Journal of Building Engineering 87 (2024) 108954

Contents lists available at ScienceDirect

Journal of Building Engineering

journal homepage: www.elsevier.com/locate/jobe

Single drone-based 3D reconstruction approach to improve public engagement in conservation of heritage buildings: A case of Hakka Tulou

Qingxiang Li, Guidong Yang [*], Chuanxiang Gao, Yijun Huang, Jihan Zhang, Dongyue Huang, Benyun Zhao, Xi Chen [**], Ben M. Chen

Department of Mechanical and Automation Engineering, The Chinese University of Hong Kong, Shatin, NT, Hong Kong

ARTICLE INFO

Keywords:
Heritage conservation
3D reconstruction
Drone photography
Tulou
Satisfaction survey

ABSTRACT

Public engagement in protecting architectural heritage is a critical component of sustainable development. This study has developed an innovative single drone-based 3D reconstruction (SD-3DR) approach for public-involved architectural heritage conservation. A new software tool named after CU-Recon is also developed to enable easy access and efficient 3D reconstruction using drone photography. Unlike traditional photogrammetry-based reconstruction, CU-Recon adopts our unique deep learning-based multi-view stereo network named LCM-MVSNet, enabling the public to employ only one drone for image capture and 3D reconstruction. LCM-MVSNet applies a learnable cost metric (LCM) to adaptively aggregate multi-view matching similarity into the 3D cost volume by leveraging sparse point hints. Its outstanding reconstruction performance for building scale applications is proved by the extensive experiments on the DTU training dataset and BlendedMVS dataset. A remarkable architectural heritage Hakka Tulou is selected to verify the effectiveness of the SD-3DR on large-scale heritage buildings. The results show our approach outperformed the other four 3D reconstruction tools. Moreover, the reconstruction quality in *CU-Recon* is evaluated from a perspective of heritage conservation using the concept of satisfaction. Practitioners expressed an overall satisfaction score of 4.175 (out of 5) for the reconstruction quality of the method. Survey results reveal a higher level of satisfaction with single drone photography for architectural heritage conservation compared to LiDAR-based scanning in terms of portability, operability, and cost. The research outcome changes the current situation of government-led top-down architectural heritage conservation by providing valuable insights for individual practitioners in creating bottom-up heritage conservation routes.

1. Introduction

Architectural heritage serves as a symbol of a region's identity, connecting the past with the present by encapsulating local culture in a comprehensive manner [1]. It holds a significant role in preserving the connection between people and the historical context of a region [2]. Architectural heritage is an invaluable, non-renewable resource in the progression of civilization and stands as a vital repository of each nation's historical legacy. As important components of sustainable development, architectural heritage possesses

[*] Corresponding author.
[**] Corresponding author.
E-mail addresses: gdyang@mae.cuhk.edu.hk (G. Yang), xichen002@cuhk.edu.hk (X. Chen).

https://doi.org/10.1016/j.jobe.2024.108954
Received 27 December 2023; Received in revised form 3 February 2024; Accepted 28 February 2024
Available online 8 March 2024

profound cultural, economic, and scientific significance [3]. Documenting and protecting architectural heritage are instrumental in upholding national culture's survival and continuity within the framework of sustainable development [4]. However, numerous threats, such as structural deterioration, environmental changes, pollution, destructive construction practices, and conflicts, imperil the sustainability of architectural heritage, leading to its degradation, and even destruction [5]. It is therefore essential to take measures that promote the sustainable preservation and utilization of architectural heritage.

Traditionally, heritage conservation has been chiefly led by government, with an emphasis on protecting well-known architectural treasures [6–9]. However, a substantial portion of architectural heritage consists of lesser-known and remote architectures, often neglected, inadequately preserved and documented [10]. As depicted in the memorable quote from the movie *CoCo*, "Death is not the end of life, forgetting is the end of life" [11]. This sentiment is also highly suitable to be used in lesser-known architectural heritage [7, 12]. They still exist worldwide, harboring history, culture, art, and narratives amidst the dust of time. But they have been largely forgotten by the world and left in obscurity [13]. Consequently, there is an urgent need to establish a bottom-up approach that actively involves the public in the preservation of lesser-known heritage sites.

The process of capturing the 3D geometries and surface textures of architectural heritage represents an easily comprehensible undertaking for the general public [14]. This type of data recording enables the presentation of heritage in a digital format, opening up possibilities for greater exposure, wider awareness, and enhanced conservation efforts. The data recording techniques described above primarily pertain to the fundamental digital three-dimensional (3D) representation of heritage buildings [15,16]. This digital 3D information model serves as a tool for recognizing potential threats, providing decision support for sound conservation strategies [17]. Consequently, it becomes essential to introduce an outstanding 3D reconstruction method for heritage buildings.

The data collection for reconstructing 3D architectural heritage models is typically accomplished through two primary tools: optical cameras and laser sensors, both of which are employed to acquire essential spatial data and texture information [16]. Among the existing 3D reconstruction technologies, two prominent approaches stand out: Air Oblique Photography Technology (OP) and Laser Scanning Technology (LT). LT is a scanning approach known for its high accuracy [18] and primarily applied with ground laser scanners [19]. Consequently, it has found extensive application in the detailed 3D reconstruction of individual structures [20]. However, because of its restricted scanning range, LT is not suitable for large-scale historical structures [21–24]. Furthermore, it presents challenges to the public in daily use due to its high cost (the Lidar approx. $30,000) [24,25], heaviness [26,27], and the need for specialized skills [25,28].

The OP technology is primarily integrated into unmanned aerial vehicle (UAV) platforms [29]. Because UAV platforms are cost-effective and portable, OP is widely applied in various building domains [21,30]. The fundamental process of OP technology involves employing sensors to capture a sequence of photographs with a predetermined degree of overlap. Subsequently, 3D data is indirectly generated through the application of corresponding 3D reconstruction algorithms. Multi-view stereo (MVS) plays a vital role as a fundamental element in the process of 3D reconstruction [31]. It is a technique that derives the 3D geometry of an object or scene from a collection of images taken from various known positions and angles. However, when it comes to the reconstruction of architectural heritage, the demand for clarity, precision, and completeness is notably high. Regrettably, the models produced by existing MVS algorithms using images acquired through OP do not meet the requirements for heritage documentation.

