A practical approach to Candi Siwa 3D reconstruction with COLMAP and Nerfstudio

Helena Widiarti, Rokhmat Febrianto, Agung Alfiansyah
Department of Computer Systems Engineering, Faculty of Science, Technology, Engineering, and Mathematics, Prasetiya Mulya University, Jakarta, Indonesia

Article Info

Article history:

Received May 14, 2025
Revised Nov 11, 2025
Accepted Dec 8, 2025

Keywords:

3D documentation
3D reconstruction
Computer vision
Crowdsourcing
Cultural heritage
Digitization

ABSTRACT

We demonstrate a practical approach for large-scale object three-dimensional (3D) reconstruction with freely available frameworks, COLMAP and Nerfstudio. We performed the reconstruction of a temple named Candi Siwa, located at Prambanan Site, which is situated between Central Java and Yogyakarta Province, Indonesia. We utilized COLMAP and Nerfstudio as platforms for 3D reconstruction from images captured by an everyday smartphone. In the 3D model construction process, COLMAP generates a dense point cloud, whereas Nerfstudio generates a scene from source images. We selected 96 images of Candi Siwa to perform reconstruction using COLMAP. As a result, a 3D model for the temple with a clear structure and color was observed. A scene rendered in MP4 format was also generated using Nerfstudio. Additionally, we performed the 3D reconstruction from 150 images taken by the public and found them insufficient for constructing the object. This occurred despite the number of images being larger than those used in the previous reconstruction. The results indicate that the success of a crowdsourcing project for reconstructing a large-scale object should consider not only the number of images but also the variation in point of view and the completeness of the whole structure.

Corresponding Author:

Helena Widiarti
Department of Computer Systems Engineering
Faculty of Science, Technology, Engineering, and Mathematics, Prasetiya Mulya University
Jakarta, Indonesia
Email: helena.widiarti@prasetiyamulya.ac.id

1. INTRODUCTION

Research in three-dimensional (3D) reconstruction has always been a fascinating field in computer vision [1]. The advancement of digital technology has increased the processing capability of devices while simultaneously lowering cost. One of the applications for this technology is the preservation of cultural heritage. Many studies have focused on the reconstruction of 3D models for cultural heritage. Recent work in Bertocci *et al*. [2] demonstrated 3D digitization of La Verna that combines laser scanning and photogrammetry. Liu and Mamat [3], 3D laser scanning was used to acquire point cloud models of the great achievement palace of the Confucian temple. Costantino *et al.* [4] utilized scan-to-heritage building information modeling (HBIM) for the conservation and preservation of San Nicola in Montedoro church, Italy.

A popular application of 3D objects is augmented reality (AR). In AR, the 3D model of objects can be observed digitally, allowing tourists to visit a historical site virtually. Spallone *et al*. [5], AR and virtual reality (VR) enable visitors to do interactive communication in the virtual museum of Museo D'arte Orientale Di Torino. Banfi *et al*. [6], data visualization with AR is presented at the Museo Nazionale Romano, Italy. Recently, the digital twin has become an attractive research topic due to its potential applications in Industry

4.0. It introduces the concept of virtualization of an object, which is not limited to only showing the structure or color in 3D form. In digital twin, the behaviour of the object can be simulated as well [7]. The applications of digital twin range from manufacturing to urban landscaping. We can conclude that the 3D model of an object has many potential applications in the future.

The common methods for constructing a 3D model of an object are photogrammetry, triangulation and time-of-flight laser scanning, structured light, stereo-calibrated optical sensors, and mobile lidar [8]. In photogrammetry, the 3D models for real objects are created from multiple images. The objects in photogrammetry are not limited to small objects, but large objects, as well as a reconstruction of a site. Examples of using photogrammetry in small objects, such as a statue or figurine, are presented in Bonora *et al*. [9]. Meanwhile, for the reconstruction of the site, Sambugaro *et al.* [10] use photogrammetry to represent industrial objects. Prasidya *et al*. [11] the multi-mode sensor placement, combining light detection and ranging (LIDAR) and photogrammetry, was used to capture the Prambanan Temple's detailed structure. Alsadik [12], crowdsourced and web-based drone videos were used as a resource for 3D reconstruction of cultural heritage sites.

