software	SW generation results (single/multi)	Deep learning generation results (single/multi)
Mountainous terrain	PCI	9.151/8.667	0/0
	ERDAS	6.765/5.331	0/0
	Correlator3D	0.003/−0.651	0/0
Flat terrain	PCI	5.51/3.409	0/0
	ERDAS	4.409/3.337	0/0
	Correlator3D	−4.165/−2.283	0/0
Urban area	PCI	5.478/5.44	0/0
	ERDAS	5.417/5.373	0/0
	Correlator3D	0.382/1.41	0/0
Nonurban area	PCI	9.482/9.036	0/0
	ERDAS	9.343/8.27	0/0
	Correlator3D	−1.504/−14.856	0/0

Table 10
Bias by terrain type for Canada sites (unit: m).

Classification	Commercial software	SW generation results (single/multi)	Deep learning generation results (single/multi)
Mountainous terrain	PCI	7.993/9.539	1.102/1.195
	ERDAS	7.752/8.979	1.138/1.102
	Correlator3D	0.029/1.077	0.004/1.010
Flat terrain	PCI	4.721/5.985	1.057/1.322
	ERDAS	3.514/4.983	0.916/1.420
	Correlator3D	−1.555/−1.931	−1.286/−0.177
Urban area	PCI	9.854/10.326	1.250/1.310
	ERDAS	8.009/8.795	1.306/1.418
	Correlator3D	0.461/−4.595	0.448/0.390
Nonurban area	PCI	6.006/6.221	0.782/0.890
	ERDAS	6.006/6.590	1.196/1.442
	Correlator3D	0.461/−2.647	0.448/−1.565

Overall, the proposed deep-learning-based DEM refinement method achieved generally superior performance relative to the major commercial software tested. In terms of *RMSE* and *MAE*, the proposed framework reduced error by approximately 35–40% on average, with particularly strong improvements in urban areas characterized by dense building coverage. In terms of Bias, our method substantially reduced the systematic offsets observed in the commercial DEM products and improved overall absolute elevation consistency. This performance can be attributed to the complementary roles of the detailed components: deep learning enhanced local detail reconstruction, SMRF reduced nonground artifacts, and low-rank correction improved scene-wide elevation alignment. Although multi-image inputs generally improved performance across most methods, the proposed framework also produced competitive results under single-image conditions, indicating that it remains effective even when stereo input is limited.

5. Conclusions

In this study, we evaluated a hybrid DEM refinement workflow integrating U-Net, SMRF, low-rank matrix factorization, and U-Net++ with an emphasis on terrain-specific performance across heterogeneous surface conditions. Experimental validation using KOMPSAT-3 stereo imagery from San Francisco, USA, and multiple Canadian sites showed that the proposed framework generally outperformed PCI and ERDAS and achieved a performance that was competitive with or better than that of Correlator3D depending on terrain type and input configuration. Overall, our framework reduced *RMSE* and *MAE* by approximately 35–40% on average and substantially reduced systematic Bias, thereby improving both relative precision and absolute elevation consistency. Most importantly, the results of this study showed that the DEM refinement performance is inherently terrain-dependent. Urban areas exhibited the largest improvements, reflecting the effectiveness of the proposed workflow in suppressing building-induced elevation artifacts. Mountainous terrain also showed meaningful improvement despite higher residual error caused by steep slopes and complex terrain geometry. Flat and nonurban areas demonstrated stable performance improvement, although the magnitude of improvement varied by study region and image configuration. In addition, the results showed that the proposed framework remained effective even under single-stereo conditions, achieving, in several cases, performance comparable to or better than multi-image commercial outputs. These findings suggest that terrain-stratified evaluation provides a more informative basis for interpreting DEM refinement performance than aggregate scene-wide accuracy metrics alone.

References

1 C. V. Tao and Y. Hu: Photogramm. Eng. Remote Sens. **68** (2002) 705.
2 J. Grodecki and G. Dial: Photogramm. Eng. Remote Sens. **69** (2003) 59. https://doi.org/10.14358/PERS.69.1.59
3 H. Hirschmüller: IEEE Trans. Pattern Analysis and Machine Intelligence **30** (2008) 328. https://doi.org/10.1109/TPAMI.2007.1166
4 J. H. Oh, D. C. Seo, and C. N. Lee: J. Korean Soc. Surv., Geodesy, Photogramm. Cartogr. **32** (2014) 19. https://doi.org/10.7848/ksgpc.2014.32.1.19
5 H. Lee and M. Hahn: Remote Sens. **11** (2019) 2340. https://doi.org/10.3390/rs11202340
6 O. Ronneberger, P. Fischer, and T. Brox: Int. Conf. Medical Image Computing and Computer-Assisted Intervention (2015) 234. https://doi.org/10.48550/arXiv.1505.04597
7 K. He, X. Zhang, S. Ren, and J. Sun: Proc. 2016 IEEE Conf. Computer Vision and Pattern Recognition (2016) 778. https://doi.org/10.1109/CVPR.2016.90
8 L. Mou and X. X. Zhu: arXiv:1802.10249 (2018) https://doi.org/10.48550/arXiv.1802.1024
9 Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang: DLMIA **11045** (2018) 3. https://doi.org/10.48550/arXiv.1807.10165
10 J. Gao, J. Liu, and S. Ji: ISPRS J. Photogramm. Remote Sens. **195** (2023) 446. https://doi.org/10.1016/j.isprsjprs.2022.12.012
11 E. J. Candes, X. Li, Y. Ma, and J. Wright: J. ACM **58** (2011) 11. https://doi.org/10.48550/arXiv.0912.3599
12 J. Hohle and M. Hohle: ISPRS J. Photogramm. Remote Sens. **64** (2009) 398. https://doi.org/10.1016/j.isprsjprs.2009.02.003
13 J. L. Mesa-Mingorance and F. J. Ariza-Lopez: Remote Sens. **12** (2020) 2630. https://doi.org/10.3390/rs12162630
14 T. J. Pingel, K. C. Clarke, and W. A. McBride: ISPRS J. Photogramm. Remote Sens. **77** (2013) 21. https://doi.org/10.1016/j.isprsjprs.2012.12.002
15 K. Zhang, S. C. Chen, D. Whitman, M. L. Shyu, J. Yan, and C. Zhang: IEEE Trans. Geosci. Remote Sens. **41** (2003) 872. https://doi.org/10.1109/TGRS.2003.810682

About the Authors

Jae Myeong Kim received his Ph.D. degree from the University of Seoul in 2012. He is currently an associate professor in the Major of Urban Planning Engineering at Seokyeong University. His research interests encompass automation and data related to smart construction, construction automation, and construction data platforms. (jm927k@skuniv.ac.kr)

Ji Yoon Kim received his Master's degree from Korea University in 2014. He currently serves as the CEO of Carbon Ace and holds a concurrent position as the vice director of the Spatial Information Research Center at Seokyeong University. His research interests are primarily focused on deep learning, AI, and computer vision technologies. (aliasgis@gmail.com)

Kwan Young Oh is currently working at Satellite Application Center, Korea Aerospace Research Institute (KARI). His research focuses on the exploration of new frontiers, technological leadership, and the development and dissemination of aerospace science and technology. (ohky@kari.re.kr)

Kwang Jae Lee is currently working at Satellite Application Center, Korea Aerospace Research Institute (KARI). His research focuses on the exploration of new frontiers, technological leadership, and the development and dissemination of aerospace science and technology.
(kjlee@kari.re.kr)