Recently, learning-based MVS approaches have demonstrated notable superiority over traditional methods in MVS benchmark assessments [32]. However, when it comes to implementing learning-based MVS approaches for the reconstruction of architectural heritage, there are still several challenges to be addressed. Firstly, most learning-based approaches make use of a feature pyramid network (FPN) to obtain features at multiple scales [33]. However, a common issue observed in these methods is the tendency to produce overly smoothed depth estimations near object boundaries. This smoothing effect is attributed to the deficiency of shallow feature information that includes low-level details. Moreover, effective cost volume aggregation is essential for maintaining consistency among multi-view photos. Learning-based cost volume aggregation (CVA) commonly involves the integration of an extra re-weighting network. This network is designed to learn and assign weights at various levels to facilitate the process of CVA. Nevertheless, the inclusion of this extra re-weighting network can introduce computational complexity and overlook the inherent correspondences that exist between multi-view images [34].

To solve the problems above, this paper develops a photogrammetry software tool *CU-Recon* based on an innovative MVS network incorporating Learnable Cost Metric (LCM) in order to meet the requirements of large-scale heritage building conservation [35]. Simultaneously, it optimizes memory usage and expedites inference speed. The network takes a set of multi-view images as input and generates a pyramid of per-view depth maps. Then, we perform filtering and fusion of multi-view depth maps to obtain a densely reconstructed point cloud. To assess the reconstruction quality, extensive experiments on benchmark datasets are conducted. However, codes may not be readily accessible to the general public and may pose challenges for individuals seeking to carry out 3D reconstructions of architectural heritage [36]. Hence, there is a pressing need to develop user-friendly software solutions to streamline the 3D reconstruction process and make it more accessible.

The reconstruction quality in CU-Recon is also required to be evaluated from the perspective of heritage conservation. We therefore invited practitioners in the field of heritage conservation to conduct a survey on the reconstruction quality. The satisfaction survey is used to quantitatively examine the reconstruction quality.

Herein, the rest of paper is as following structure: The methodology is illustrated in Section 2; Next, the case study and survey are shown in Section 3; Section 4 shows the results of the reconstruction of Tulou and the satisfaction survey.

2. Methodology

This study proposes a single-drone 3D reconstruction (SD-3DR) approach for public engagement into architectural heritage conservation. Fig. 1 shows the overview of the approach. Firstly, drone scanning of the target building is completed by the public. Then, the photos are input into the software tool CU-Recon to generate a 3D model. Finally, the reconstructed building model is upload to the WebGIS platform to realize the online visualization and interaction. In this paper, the software tool CU-Recon and its reconstruction algorithm (LCM based MVS Network) are highlighted.

2.1. *Software tool CU-Recon*

CU-Recon is a 3D reconstruction software tool based on an LCM-based MVS network, which has photogrammetry as its key technology. It can generate accurate 3D models of various subjects and surroundings by utilizing photographs captured from diverse angles with a single drone. Its primary application lies in the creation of 3D models for a wide range of subjects, with a particular aptitude for modeling building, terrains, and landscapes. Furthermore, this 3D reconstruction software offers a user-friendly interface, streamlining the entire photogrammetric pipeline. Users can directly import their images into the software, which then generates 3D models, simplifying the process. The key features of the CU-Recon include.

1. It supports a range of accurate and efficient 3D point cloud reconstruction.
2. A viewer that can be used directly to visualize 3D point cloud model.
3. It is suitable for the reconstruction of building-scale scenes with competitive performance.
4. It requires a limited memory footprint and reduces computational burden.
5. A set of innovative algorithms integrated well-known algorithms for processing point cloud reconstruction.

2.2. *LCM-based MVS network*

In this section, the new LCM-based MVS approach for 3D reconstruction is proposed to improve the reconstruction quality and adapted to a large-scale heritage building. The LCM aims to strike a balance between existing methodologies. In addressing pixel differences arising from diverse views, the LCM computes per-view features to accommodate scene variations. To minimize both memory consumption and computational burden, the LCM incorporates sparse point hints from Structure from Motion (SfM) into the aggregation process, facilitating the direct computation of source-view features.

LCM-based MVS approach has two stages. The first stage is to use LCM MVS to realize depth inference. The second stage is to develop the point cloud by the depth map filtering & fusion. The overview of the proposed network is shown in Fig. 2.

2.2.1. *Stage 1: depth inference*

Here, four procedures to realize the depth inference are introduced.

2.2.1.1. Feature pyramid extraction. Compared to previous methodology of 3D reconstruction, this study strengthened the shallow feature information flow by designing a bottom-up pathway. This augmentation facilitates the transmission of low-level features and extends the receptive field, enabling the integration of global context information. Consequently, this approach enhances the accuracy and robustness of feature matching in scenarios with limited surface texture.

In this network, multi-view images $\{\mathbf{I}i\}_{i=0}^{N}$ should be given first. $(L+1)$-level features $\{\mathbf{f}_{l,i} \in \mathbb{R}^{F_l \times H/2^l \times W/2^l}\}_{l=0}^{L}$ are extracted for each image $\mathbf{I}_i$, where F_l is the channel number at l-th level, H and W are the height and width, respectively. L is set as 3-level in this study. In this setup, the spatial resolution of the 3-level feature extraction module is as follows: $H \times W$, $H/2 \times W/2$, and $H/4 \times W/4$, respectively. Values of 8, 16, and 32 have been assigned to F_l for $l = 0$, 1, and 2, respectively.

2.2.1.2. Adaptive cost volume aggregation. Next, the image features $\{\mathbf{f}_{l,i} \in \mathbb{R}^{F_l \times H/2^l \times W/2^l}\}_{i=0}^{N}$ of $(N+1)$-view images $\{\mathbf{I}_i\}_{i=0}^{N}$ and camera parameters are required to encode into the network. This process is pivotal in constructing multi-view feature volumes and aggregating cost volumes.