One of the popular platforms for photogrammetry is COLMAP. It was introduced first in 2014 by Schönberger and Frahm [13]. It utilized structure-from-motion (SfM), which estimates 3D structure from 2D images from multiple points of view. COLMAP is available freely under an open-source license. Because of this, COLMAP has been used in many research projects [14], [15]. Recently, neural radiance fields (NerF), a method that utilizes artificial intelligence (AI) to perform 3D reconstruction, has gained popularity as well. While COLMAP employs a classic approach in making 3D objects, NerF utilizes a deep learning method to generate a scene based on two-dimensional (2D) images. Together, COLMAP and Nerf can be used to create a point cloud file, a mesh file, and a scene in MP4 format. NerF uses the camera poses generated by COLMAP to create a scene [16]. COLMAP can estimate camera parameters and the sparse point clouds. NerF uses them for creating 3D models. Thus, with both tools, the digitization of cultural heritage can be generated in 3D visualization and video.

In this work, we present a workflow to use COLMAP and Nerfstudio for photogrammetry. We used a Linux-based server with an NVIDIA graphics processing unit (GPU) with compute unified device architecture (CUDA) support. The object of interest is one of the temples located inside the Candi Prambanan area, named Candi Siwa (Shiva Temple). We took around 150 images and performed 3D reconstruction using COLMAP. Later on, we created a scene based on the images with NerF. We also performed crowdsourcing, where we inquired number of people visiting the temple to take multiple pictures. Based on those images, we created a 3D model. The result shows that the 3D model generated from crowdsourcing fails to represent the structure of the object.

We propose a practical approach for utilizing COLMAP and NerF to reconstruct a 3D large-scale object with sufficient images. Our workflow can be replicated to get the 3D model of an object since both the framework (COLMAP, NerF) and hardware (computer, smartphone) are easily accessible. The data collection was done with smartphones, a device that is ubiquitous today, therefore allows everyone to contribute to the project. Furthermore, the framework used in this work is freely available.

COLMAP is an implementation of the SfM technique, where the geometrical feature of a structure is estimated from multiple images from different points of view [13]. COLMAP is an open-source platform for creating a 3D structure and camera motions from multiple images. COLMAP can run on a computer with a central processing unit (CPU), but it will produce a better result on a computer with a GPU.

The reconstruction process follows the steps as shown in Figure 1 [13]. First, feature extraction and matching are performed. The results are feature descriptors that are stored as an array to describe a local feature of an image. During matching process, COLMAP discovers images and performs a naive search for a pair of images. It identifies a potential match by comparing them to each other by calculating the minimum distance. Next, image registration matches the multiple images into a 3D model. The method detects and matches key features such as corners and edges between images. If a valid transformation maps a sufficient number of features between images, they are considered geometrically verified. The transformation parameters are computed based on these features. The result of the reconstruction is a mesh file that is formed by point clouds calculated in the previous step. The reconstruction is done incrementally through image registration, triangulation, and bundle adjustment.

NerF used deep learning to create a scene based on multiple images. NerF represents a scene with five-dimensional (5D) function (x, y, z), which are spatial coordinates and polar angles (ϑ and ϖ). The outputs are volume density and red, green, blue (RGB) color values, symbolized by c and σ [17]. The learning in NerF is based on the function f that estimates the parameters as shown in (1).

$$F_\theta: (x, d) \rightarrow (c, \sigma) \quad (1)$$

Two significant scientific works in the development of NeRF are [17] which improves training efficiency using depth information obtained from COLMAP, and [18] which introduced the foundational NeRF method for realistic scene reconstruction and novel view synthesis.

