Sensors and Materials, Vol. 37, No. 2 (2025) 773–782
MYU Tokyo

S & M 3955

Algorithm for Generating Orthophotos from Unmanned Aerial Vehicle Imagery Based on Neural Radiance Fields

Junxing Yang,[1] Tianjiao Wang,[1] Renzhong Wang,[1]
He Huang,[1*] Gwangjae We,[2] and Dongha Lee[3**]

[1]School of Geomatics and Urban Spatial Informatics, Beijing University of Civil Engineering and Architecture, Beijing 102616, China
[2]Technology Research Headquarters, Geostory Co., Ltd., Seoul 07566, Republic of Korea
[3]Department of Integrated Energy and Infra System, Kangwon National University, Chuncheon 24341, Republic of Korea

(Received November 29, 2024; accepted February 18, 2025)

Keywords: orthophoto, neural radiance field (NeRF), photogrammetry algorithms, instant neural graphics primitives (instant-ngp)

Digital orthophotos are renowned for their high geometric accuracy and distortion-free characteristics, captured from a parallel perspective. They are widely used in map making, urban planning, and related fields. In this study, we employed the neural radiance field (NeRF) technology to generate highly realistic orthophotos through an end-to-end image generation process, eliminating the need for prior 3D geometric information or auxiliary data. We compared the NeRF-based approach with current mainstream photogrammetric methods for orthophoto generation. The experimental results demonstrate that the NeRF-based algorithm meets the measurement accuracy requirements and surpasses traditional methods in terms of detail and texture quality. We analyzed the performance characteristics of forward-facing and 360° object-centric camera shooting methods. Our findings indicate that combining these techniques yields high-quality orthophotos, demonstrating the advantages of implicit methods in orthophoto generation. Moreover, the findings provide valuable guidance for the efficient production of digital orthophotos.

1. Introduction

Digital orthophotos are distinguished by their high geometric accuracy and distortion-free characteristics, captured from a bird's-eye perspective. They are widely used in map making, urban planning, and other fields. Early methods for generating digital orthophotos primarily depended on the transformation of digital images.(1) These approaches involved calculating the parameters of the transformation model using ground control points and applying these transformations to the digital images. However, these transformation models are often overly simple, leading to inaccuracies in the resulting digital orthophotos, frequently requiring manual adjustments.

*Corresponding author: e-mail: huanghe@bucea.edu.cn
**Corresponding author: e-mail: geodesy@kangwon.ac.kr
https://doi.org/10.18494/SAM5492

The advancement in 3D reconstruction technology has made it possible to generate digital orthophotos using 3D models and project them onto a horizontal plane. Among various image-based 3D reconstruction methods, photogrammetry is a widely recognized technique capable of producing dense and geometrically accurate 3D point clouds of real-world scenes from images captured from multiple angles. The generation of digital orthophotos relies primarily on the 3D reconstruction process. By applying the principles of photogrammetry, 3D objects are projected onto different images using internal and external camera parameters. During this process, image blocks are extracted with the reprojection point at the center. However, the likelihood of objects at the current elevation is inferred by quantitatively assessing the similarity between scenes, thereby reconstructing the crucial spatial structure information. Once the 3D reconstruction is complete, a digital orthophoto can be generated by downward parallel projection, orthogonally projected onto the horizontal plane to yield the final results.(2)