At level l, a uniform sampling strategy is applied to generate $(M_l + 1)$ depth hypotheses distributed across a 3D space. These hypotheses are sampled from the depth range $[d_{\min,l}, d_{\max,l}]$ within the reference camera frustum. The normal vector $\mathbf{n}_0$ coincides with the principal axis of the reference camera.

$$d_{m,l} = d_{\min,l} + m\frac{d_{\max,l} - d_{\min,l}}{M_l} \quad m \in \{0, 1, ..., M_l\} \tag{1}$$

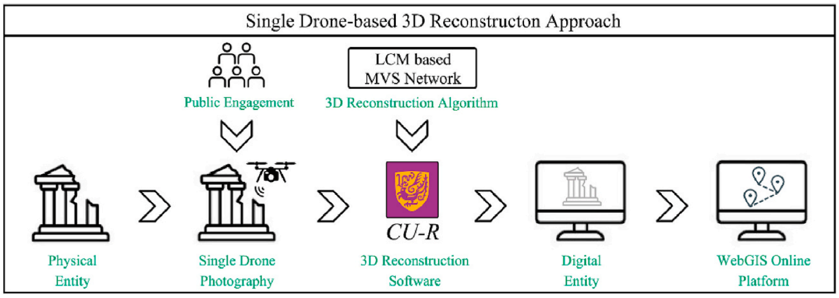


Fig. 1. The process of the single-drone 3D reconstruction approach for public involved sustainable architectural heritage conservation.

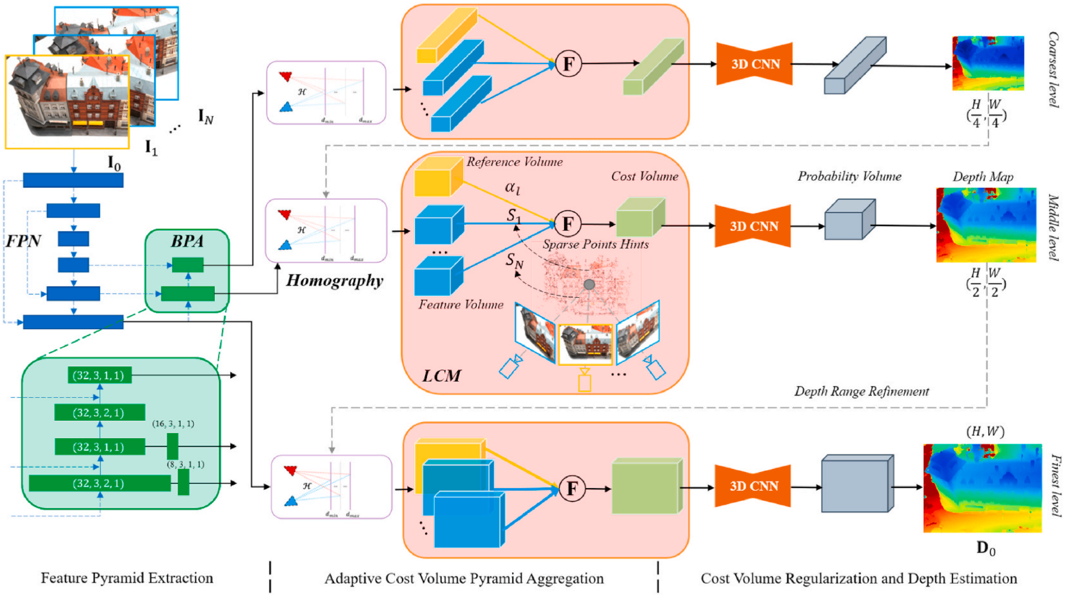
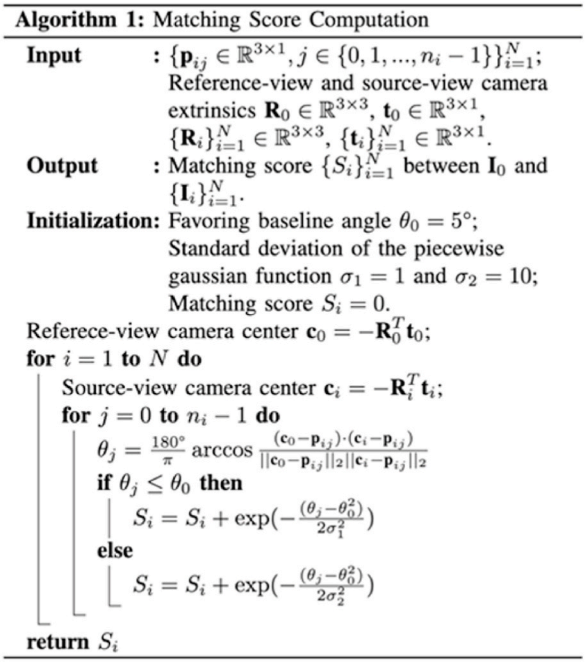


Fig. 2. Overview of the proposed network.

By using the sampled depth hypotheses, source-view features are developed in the 3D space by homography transformation, which involves warping the 2D image features extracted from the source-view into the reference camera frustum to construct the $(N+1)$-view features $\{\mathbf{V}_{l,i} \in \mathbb{R}^{F_l \times M_l \times H/2^l \times W/2^l}\}_{i=0}^{N}$.

To accommodate any amount of input views, the $(N+1)$-view cost volumes $\{\mathbf{V}_{l,i} \in \mathbb{R}^{F_l \times M_l \times H/2^l \times W/2^l}\}_{i=0}^{N}$ are consolidated to a unified volume $\mathbf{C} \in \mathbb{R}^{F_{l,c} \times M_l \times H/2^l \times W/2^l}$ for the purpose of measuring the matching degree of multi-view features. This step can be designed as a mapping function $M : \mathbb{R}^{F_l \times M_l \times H/2^l \times W/2^l} \times \cdots \times \mathbb{R}^{F_l \times M_l \times H/2^l \times W/2^l} \rightarrow \mathbb{R}^{F_{l,c} \times M_l \times H/2^l \times W/2^l}$.