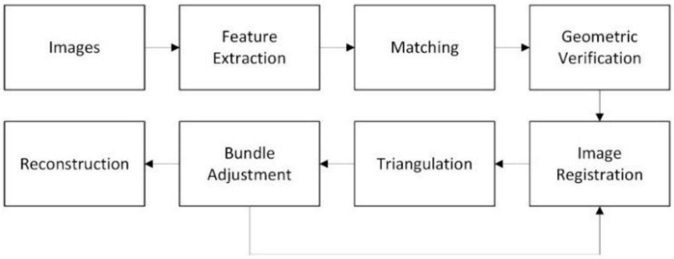


Figure 1. COLMAP pipeline

2. METHOD

We performed data collection and reconstruction for a large-scale object. The object of research is Candi Siwa, a temple located in the center of the Prambanan Site. There are two reasons for this selection. First, Candi Siwa is the largest temple inside the Prambanan Site. Second, the surrounding perimeter of Candi Siwa is an area with no obstacles or other objects. People can capture images that include the whole structure of the temple with ease. Prambanan is a popular tourist attraction in the region. In the Prambanan area, the temples are distributed across the area in various sizes. It is located in Central Java and Yogyakarta Province [19]. We conducted our data collection for a day in December 2024.

Data collection was done in two parts. The first part was done by a research team, and the public did the second step. For the first part, the people in charge of taking the pictures circled around the temple and took the images with the purpose of having images with many overlapping areas. The images shall cover the whole structure of the temple. For the second part, we asked 10 volunteers to take around 50 images. In this crowdsourcing project, the volunteers were provided with only general guidance regarding the image collection process. Specifically, they were instructed to capture photographs while moving around the temple in order to obtain coverage of the entire structure. However, detailed methodological instructions such as prescribing the interval of movement between successive shots (e.g., taking a short step before each capture) were not conveyed. Consequently, the guidance emphasized overall spatial coverage rather than adherence to a standardized or systematic acquisition protocol.

For 3D reconstruction, we selected COLMAP and NerF as our tools. During reconstruction with COLMAP, the parameters are selected automatically. The methods used in automatic reconstructor for feature extraction, feature matching, and geometric verification steps are scale-invariant feature transform (SIFT), exhaustive matcher, and random sample consensus (RANSAC) with epipolar constraints. The result generated by COLMAP that has been observed is a file with a ply extension. The ply file is a mesh representation of the object of interest.

Nerfstudio [20] was selected for the scene creation. The framework requires NVIDIA with CUDA installed on the server. Nerfstudio also requires PyTorch with CUDA as a dependency. We used Nerfstudio to generate a scene in MP4 format, with the source images similar to those used in COLMAP. Nerfstudio needs COLMAP to obtain the external and internal parameters of the camera and generate point clouds. Both COLMAP and NerF run on one server. Figures 2(a) and (b) show the overall workflow of our research, and the dependency and result files of COLMAP and Nerfstudio.

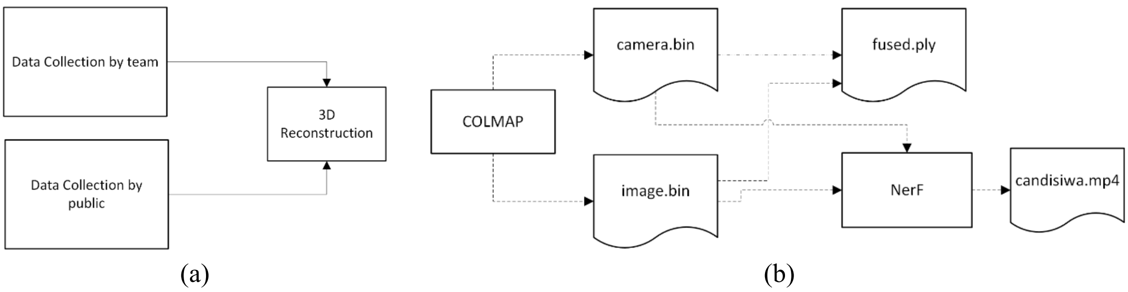


Figure 2. Flowchart of work in (a) general and (b) 3D reconstruction

We installed the COLMAP and Nerfstudio on a server equipped with an NVIDIA GPU card. During the experiment, we ran the server with the client host connected within the local area network (LAN). All the commands in COLMAP and Nerfstudio were done with a command line interface, except for the MP4 generating process, where we used graphical user interface (GUI) for Nerfstudio. The reconstruction process in real time can be monitored through Nerfstudio viewer that is forwarded to the default port 7007.