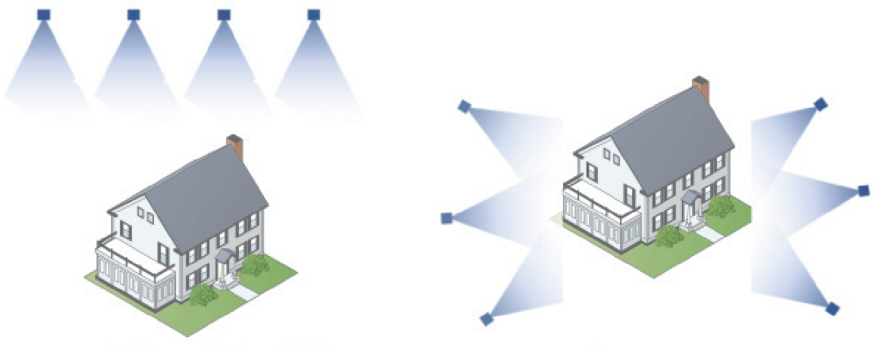
Neural radiance fields (NeRFs), a rapidly evolving neural rendering technique, have attracted significant attention in recent years and demonstrated considerable potential.(3) NeRFs facilitate new perspectives and are suitable for generating digital orthophotos for any scene, provided that scene reconstruction is completed. However, traditional NeRFs utilize a straightforward multi-layer perceptron (MLP) to represent a continuous radiance field by sampling spatial points along the ray direction from the camera to the image plane. This approach retrieves the density and irradiance values from the MLP, which are used to render a new perspective through volume-rendering equations. This method requires large computational memory and resources, making acceleration essential.

Several studies have optimized this process by enhancing encoding techniques. For instance, the TensorF method proposed by Chen *et al.* models and reconstructs a scene's radiance field as a 4D tensor, representing a 3D voxel grid with multichannel features.(4) This approach achieves exceptional rendering quality and reduces memory usage compared with earlier methods. Müller *et al.* introduced the concept of instant neural graphics primitives (instant-ngp) with multiresolution hash coding, enabling the rapid and efficient generation of 3D models.(5)

A commonly adopted strategy for representing unbounded scenes involves using a spacewarping method that maps an unbounded space onto a bounded space.(3,6,7) Two different warping functions exist: (1) For forward-facing scenes [Fig. 1(a)], normalized device coordinate (NDC) warping is employed to map an infinitely distant view frustum to a bounded box by compressing the space along the z-axis.(6) (2) For 360° object-centric unbounded scenes [Fig. 1(b)], inverse-sphere warping involves mapping an infinitely large space to a bounded sphere through the sphere inversion transformation.(7) The expansion of neural-radiance scenes into urban orthophotos has become a reality.

In this study, we utilized instant-ngp with multiresolution hash coding to generate orthophotos by enhancing the rendering equation. First, we compared the generation quality of the forward-facing camera dataset with that of the surrounding dataset. The results demonstrate that this combination significantly improves the overall data quality.

Moreover, the results reveal that our approach is more advantageous owing to generation speed compared with photogrammetry-based methods, achieving suitable measurement accuracy and rendering quality.



(a) (b)

Fig. 1. (Color online) (a) Forward-facing and (b) 360° object-centric camera trajectories.

2. Methods

NeRFs utilize a mapping function F that transforms a 3D spatial location X and a viewing direction d into volume density σ and color value c. The ability of MLPs to effectively map low-frequency signals is limited, prompting NeRFs to employ positional encoding. Conventional NeRFs often require time-consuming training and evaluation processes owing to the involvement of fully connected neural networks.

Therefore, to address this challenge, instant-ngp employs MLPs with hash encodings for coordinates and a spherical harmonic function for viewing directions. The scene space is divided into L different grid resolutions for coordinate encoding, where each resolution Nl is calculated using a geometric sequence with scale factor b. For direction encoding, the view-dependent color (c), after encoding the coordinate $h(X)$, is determined by the spherical harmonic (SH) function $Y(d)$. Here, k represents the SH coefficients, from which the colors can be further derived.

We generated a novel view synthesis method based on NeRFs. However, a significant limitation arises with central projection techniques, which introduce a perspective effect causing objects to appear larger when closer and smaller when farther away. Although this effect is common, it is not ideal for accurately representing terrain features. Therefore, we replaced the central projection with a parallel projection. We defined the position and direction of each rendered ray within the world's coordinate system. Subsequently, the rays were transformed into a 3D coordinate space and projected to produce true orthophotos, as illustrated in Fig. 2.