It is observed that an image taken close to the reference view and free from occlusions can provide more precise photometric and geometric data compared to a distant image that is partially obscured. Based on this trend, the LCM is proposed and explained here.



```
Algorithm 1: Matching Score Computation
Input          : $\{\mathbf{p}_{ij} \in \mathbb{R}^{3\times1}, j \in \{0, 1, ..., n_i - 1\}\}_{i=1}^{N}$;
                 Reference-view and source-view camera
                 extrinsics $\mathbf{R}_0 \in \mathbb{R}^{3\times3}$, $\mathbf{t}_0 \in \mathbb{R}^{3\times1}$,
                 $\{\mathbf{R}_i\}_{i=1}^{N} \in \mathbb{R}^{3\times3}$, $\{\mathbf{t}_i\}_{i=1}^{N} \in \mathbb{R}^{3\times1}$.
Output         : Matching score $\{S_i\}_{i=1}^{N}$ between $\mathbf{I}_0$ and
                 $\{\mathbf{I}_i\}_{i=1}^{N}$.
Initialization: Favoring baseline angle $\theta_0 = 5°$;
                 Standard deviation of the piecewise
                 gaussian function $\sigma_1 = 1$ and $\sigma_2 = 10$;
                 Matching score $S_i = 0$.
Referece-view camera center $\mathbf{c}_0 = -\mathbf{R}_0^T \mathbf{t}_0$;
for $i = 1$ to $N$ do
    Source-view camera center $\mathbf{c}_i = -\mathbf{R}_i^T \mathbf{t}_i$;
    for $j = 0$ to $n_i - 1$ do
        $\theta_j = \frac{180°}{\pi} \arccos \frac{(\mathbf{c}_0 - \mathbf{p}_{ij}) \cdot (\mathbf{c}_i - \mathbf{p}_{ij})}{||\mathbf{c}_0 - \mathbf{p}_{ij}||_2 ||\mathbf{c}_i - \mathbf{p}_{ij}||_2}$
        if $\theta_j \leq \theta_0$ then
            $S_i = S_i + \exp(-\frac{(\theta_j - \theta_0^2)}{2\sigma_1^2})$
        else
            $S_i = S_i + \exp(-\frac{(\theta_j - \theta_0^2)}{2\sigma_2^2})$
return $S_i$
```

Fig. 3. Algorithm 1: Matching score computation.

2.2.1.3. Learnable Cost Metric. The *LCM* at network level l is set as:

$$\begin{aligned}\mathbf{C}_l &= M\left(\mathbf{V}_{l,o}, \cdots, \mathbf{V}_{l,N}\right)\\ &= M\left(\mathbf{B}_{l,o}, \cdots, \mathbf{B}_{l,N}\right)\\ &= AvgPool\left(\alpha l\mathbf{B}l, 0 \odot \sum_{i=1}^{N} \frac{S_i}{\sum\limits_{i=1}^{N} S_i}\mathbf{B}l, i\right)\end{aligned} \tag{2}$$

Where $\mathrm{B}_{l,i} \in \mathbb{R}^{K\times(F_l/K)\times M_l\times H/2^l\times W/2^l}$ represents the batched volumes after evenly separating the original volumes $\mathbf{V}_{l,i}$ into K batches along the channel dimension. At various network levels, denoted as $l \in \{0, 1, ..., L\}$, we assign distinct learnable values to α_l to capture the significance of the reference view. Additionally, we utilize the normalized matching score $\frac{S_i}{\sum_{i=1}^{N} S_i}$ as the source-view significance to enable the network to adapt to variations within the scene. The Hadamard product ($\odot$) is used to combine the weighted volumes from multiple views. We then employ average pooling along the channel dimension to calculate the multi-view feature matching similarity, which is used to derive the cost volume $\mathrm{C}_l \in \mathbb{R}^{K\times M_l\times H/2^l\times W/2^l}$.

The computational procedure (see Fig. 3) of the matching score $\{S_i\}_{i=1}^{N}$ between the i_{th} source image $\{\mathbf{I}_i\}_{i=1}^{N}$ and the reference image $\mathbf{I}_0$ is detailed in the following algorithm, $\{\mathrm{p_{ij}} \in \mathbb{R}^{3\times 1}, \mathrm{j} \in \{0, 1, ..., \mathrm{n_i}-1\}\}_{\mathrm{i}=1}^{\mathrm{N}}$ is the inhomogeneous coordinates of the common 3D points visible in both reference image and i_{th} source image, where n_i is the total number of points triangulated by reference view and i_{th} source view. $\{\mathbf{c}_0, \mathbf{c}_i\} \in \mathbb{R}^{3\times 1}$ is the inhomogeneous coordinates of the reference-view and i_{th} source-view camera center, respectively, and θ_j is the baseline angle of $\mathbf{p}_{ij}$. The matching score S_i is accumulated based on a piecewise gaussian function favoring the particular baseline angle θ_0. The normalized matching score $\frac{S_i}{\sum_{i=1}^{N} S_i}$ as the i_{th} source-view significance is defined to ensure adaptability of the network to variations in the input scene.

2.2.1.4. Cost volume regularization and depth estimation. A four-scale 3D convolutional neural network (CNN) is applied to regularize the aggregated cost volume pyramid $\{\mathbf{C}_l\}_{l=0}^{L}$ and generate the probability volume pyramid $\{\mathbf{P}_{l,est}\}_{l=0}^{L}$ using the sigmoid activation function. To achieve continuous depth estimation, we refine the discrete depth estimation by considering the estimated bias between the target depth and the discretized depth:

$$\mathbf{D}_{l,est} = \underset{dm,l\in[d\ min,l,d\ max,l]}{\operatorname{argmax}} \mathbf{P}l, est(dm, l) + \frac{(d\ max, l - d\ min, l)}{Ml} \max \mathbf{P}l, est(dm, l) \tag{3}$$

where $\mathbf{P}l, est(dl)$ is the probability map at depth hypothesis d_l. $\mathbf{D}_{l,est}$ is the depth estimation at level l. $\underset{dm,l\in[d\ min,l,d\ max,l]}{\operatorname{argmax}} \mathbf{P}l, est(dm, l)$ is the discrete depth. $\frac{(d\ max,l-d\ min,l)}{Ml}$ is the depth interval. $\max \mathbf{P}l, est(dm, l)$ is the normalized bias. So, it can be known that $\frac{(d\ max,l-d\ min,l)}{Ml} \max \mathbf{P}l, est(dm, l)$ is the estimated bias.

2.2.1.5. Loss function. For network training, this study uses the focal loss to provide direct supervision for the probability volume [37, 38].