3. RESULTS AND DISCUSSION

From the researcher team's dataset, 96 images were selected among 150 images to perform the 3D reconstruction that represents the object with good quality and shows the structure well. Only a subset of the images was selected, since many of the images exhibit overlap with minimal variation. The size of each picture ranges from 597 bytes to 1 MB. The samples of images presented in Figures 3(a) and (b) captured from different points of view. The camera used to take the images is Samsung SM-A528B smartphone.

We asked tourists who visited the site on the same day to take images and submit them directly. Overall, we obtained around 240 images from the public. Among them, we selected 150 images to perform 3D reconstruction, excluding images with minimal overlap. The samples of images are shown in Figures 4(a) and (b) both show different points of view with incomplete structure.

(a) (b)

Figure 3. Samples of images taken by researcher team: (a) left view and (b) right view

(a) (b)

Figure 4. Samples of images taken by the public with incomplete structure: (a) front view and (b) left view

The 3D reconstruction of Candi Siwa was done using COLMAP and Nerfstudio. We run both frameworks on a server equipped with an NVIDIA GPU in the local network. The specification of the server is shown in Table 1. The proposed hardware configuration, comprising an NVIDIA GeForce RTX 3060 GPU, an AMD Ryzen 7 5700X 8-core processor, and 32 GB of RAM, provides an effective balance between computational performance and cost. This setup is sufficient to perform the required 3D reconstruction tasks efficiently while remaining financially accessible to most research institutions and individual practitioners. Compared to high-end servers or dedicated GPU clusters, the configuration represents a practical and cost-effective solution for small- to medium-scale reconstruction projects.

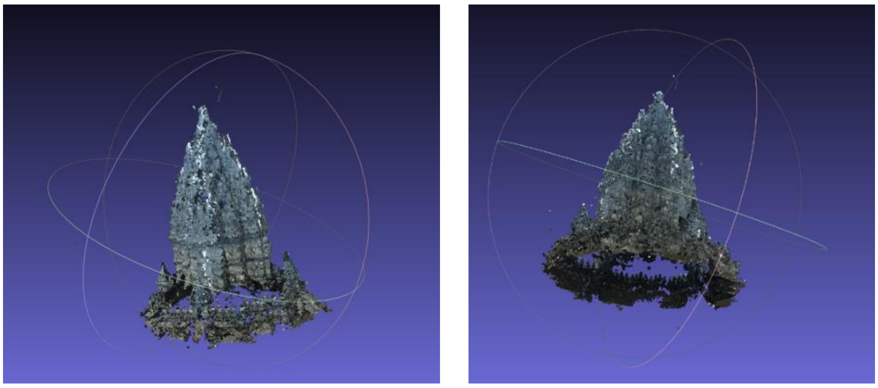
Table 1. Reference server specifications demonstrating an affordable and scalable hardware setup

Hardware component	Specification
Graphic card	NVIDIA GeForce RTX 3060
CPU model	AMD Ryzen 7 5700X 8-core processor
RAM	32 GB

We set the parameter to automatic to run COLMAP. The default value of parameters in automatic_reconstructor are optimized to refine all focal length, principal point, and distortion parameters. The manually set parameters are used for the case when the camera calibration parameters are known beforehand. Since we used an everyday smartphone to take the pictures, the automatic reconstruction parameters are already the most suitable choice, since smartphones have unknown and variable intrinsic.

The command for automatic reconstruction is: colmap automatic_reconstructor --workspace_path $DATASET_PATH --image_path $DATASET_PATH/images, where we changed parameters $DATASET_PATH and $DATASET_PATH/images to the directory of the source images file. We observed the camera poses from “images.txt” and the intrinsics from “cameras.txt”. These files are generated from “camera.bin” and “image.bin” in COLMAP. The image.txt file noted that the number of images is equal to 96, while the mean observations per image is 4407.437. From “cameras.txt”, we obtain that the model used in the camera is SIMPLE_RADIAL. The height of the images ranges from 1500 to 4169 pixels. The width of the images ranges from 3000 to 3120 pixels.