3. Experiments and Analysis

The data used in this study were obtained from an image dataset captured using a DJI FC330 drone. The dataset comprises forward-facing images captured vertically, as referenced in the LLFF. Consequently, our data comprised zenith-angle images captured with a vertically downward-oriented camera lens. This was at an angle of approximately 90° with respect to the

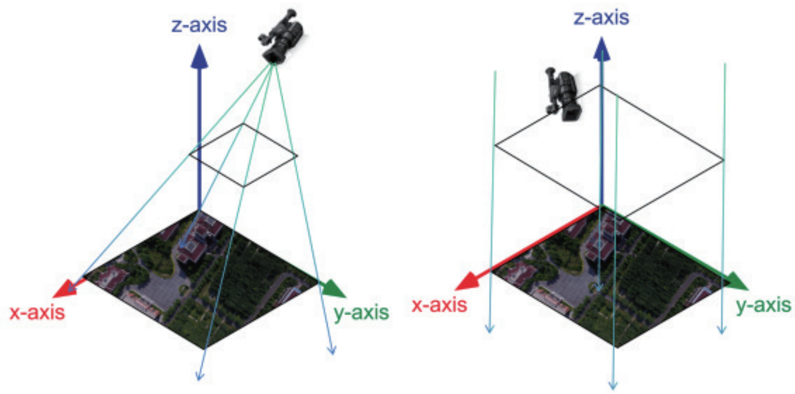


Fig. 2. (Color online) Multiple parallel rays are evenly projected upward onto a virtual plane to render a true orthophoto from (b) an orthographic projection, compared with (a) the central projection.

ground. However, incorporating data from 360° object-centric images is crucial for accurately generating orthophoto images and unbounded scenes. These perspectives are essential for reconstructing buildings or complex terrain because 360° object-centric images can compensate for the limitations of zenith-angle images, addressing issues related to missing perspectives caused by building occlusion. For a detailed description of the dataset, please refer to Table 1.

In our experiments, the model was trained on the complete dataset, with both the training and validation phases integrated into a self-training framework. All experiments were conducted on an RTX A5000 GPU, utilizing a batch size of 4096 rays to optimize GPU memory usage. The training process consisted of 30 epochs per session, with each session lasting approximately 5 to 7 min. The initial learning rate was set to 0.001 and was dynamically adjusted on the basis of the model's convergence behavior. The Adam optimizer was used, and the loss function followed the implementation in the Instant-NGP framework. To evaluate the effectiveness of our algorithm, we compared the orthophotos generated by conventional methods with those produced by our approach, using ground truth images as references.

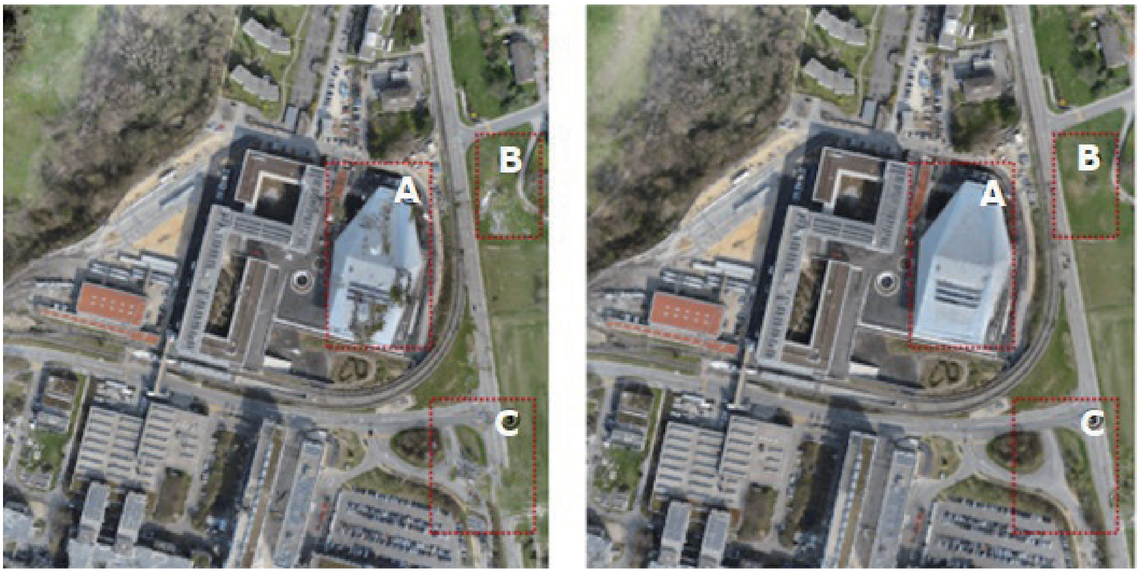
For benchmarking, we selected Metashape (formerly PhotoScan), a widely recognized commercial software program for generating digital orthophotos using photogrammetry.(8) Metashape's established effectiveness in this domain made it a suitable reference method for our study. We used the results generated by Metashape as a benchmark to evaluate the accuracy of our approach.