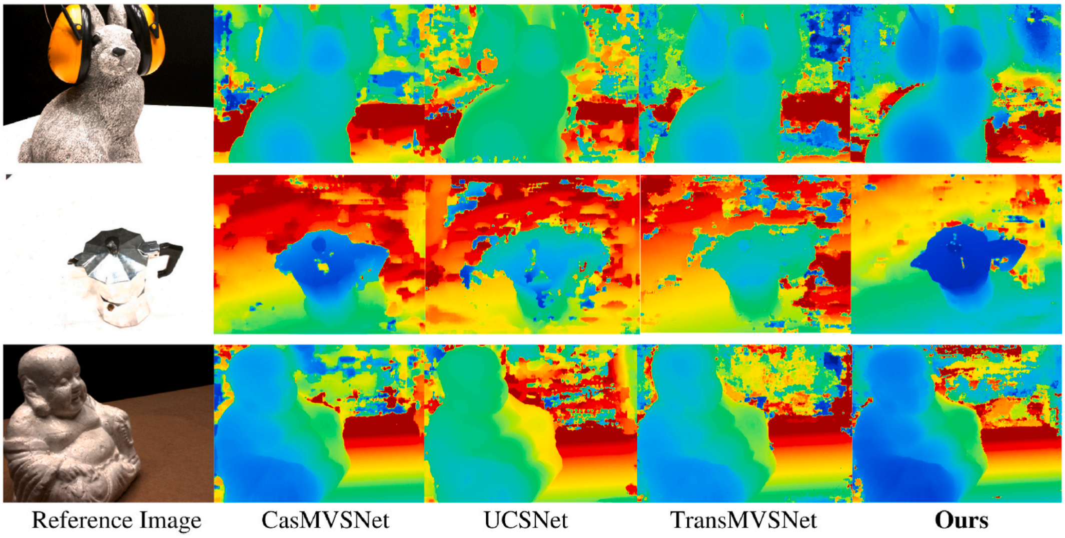


Fig. 4. Qualitative comparison of the depth map estimations on the DTU evaluation set.

2.2.2. Stage 2: depth map fusion

Depth map fusion are carried out to merge the inferred multi-view depth maps $\{\mathbf{D}_i\}_{i=0}^{N}$. When we filter depth maps, a probability threshold τ is set to eliminate depth outliers and specifying the amount of consistent views N_c to mitigate depth inconsistencies. The photometric constraint evaluates the matching quality of multi-view depth maps, while the geometric constraint assesses the consistency of multi-view depth maps. Then, the inferred depth maps are combined to generate the ultimate point cloud.

2.2.3. Experiment on benchmarking dataset

This study conducts experiments on DTU training dataset to evaluate the reconstruction performance and on the BlendedMVS database to assess the depth estimation performance. Moreover, the reconstruction in large-scale scenarios is also assessed on the Tanks and Temples advanced database.

2.2.3.1. Reconstruction performance. Our MVS network is trained using the DTU evaluation set and subsequently evaluated on the DTU evaluation set to quantitatively benchmark its reconstruction performance. Qualitative comparison of the depth map estimations is shown in Fig. 4. In the DTU dataset, the assessment of MVS reconstruction quality relies on accuracy and completeness, both measured in mean error distance metrics (in millimeters, with lower values indicating better results). To obtain a concise evaluation, the DTU dataset calculates the arithmetic mean of accuracy and completeness, which is referred to as the overall score.

Then the method is rigorously benchmarked against the DTU evaluation dataset, where it undergoes comprehensive comparisons with traditional geometric approaches and recent learning-based MVS methods. The standard evaluation procedure [39] is followed, yielding quantitative benchmark results presented in Table 1. These results are measured in millimeters, with lower values indicating enhanced reconstruction accuracy, completeness, and overall score.

The approach consistently shows a commendable balance between accuracy and completeness in reconstruction, spanning various scenarios. It consistently exhibits superior performance compared to both traditional and learning-based methods currently available, excelling in accuracy, completeness, and overall scores, thus establishing its state-of-the-art performance.

Furthermore, a qualitative analysis has been conducted to evaluate depth estimation and reconstruction outcomes for scenes with varying reflectivity, low-textured surfaces, and lighting variations. These assessments have been carried out on the DTU evaluation set. Fig. 4 shows the method's ability to achieve comprehensive depth estimation and generate dense point cloud reconstructions while preserving fine-grained details. The achievement can be attributed to the proposed LCM scheme and provides qualitative confirmation of the quantitative comparative results.

2.2.3.2. Depth estimation performance. The proposed network is specifically designed for depth map estimation. So, the quality of depth map estimation is assessed using the BlendedMVS set. In the BlendedMVS dataset, the quality of depth estimation is evaluated using several metrics, including *end-point error* (EPE), 1-threshold error e_1, and 3-threshold error e_3. These metrics help assess the accuracy and precision of depth estimation. Comparison on depth estimation performance based on BlendedMVS are shown in Table 2.

To highlight the superior accuracy of the proposed network in depth map estimation, a quantitative comparison was conducted using the BlendedMVS validation set. All methods, including ours, have been evaluated using the original input image resolution of 768 × 576, and the number of input views has been uniformly set to 5 to maintain a fair basis for comparison. In Table 2, the results highlight the impressive performance of our method, evident in the lowest EPE, e_1, and e_3 values. This underscores our method's capability to infer high-quality depth maps effectively.

2.2.3.3. Reconstruction error. We carried on experiments on the Tanks and Temples benchmark (advanced set) and obtained the F-score (in %, higher the better). The visualization of the errors is provided in Fig. 5. The proposed MVS network shows a competitive performance. Compared to CasMVSNet, TransMVSNet, and NR-MVSNet, our MVS network achieves the highest F-score.

2.2.3.4. Scalability. To assess the method's adaptability on the building scale, three building scenes, including a 5-floor historical tower, a 4345 m^2 temple, and 18-floor modern building, are selected for the reconstruction using the proposed network strategy. In Fig. 6, results show large-scale buildings are reconstructed by our network in high completeness with fine details. In summary, our proposed method is well-suited for scenes at the building scale, demonstrating a competitive performance in 3D reconstruction.

Table 1
Comparison on reconstruction performance based on DTU dataset.

Methods	Mean Error Distance (mm)			Reference
	ACC. ↓	Comp. ↓	Overall ↓	
Gipuma	0.283	0.873	0.578	[40]
COLMAP	0.400	0.664	0.532	[41]
CasMVSNet	0.325	0.385	0.355	[35]
TransMVSNet	0.360	0.271	0.316	[42]
IGEV-MVSNet	0.331	0.316	0.324	[43]
N2MVSNet	0.336	0.295	0.316	[44]
DispMVS	0.354	0.324	0.339	[45]
Ours ($N = 5$, $N_c = 6$)	0.263	0.539	0.401	
Ours ($N = 5$, $N_c = 3$)	0.368	0.263	0.315	

↓ means lower the better reconstruction performance.