Figures 5(a) and (b) illustrates the output from “fused.ply” file from different viewpoints. From the result, we conclude that the 3D mesh for Candi Siwa has succeeded in showing color and the overall structure of the temple. The server needed around 2-3 hours to generate the dense mesh. The time needed for this reconstruction project is relatively slow and may not be suitable for real world scalability. Further improvements on the hardware is needed to make this solution applicable in real world for reconstruction projects, especially with the participation of public.



(a) (b)

Figure 5. SfM 3D model of the object captured from two perspectives: (a) left view and (b) right view

We used around 150 images to generate the mesh file. The result is shown in Figure 6. As the result shows, the 3D reconstruction failed to capture the whole structure of the object. This result is attributed to the lack of variation in point of view, which reduced geometric information and parallax essential for depth estimation [21]. The photographer in crowdsourcing groups take several images by moving their hand position while standing at the same point. Furthermore, some of the images failed to show the whole structure of object, because the photographer is standing too close to the object. This finding was reported also in [22], [23] in which most of images from crowdsourcing were taken close to the object, minimal invariance to illumination and only around subset of them can be used for reconstruction.

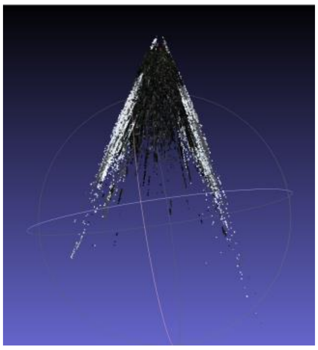


Figure 6. 3D reconstruction of Candi Siwa using the public-captured images

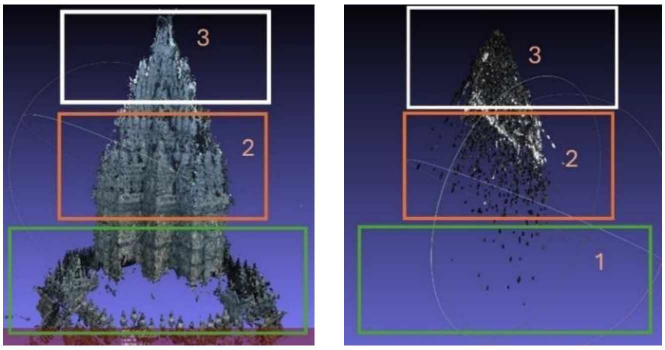
To increase the quality of reconstruction, filtering was applied to the source images for reconstructing the object [24] into two categories, inlier and outlier. The classification was done with parameter d, define as the distance of a point to its nearest neighbor. If d is larger than global mean and standard deviation of all d, then it is considered as an outlier. Upon this however, results are not much improved and still fail to show the whole structure.

To compare the results of the research team and crowdsourcing datasets quantitatively, the Open3D library in Python was used to calculate point density and reprojection error. The results, summarized in Table 2, show that the researcher team's dataset achieves slightly higher point density and lower mean reprojection error. This indicates more accurate and complete reconstructions when images are systematically collected. As for the average overlap in researcher team is 482.79 which is higher than crowdsourcing 378.79, thus producing better reconstruction.

Table 2. Comparison of 3D reconstruction quality metrics between the researcher team and crowdsourcing datasets

Parameters	Researcher team	Crowdsourcing
Point density	3382788	3037765
Mean reprojection error	1.329412	1.50674
Average overlap (shared 3D points per image pair)	482.97	378.79

Figures 7(a) and (b) shows side by side comparison of both research team and crowdsourcing results. The crowdsourcing result failed to show the structure and texture of the temple for all three points. Thus, our work demonstrates the difference in the quality of reconstruction objects between expert (researcher team) and public images through crowdsourcing. Similar findings were observed in [25], in which the data that had been contributed without clear guidance often suffers from reduced quality and completeness, which aligns with challenges in image-based reconstruction when contributors do not vary their capture viewpoints sufficiently. A more common approach to achieving higher public engagement and obtaining reconstruction with good results simultaneously is by combining images from both teams, as described in [26].