3.1 Test on various scenes

All images were oriented and undistorted using Metashape. Initially, we tested the forward-facing images and combined them with the 360° object-centric images, as shown in Fig. 3. We then evaluated the generation of digital orthophotos across various typical scenes, including reflective buildings, building blocks, and villages, to assess both the image quality and the algorithm's robustness in different environments. The results of this evaluation are presented in

Table 1
Data details.

Scene	a	b	c
f (mm)	4	4	4
Resolution	4000*3000	4000*22500	4000*3000
Number	221	96	82



(a) (b)

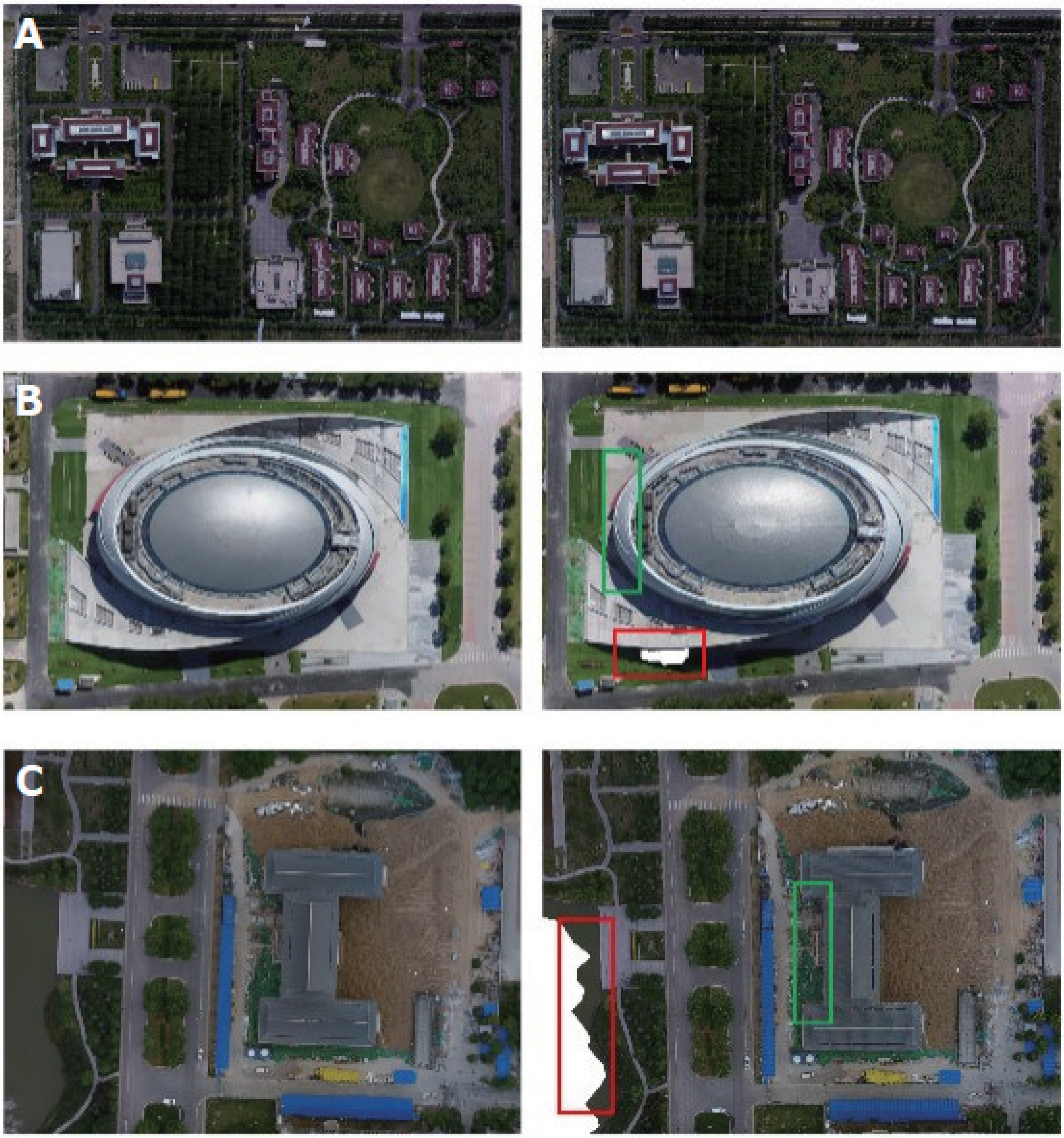
Fig. 3. (Color online) (a) Generation of digital orthophotos from forward-facing images and (b) their integration with 360° object-centric images. The areas of the red square frame are represented by the labels A, B, and C.