Table 2
Comparison on depth estimation performance based on BlendedMVS.

Methods	EPE ↓	e_1 ↓	e_3 ↓	Reference
EPP-MVSNet	1.17	12.66	6.20	[46]
UniMVSNet	1.17	11.27	4.96	[45]
TransMVSNet	0.05	13.74	5.47	[42]
Ours	1.02	10.15	4.54	

↓ means lower the better depth estimation performance.

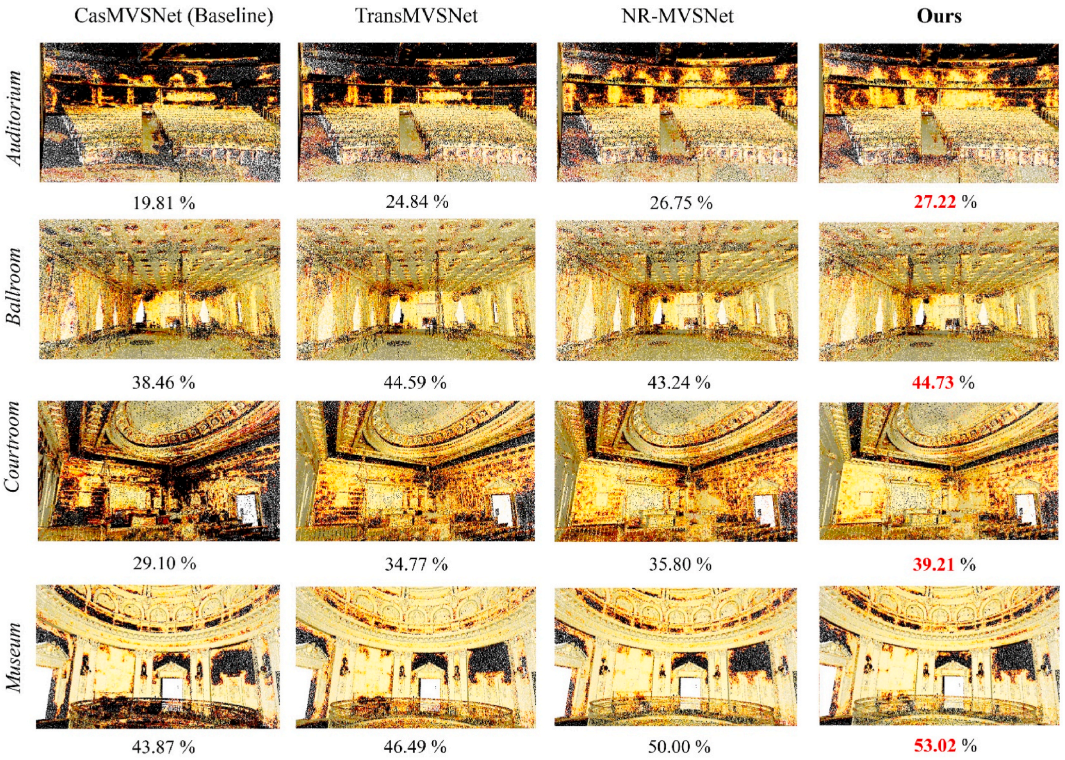


Fig. 5. Visualization of point cloud reconstruction error. The F-score is given blow the reconstruction model.

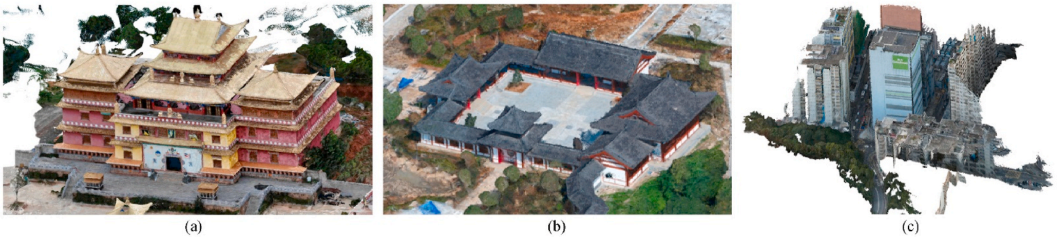


Fig. 6. 3D reconstructed models of three large-scale building by our method.

3. Case study and satisfaction survey

3.1. Hakka Tulou - Zhenchenglou

The Hakka Tulou structures, located in the southwestern region of Fujian province and Guangdong province, China, stand out as a remarkable part of international built heritage (Fig. 7) [47]. In 2008, 46 buildings of Hakka Tulou were formally listed on the World Heritage Sites [48]. The Hakka Tulou structures are the foremost, most representative, and exceptionally well-preserved illustrations found in the mountainous areas of southeastern China. These large, technically advanced, and striking earthen buildings were constructed over the course of several centuries, spanning from the 13th to the 20th century. Specially, these vernacular buildings used earthen materials and were primarily constructed for defensive purposes, featuring a distinctive layout with a central open courtyard, a

Fig. 7. Hakka Tulou [50].

single entrance, and windows on the top floor only. They are recognized as exceptional exemplars of a specific architectural tradition and functional design, manifesting a distinctive manifestation of communal habitation and defensive structuring [49]. Furthermore, their harmonious integration with the natural environment makes them an outstanding representation of human settlement.

The case Tulou of this research is Zhenchenglou, which is located at coordinates 24.66°N and 116.98°E in Fujian province, as shown in Fig. 8. Zhenchenglou is distinguished as a double-circle Tulou, characterized by its outer circle consisting of four stories with a total of 184 rooms, and the inner circle, which is two stories high and contains 32 rooms. It is recognized for its imposing fortified mud walls crowned with tiled roofs featuring wide overhanging eaves.

Zhenchenglou stands as one of the paramount historical structures in Fujian province, demanding immediate and substantial attention for preservation and documentation [50]. The scale of Zhenchenglou is a big challenge for traditional LiDAR scanning method. Zhenchenglou stands at a height of over 10 m and has a diameter of nearly 60 m, covering an area of approximately 5000 square meters. Moreover, the surrounding environment is complex and not friendly for ground scanning. The building is nestled amidst the picturesque backdrop of rice, tea, and tobacco fields. Qingchanglou (another Tulou) is around 30 m far away on the northwest of Zhenchenglou. If employing ground scanning, it is necessary to set up a greater number of LiDAR scanning stations strategically around the building in confined and obstructed street environments in order to ensure comprehensive data collection for the 3D geometric properties of the building and enable the capture of intricate details. This would also cause inference with the daily life of local residents.