(a) (b)

Figure 7. Comparison of reconstruction result for (a) research team and (b) crowdsourcing with annotation

Next, we use NerF to create a scene based on images taken by the researcher team. First, we put camera.bin and image.bin files generated by COLMAP in the workspace folder. Then we run the command: ns-train nerfacto colmap --data candisiwa --center-method focus. This step generates “config.yml” file, which contains parameters for training. We found that the utilization of GPU is around 98% for this process. Despite this, the training process was completed successfully. Figure 7 shows the example of real time viewing of 3D reconstruction process in NerfStudio.

After the training process is done, we use the render menu from Nerfstudio GUI to generate a short video. In order to do this, we selected one of the training cameras that correspondent to the starting viewpoint of the scene. The result is presented in Figure 8. The scene reconstruction effectively captures both the structural details and the color information of the surrounding environment.

Figure 8. 3D reconstruction real time monitoring with Nerfstudio GUI

4. CONCLUSION

In this work, we present a practical approach to utilize COLMAP and Nerfstudio as tools to perform 3D reconstruction. The object of interest is Candi Siwa, a temple located between Central Java Province and Yogyakarta Province, Indonesia. We used Linux-based server equipped with NVIDIA GPU card. Two reconstruction results are produced. The first result, using images taken by the team (96 images), successfully performed 3D reconstruction with clear structure and color with COLMAP. The scene reconstruction in the form of an MP4 file has been created successfully as well with Nerfstudio. However, the second reconstruction result, using images taken by the public, was shown to be insufficient in constructing the object, with an unclear structure and color, despite the larger number of images (150 images). This result is attributed to a lack of variation in point of view, where photographers take several images by moving their hand position while standing at the same point. Additionally, the images do not capture the whole structure of the object. Therefore, in the crowdsourcing approach, it is essential to provide clear guidance to the participants in capturing the object to ensure the coverage of the whole structure from various points of view. Such guidance can be shown as a part of a crowdsourcing project through videos or posters. Our reconstruction task took around 2-3 hours which is relatively slow and may limit the scalability of this approach in real-world applications. Hardware upgrades or computational optimizations are therefore necessary to make this reconstruction solution practical, particularly when incorporating large-scale public participation.

ACKNOWLEDGMENTS

Authors would like to thank Naumi Sasmita, a student in Computer System Engineering, Universitas Prasetiya Mulya for her contribution during data collection. This work was supported by the Prasetiya Mulya University through 2024 Internal Research Grant scheme.

FUNDING INFORMATION

This research was funded by Internal Research Grant scheme in Prasetiya Mulya University.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Helena Widiarti	✓	✓	✓	✓	✓	✓		✓	✓	✓	✓	✓	✓	✓
Rokhmat Febrianto	✓	✓	✓	✓		✓	✓	✓		✓	✓	✓		✓
Agung Alfiansyah	✓	✓		✓						✓	✓	✓		✓

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author, [HW], upon reasonable request.

REFERENCES

[1] L. Zhou, G. Wu, Y. Zuo, X. Chen, and H. Hu, "A comprehensive review of vision-based 3D reconstruction methods," *Sensors*, vol. 24, no. 7, Apr. 2024, doi: 10.3390/s24072314.

[2] S. Bertocci, A. Lumini, and A. Cottini, "Virtual tour and 3D digitisation as tools for cultural heritage fruition. The robbiane at the sanctuary of La Verna," in *Proceedings of the 21th International Conference on Culture and Computer Science: from Humanism to Digital Humanities*, Oct. 2024, pp. 1–12, doi: 10.1145/3719236.3719275.

[3] S. Liu and M. J. B. Mamat, "Application of 3D laser scanning technology for mapping and accuracy assessment of the point cloud model for the Great Achievement Palace heritage building," *Heritage Science*, vol. 12, no. 1, May 2024, doi: 10.1186/s40494-024-01252-y.

[4] D. Costantino, M. Pepe, and A. G. Restuccia, "Scan-to-HBIM for conservation and preservation of Cultural Heritage building: the case study of San Nicola in Montedoro church (Italy)," *Applied Geomatics*, vol. 15, no. 3, pp. 607–621, Sep. 2023, doi: 10.1007/s12518-021-00359-2.