Fig. 4. Next, we compared the accuracy of the proposed method with those of two other approaches. Finally, we measured the time required to generate scenes of varying sizes, allowing us to evaluate the efficiency of both techniques.

We evaluated the generation of digital orthophotos for urban scenes using forward-facing images and their integration with 360° object-centric images, as illustrated in Fig. 3. Moreover, the overall rendering quality of the buildings was satisfactory when only the forward dataset was utilized. Particular details exhibited blurriness and artifacts, particularly in the three areas highlighted by the red square frame (Fig. 3).

In the area of the red square frame A, a green artifact is observed in the blue building because the scene incorrectly renders the vegetation atop the structure. In the area of the red square frame B, the white spots are attributable to the improper mapping of light reflections. In the area of the red square frame C, the building shows shadow artifacts resulting from insufficient edge data, which leads to blurriness and inaccurate texture mapping during viewpoint changes. However, the combination of forward and surrounding data effectively mitigates these issues, rendering high-quality orthophotos.

We then tested the generation of digital orthophotos for various scenes using both instant-ngp and photogrammetry-based methods, as illustrated in Figs. 4(a) and 4(b), respectively. The



(a) (b)

Fig. 4. (Color online) Digital orthophotos of scenes generated using two different methods, (a) instant-ngp and (b) photogrammetry methods. The areas are represented by the labels A, B, and C.

photogrammetry method exhibited significant inaccuracies in areas with sudden height variations, such as the edges of roofs, where sharp changes in elevation resulted in misalignments and artifacts in the generated orthophotos. In contrast, our method effectively captured these sudden changes, as highlighted by the green box in Area B, demonstrating superior performance in accurately modeling complex structures with rapid elevation transitions.