3.2. Data collection

A drone, DJI Mavic 2 enterprise advanced, was applied to scan the Zhenchenglou. The drone was outfitted with a CMOS sensor boasting 48 million effective pixels, capturing images with a maximum resolution of 8000 × 6000 pixels. The DJI mavic 2 enterprise advanced are shown in Fig. 9 (a). One flight mission was performed around the heritage building with planned route. The planned route is shown in Fig. 9 (b). During the flight mission, planned using the DJI Pilot software [51], the drone maintained an altitude of approximately 20 m above the historical heritage structure. The parameters for lateral and frontal image overlap were thoughtfully configured at 40%. As a result, a total of 248 high-resolution images were captured specifically for the purpose of 3D reconstruction of the Zhenchenglou.

3.3. Satisfaction survey

In this section, the survey is provided to assess the satisfaction level of heritage conservation practitioners in the reconstruction quality by CU-Recon.

3.3.1. Satisfaction

At times, the perceived shared language among individuals may not be as shared as presumed, revealing itself through subtle nuances or, in some instances, through significant disparities. Users always hold unbeknownst the disciplinary subjectivities to themselves. Therefore, the degree of satisfaction is introduced in this study to evaluate the reconstruction quality by CU-Recon.

The concept of satisfaction has been subjects of frequent discussions in previous academic works, with several definitions being advanced. There are three main perspectives for satisfaction [52]. The first perspective, referred to as the purposive approach, defines satisfaction as an assessment of the extent to which an entity either facilitates or inhibits users in achieving their objectives. The second

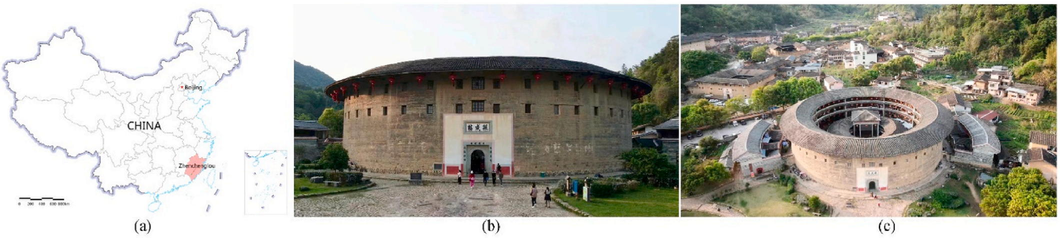


Fig. 8. Hakka tulou. (a) Location. (b)(c) Zhenchenglou.

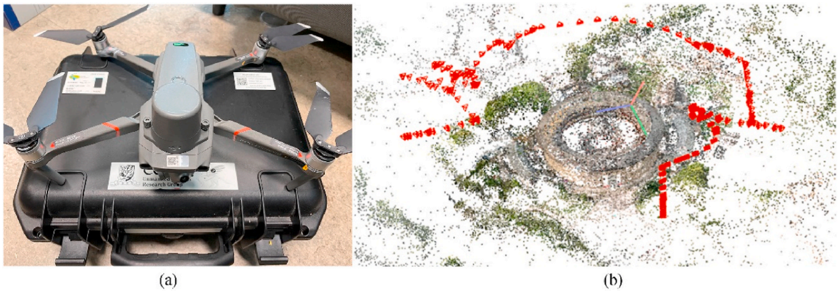


Fig. 9. (a) DJI Mavic 2 enterprise advanced. (b) Drone planned route.

perspective, referred to as the actual-aspiration gap approach, defines satisfaction as an assessment of the disparity between users' existing and desired needs. The third perspective conceptualizes satisfaction as a multi-faceted construct with cognitive, affective, and conative dimensions, forming an attitude towards the subject. This paper uses the second perspective, assuming participants recognize significant attributes of the reconstructed heritage and assess them through a comparative analysis. The comprehensive level of satisfaction is ultimately shaped by two distinct sets of objective factors. One set is "contextual": characteristics of product-related experience. Another is "compositional": characteristics of the users. These two sets of factors can have a direct impact on satisfaction or work indirectly by affecting users' subjective attitudes and evaluations of particular aspects of the product, subsequently affecting overall satisfaction.

3.3.2. *Pilot questionnaire*

Prior to distributing the official questionnaire, we initiated a pilot survey in September 2023 involving five practitioners specializing in heritage preservation. The primary objective of this pilot survey was to assess the questionnaire's suitability in terms of length, clarity of terminology and questions, and coverage of essential factors. Following the feedback obtained from the participants in the pilot survey, we made necessary adjustments, including reordering questions and refining those that lacked clarity.

3.3.3. *Formal questionnaire*

For the formal questionnaire, based on the pilot survey, an ad-hoc survey was built to collect information on (a) Personal data, (b) Drone experience, (c) Satisfaction levels in reconstruction quality of CU-Recon, and (d) Satisfaction levels in scanning technologies. (a)(b) stand for the survey of characteristics of the users. (c)(d) stand for the survey of product-related experience.

Their personal data, including the age, gender, working years in heritage protection, education level (bachelor, master, Ph.D.), and countries were collected after obtaining their informed consent, in order to show compositional effects. Assessment of participant experience on drone is essential to show contextual effects on the satisfaction levels.

According to the concept on the conservation of historical sites in Venice Charter, London, Charter, the Principles of Seville, Xi'an Declaration and other documents [53–55], four terminologies are selected to evaluate the reconstruction quality in software CU-Recon, including clearness, accuracy, integrity, and authenticity. The mean satisfaction value of the four aspects is defined as the satisfaction of the reconstruction quality.

Table 3
Questionnaire.