[5] R. Spallone, F. Lamberti, M. Guglielminotti Trivel, F. Ronco, and S. Tamantini, "3D reconstruction and presentation of cultural heritage: AR and VR experiences at the museo D'arte orientale Di Torino," *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. XLVI-M-1–2, pp. 697–704, Aug. 2021, doi: 10.5194/isprs-archives-XLVI-M-1-2021-697-2021.

[6] F. Banfi, M. Pontisso, F. R. Paolillo, S. Roascio, C. Spallino, and C. Stanga, "Interactive and immersive digital representation for virtual Museum: VR and AR for semantic enrichment of museo Nazionale Romano, Antiquarium di Lucrezia Romana and Antiquarium di Villa Dei Quintili," *ISPRS International Journal of Geo-Information*, vol. 12, no. 2, Jan. 2023, doi: 10.3390/ijgi12020028.

[7] S. Boschert and R. Rosen, "Digital twin—the simulation aspect," in *Mechatronic Futures*, Cham: Springer International Publishing, 2016, pp. 59–74, doi: 10.1007/978-3-319-32156-1_5.

[8] R. G. Boboc, E. Băutu, F. Gîrbacia, N. Popovici, and D.-M. Popovici, "Augmented reality in cultural heritage: An overview of the last decade of applications," *Applied Sciences*, vol. 12, no. 19, Sep. 2022, doi: 10.3390/app12199859.

[9] V. Bonora, G. Tucci, A. Meucci, and B. Pagnini, "Photogrammetry and 3D Printing for Marble Statues Replicas: Critical Issues and Assessment," *Sustainability*, vol. 13, no. 2, Jan. 2021, doi: 10.3390/su13020680.

[10] Z. Sambugaro, L. Orlandi, and N. Conci, "3D reconstruction methods in industrial settings: a comparative study for COLMAP, NeRF and 3D Gaussian Splatting," in *Ceur Workshop Proceedings*, 2024, pp. 212–217.

[11] A. S. Prasidya, I. Gumilar, I. Meilano, I. Ikaputra, R. Muryamto, and E. N. Arrofiqoh, "Three-dimensional digital documentation for the conservation of the Prambanan temple cluster using guided multi-sensor techniques," *Heritage*, vol. 8, no. 1, Jan. 2025, doi: 10.3390/heritage8010032.

[12] B. Alsadik, "Crowdsource drone imagery – A powerful source for the 3D documentation of cultural heritage at risk," *International Journal of Architectural Heritage*, vol. 16, no. 7, pp. 977–987, Jul. 2022, doi: 10.1080/15583058.2020.1853851.

[13] J. L. Schönberger and J. -M. Frahm, "Structure-from-Motion Revisited, " *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 4104-4113, doi: 10.1109/CVPR.2016.445.

[14] L. Morelli, F. Ioli, R. Beber, F. Menna, F. Remondino, and A. Vitti, "COLMAP-SLAM: A framework for visual odometry," *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. XLVIII-1/W, pp. 317–324, May 2023, doi: 10.5194/isprs-archives-XLVIII-1-W1-2023-317-2023.

[15] C. Bai, R. Fu, and X. Gao, "Colmap-PCD: An open-source tool for fine image-to-point cloud registration," in *2024 IEEE International Conference on Robotics and Automation (ICRA)*, May 2024, pp. 1723–1729, doi: 10.1109/ICRA57147.2024.10611582.

[16] C. Llull, N. Baloian, B. Bustos, K. Kupczik, I. Sipiran, and A. Baloian, "Evaluation of 3D reconstruction for cultural heritage