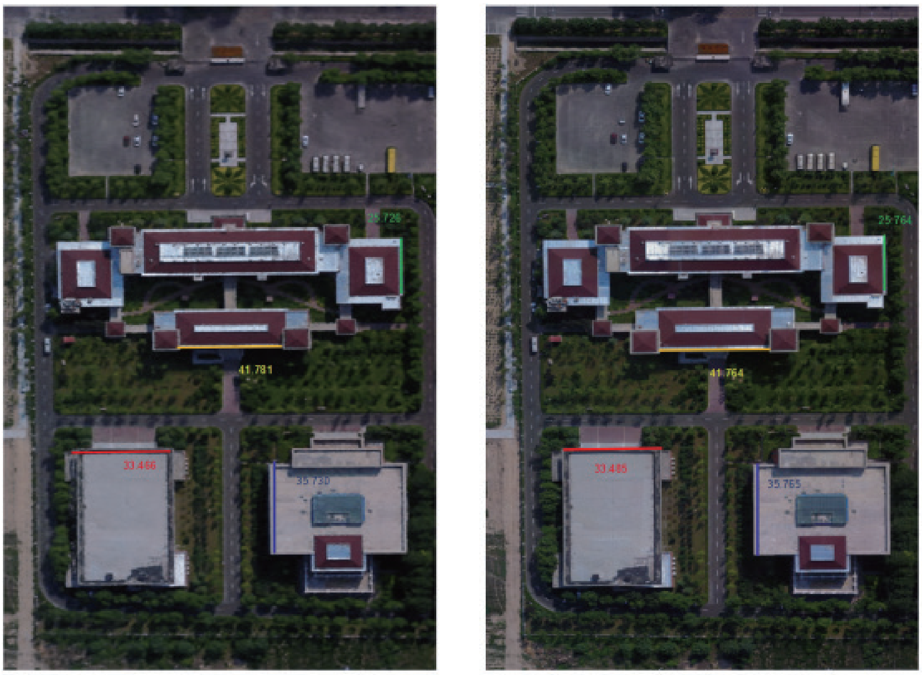
Furthermore, the photogrammetry method faced challenges with ghosting in areas containing moving objects, a common issue in photogrammetry where dynamic elements disrupt the reconstruction of static 3D scenes. Although this phenomenon can result in artifacts, such as blurred or duplicated objects in the final image, its impact was minimal in our approach. As demonstrated in the relevant areas, our method managed these dynamic elements more effectively, preserving the integrity of the scene without introducing notable distortions.

Additionally, in areas with reflective or highly textured surfaces, the photogrammetry method often produces artifacts, such as holes or inaccurate surface representations, owing to its challenges in accurately reconstructing scenes with low texture or highly reflective materials. In contrast, our method demonstrated greater robustness in these environments, as illustrated in the red box in area C, where it provided a more accurate and visually consistent representation of the scene.

3.2 Accuracy assessment

Geometric accuracy is a crucial feature for digital orthophotos, making distance measurement precision essential. We selected specific areas as test scenes to verify the measurement accuracy of various digital orthophoto generation methods. Digital orthophotos were generated by instant-ngp-based and photogrammetry-based methods, followed by length measurements.

Figure 5 shows a comparison of the distance measurements between the two methods, with the error between the instant-ngp-based and photogrammetry-based methods being approximately 0.03 m. Thus, this result suggests that digital orthophotos generated by the instant-ngp-based method meet the requirements for surveying and mapping.



(a) (b)

Fig. 5. (Color online) Digital orthophotos produced by (a) instant-ngp-based and (b) photogrammetry-based methods exhibit segments with consistent colors and corresponding values, indicating identical distances measured in both results.

Table 2
Processing time comparison between our proposed method and photogrammetry algorithms (unit: min).

Scene	Photogrammetry	Proposed method
a	42	15
b	25	12
c	20	10

3.3 Evaluation efficiency

Processing time is a critical metric for evaluating the effectiveness of orthophoto-generation algorithms. Our hardware setup included an Intel® Core™ i7-13700F CPU and an NVIDIA GTX A5000 GPU, which leveraged GPU acceleration. We compared the times required by the generation algorithms using the same dataset. Additionally, our approach was implemented in Python, which yielded slower performance than implementations in C++. However, the processing of our proposed method was faster than that of photogrammetry algorithms. Table 2 shows the processing time between our method and the photogrammetry algorithms (unit: minutes).

4. Conclusions

In this study, we proposed a method for generating true orthophotos in complex environments, addressing several challenges commonly encountered in photogrammetry techniques, such as distortions in building façades and areas with weak textures, including water bodies. We analyzed the performance of forward-facing and 360° object-centric camera shooting methods, demonstrating that their combination produces high-quality orthophotos.