applications," in *2023 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)*, Oct. 2023, pp. 1634–1643, doi: 10.1109/ICCVW60793.2023.00179.
[17] K. Deng, A. Liu, J.-Y. Zhu, and D. Ramanan, "Depth-supervised NeRF: Fewer views and faster training for free," in *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2022, pp. 12872–12881, doi: 10.1109/CVPR52688.2022.01254.
[18] B. Mildenhall, P. P. Srinivasan, M. Tancik, J. T. Barron, R. Ramamoorthi, and R. Ng, "NeRF: Representing scenes as neural radiance fields for view synthesis," in *Communications of the ACM*, 2020, pp. 405–421, doi: 10.1007/978-3-030-58452-8_24.
[19] H. H. Parlindungan, T. Sukwika, and H. Manurung, "Prambanan temple tourist destination development in Indonesia as world cultural heritage," *European Journal of Science, Innovation and Technology*, vol. 1, no. 3, pp. 39–56, 2021.
[20] Nerfstudio Project, "nerfstudio," GitHub repository, 2022–2025. (Accessed Aug. 01, 2025). [Online]. Available: https://github.com/nerfstudio-project/nerfstudio.
[21] D. Cheng and E. Ch'ng, "Harnessing collective differences in crowdsourcing behaviour for Mass photogrammetry of 3D cultural heritage," *Journal on Computing and Cultural Heritage*, vol. 16, no. 1, pp. 1–23, Mar. 2023, doi: 10.1145/3569090.
[22] E. K. Stathopoulou, A. Georgopoulos, G. Panagiotopoulos, and D. Kaliampakos, "Crowdsourcing lost cultural heritage," *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. II-5/W3, pp. 295–300, Aug. 2015, doi: 10.5194/isprsannals-II-5-W3-295-2015.
[23] L. Morelli *et al.*, "The legacy of sycamore gap: the potential of photogrammetric AI for reverse engineering lost heritage with crowdsourced data," *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. XLVIII-2–2, pp. 281–288, Jun. 2024, doi: 10.5194/isprs-archives-XLVIII-2-2024-281-2024.
[24] A. Doulamis, A. Voulodimos, E. Protopapadakis, N. Doulamis, and K. Makantasis, "Automatic 3D modeling and reconstruction of cultural heritage sites from Twitter images," *Sustainability*, vol. 12, no. 10, May 2020, doi: 10.3390/su12104223.
[25] F. Biljecki, Y. S. Chow, and K. Lee, "Quality of crowdsourced geospatial building information: A global assessment of OpenStreetMap attributes," *Building and Environment*, vol. 237, Jun. 2023, doi: 10.1016/j.buildenv.2023.110295.
[26] K. Themistocleous, V. Evripidou, and K. Toumbas, "Monitoring cultural heritage sites at risk using citizen engagement through crowdsourcing," *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. XLVIII-G-2, pp. 1435–1440, Aug. 2025, doi: 10.5194/isprs-archives-XLVIII-G-2025-1435-2025.

BIOGRAPHIES OF AUTHORS

Helena Widiarti received the Bachelor of Engineering degree in Electrical Engineering from the University of Indonesia in 2006 and the Master of Science degree in Electrical Engineering from the Korea Advanced Institute of Science and Technology (KAIST), South Korea, in 2009. She is currently a Faculty Member at Universitas Prasetiya Mulya, where she teaches and conducts research in the Department of Computer Systems Engineering. Her research interests include artificial intelligence for next-generation network and IoT. She can be contacted at email: helena.widiarti@prasetiyamulya.ac.id.

Rokhmat Febrianto received the Bachelor of Engineering degree in Mechatronics Engineering from Politeknik Elektronika Negeri Surabaya (PENS), Surabaya, Indonesia, and the Master of Engineering degree in Electrical Engineering with a specialization in Multimedia Intelligent Network from Institut Teknologi Sepuluh Nopember (ITS), Surabaya, Indonesia. He is currently a Faculty Member in the Department of Computer Systems Engineering at Universitas Prasetiya Mulya, BSD, Indonesia. His research interests include robotics, artificial intelligence, machine learning, and computer vision, with a particular focus on developing intelligent robots that incorporate AI to enhance their functionality. He is actively involved in multidisciplinary research and collaboration, aiming to foster innovative projects in these fields. He can be contacted at email: rokhmat.febrianto@prasetiyamulya.ac.id.

Agung Alfiansyah received the Ph.D. degree in Mathematics and Informatics from the University de la Mediterrane, Marseille, France. He is currently a Faculty Member in the Department of Computer Systems Engineering, School of Applied STEM, Universitas Prasetiya Mulya, Tangerang, Indonesia. His current research interests include artificial intelligence, computer vision, medical image processing, 2D/3D image registration, and computational methods in engineering. He can be contacted at email: agung.alfiansyah@prasetiyamulya.ac.id.