The proposed method demonstrated excellent performance compared with traditional algorithms and related approaches. However, we identified certain limitations regarding execution speed when implemented in Python. To address this, future work will focus on transitioning our method to C++ to enhance efficiency and broaden its applicability, including the integration of prior point-cloud data and the management of datasets with reduced overlap.

By incorporating these improvements, we expect our method to further enhance UAV-based photogrammetry applications in fields such as mapping, surveying, and environmental monitoring, contributing to more accurate and efficient orthophoto generation.

Acknowledgments

This work is supported by the Korea Agency for Infrastructure Technology Advancement (KAIA) grant R&D program of Digital Land Information Technology Development funded by the Ministry of Land, Infrastructure and Transport (MOLIT) (Grant RS-2022-00142501).

References

1 B. A. DeWitt and P. R. Wolf: Elements of Photogrammetry (with Applications in GIS) (McGraw-Hill Higher Education, 2000).
2 Q. Wang, L. Yan, Y. Sun, X. Cui, H. Mortimer, and Y. Li: Photogramm. Rec. **33** (2018) 113.
3 B. Mildenhall, P. P, Srinivasan, M. Tancik, J. T. Barron, R. Ramamoorthi, and R. Ng: Commun. ACM **65** (2021) 99.
4 A. Chen, Z. Xu, A. Geiger, J. Yu, and H. Su: Proc. European Conf. Computer Vision (Cham: Springer Nature Switzerland, 2022) 333–350.
5 T. Müller, A. Evans, C. Schied, and A. Keller: ACM Trans. Graphics, **41** (2022) 1.
6 J. T. Barron, B. Mildenhall, and D. Verbin D: Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition (2022) 5470–5479.
7 K. Zhang, G. Riegler, and N. Snavely: arXiv preprint arXiv:2010.07492 (2020).
8 Agisoft: Agisoft PhotoScan (2016). https://www.agisoft.com/

About the Authors

Junxing Yang received his Ph.D. degree from Wuhan University, focusing on photogrammetry and remote sensing. His research pursuits extend to diverse domains such as 3D reconstruction, image stitching, and scene understanding. (yangjunxing@bucea.edu.cn)

Tianjiao Wang is currently pursuing his master's degree at the School of Geomatics and Urban Spatial Informatics, Beijing University of Civil Engineering and Architecture. His research interests include 3D reconstruction, photogrammetry, and orthophoto generation.
(2108160122006@stu.bucea.edu.cn)

Renzhong Wang is currently pursuing his master's degree at the School of Geomatics and Urban Spatial Informatics, Beijing University of Civil Engineering and Architecture. His research interests include LIDAR simultaneous localization and mapping. (2108570022064@stu.bucea.edu.cn)

He Huang received his B.S. degree from Wuhan University, China, in 2000, and pursued further education abroad, earning both his M.S. and Ph.D. degrees from Sungkyunkwan University, South Korea, in 2004 and 2010, respectively. Since 2010, he has held positions as a lecturer and associate professor at Beijing University of Civil Engineering and Architecture, China. His research focuses on areas such as autonomous driving, high-precision navigation maps, and visual navigation and positioning.
(huanghe@bucea.edu.cn)

Gwangjae We received his B.E., M.S., and Ph.D. degrees from Sungkyunkwan University, Korea, in 1997, 1999, and 2012, respectively. He founded Geostory Co., Ltd. in 2012 and currently serves as the Chief Technology Officer at the Technology Research Headquarters. His research interests include airborne bathymetric LiDAR, HCMI map, and geospatial information.
(gjwie@geostory.co.kr)

Dongha Lee received his B.E., M.S., and Ph.D. degrees from Sungkyunkwan University, Korea, in 2000, 2003, and 2008, respectively. He is currently an associate professor at Kangwon National University, Korea. His research interests are in geodesy, surveying, geospatial information, and natural hazard analysis with remote sensing. (geodesy@kangwon.ac.kr)