	Questions			
Personal data				
A1	Age (S.D.)			
A2	Gender	Male	Female	
A3	Working years in heritage protection	0–5	6–10	>10
A4	Education level	Bachelor	Master	Ph.D.
A5	Working country	Developed	Developing	
Drone experience				
B1	Do you have one drone?	0 = No 1 = Yes		
B2	Do you have a drone driving license?	0 = No 1 = Yes		
B3	Did you operate the drone before?	0 = No 1 = Yes		
B4	Did you take photos for buildings by drone?	0 = No 1 = Yes		
Satisfaction levels on reconstruction quality of CU-Recon				
C1	How satisfied are you with the three aspects of reconstructed models by SD-3DR? a) Clearness; b) Accuracy; c) Integrity; d) Authenticity	1–5		
Satisfaction levels on scanning technologies.				
D1	How satisfied are you with LiDAR-based scanning? a) Portability; b) Operation; c) Cost	1–5		
D2	How satisfied are you with single drone photography? a) Portability; b) Operation; c) Cost	1–5		

The evaluation uses a five-point Likert scale consisting of five points, ranging from "very dissatisfied" at 1 to "very satisfied" at 5. 1 stands for "very dissatisfied". 2 stands for "dissatisfied". 3 stands for "neutral". 4 stands for "satisfied". 5 stands for "very satisfied".

The LiDAR scanning technology and single drone photography are also compared from three aspects, including the portability, operation, and cost. The participants' responses were evaluated using a five-point Likert scale consisting of five points, ranging from "very dissatisfied " at 1 to " very satisfied " at 5. The questionnaire is shown in Table 3.

3.3.4. Participants

Participants were chosen employing the virtual snowball sampling method. This approach is notably advantageous due to its efficacy in expanding the geographical scope and reaching individuals facing accessibility challenges [56]. The rationale behind employing the virtual snowball sampling method lies in its capacity to augment the number of cases in the sample while enhancing representativeness, as it allows for control over the quantity and diversity of responses throughout the process.

Following the prescribed steps of the snowball sampling technique [57], we initially identified the primary respondents, who served as referrers and were drawn from the friends known to the researchers and respondents known through the pilot study mentioned above. Subsequently, we conducted an eligibility assessment for these referrers and selected a group of 10 individuals who expressed their willingness to participate as research assistants, aiding in the dissemination of the questionnaire. It is noteworthy that the 15 assistants were distributed across various countries in the world, ensuring a comprehensive geographic representation.

The distribution of the questionnaire commenced in September 2023, primarily via email, WhatsApp, WeChat and other virtual social applications. To maintain the sample's representativeness, we also made efforts to regulate the direction of the referral chain in terms of educational backgrounds and geographical locations, to the greatest extent possible. By October 20, 2023, we had accumulated a total of 64 responses, thereby fulfilling our stipulated sample size requirement. All participants were informed that participation in the study was voluntary. These responses exhibited a commendable level of representativeness, particularly concerning education levels and geographical origins. In the interest of data quality, 3 responses were excluded due to their low response quality, characterized by patterns such as selecting the first option for all questions, and notably brief response times of less than 5 min. This meticulous filtering process ultimately resulted in a dataset of 61 valid responses.

3.3.5. Statistical methods

The mean value of the four items in questions C1 was calculated to evaluate satisfaction levels on the reconstruction quality. The mean value of the three items in questions D1 and D2.

To test the appropriateness of averaging the items, Cronbach's alpha, a reliability statistic, was utilized. Typically, an alpha value exceeding 0.70 is indicative of a reliable scale [58]. The results revealed a strong correlation (Cronbach's alpha = 0.774) among these items, based on a sample size of 61. This suggests that elements are internally connected and can be consolidated into a single composite score for assessing the satisfaction level of the reconstruction quality.

The survey aims to investigate the satisfaction level on the reconstruction quality of the heritage building in CU-Recon software and compare the single drone photography with LiDAR-based scanning technology from the perspective of practitioners. We used IBM SPSS Statistics19 software to analyze the data. First, a one-way ANOVA was carried out on to evaluate satisfaction in the single drone-based 3D reconstruction approach. Next, a multivariate regression analysis was conducted to figure out the factors influencing satisfaction level in practitioners. To simplify the model, we employed the backward elimination-by-hand procedure [59]. It serves as entering all variables simultaneously into the regression analysis. Moreover, Ordinary Least Squares (OLS) regression analysis was conducted in software SPSS. Subsequently, the predictor which shows the highest p-value was systematically excluded and this step-wise elimination was iterated. The meticulous application of this method continued until only statistically significant predictors

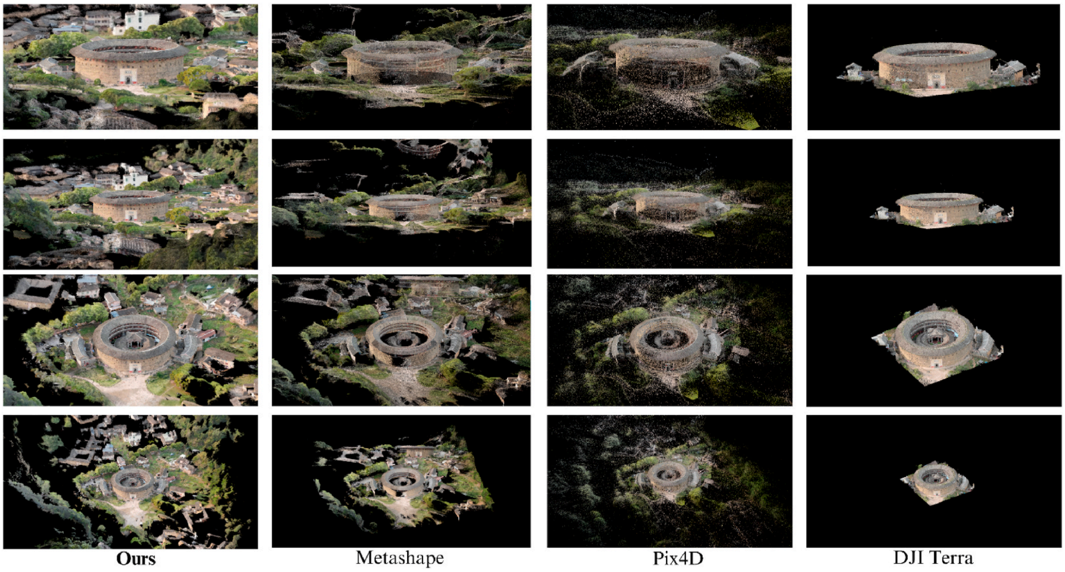


Fig. 10. Comparison of the 3D reconstruction quality in different software.

remained. Dummy variables were created for all categorical variables, and missing values were coded into the "others" category to maximize inclusion of respondents in the regression analysis.

The mean scores of (a)(b)(c) and (d) can directly show the satisfaction level on the reconstructed models by CU-Recon and effectiveness of the SD-3DR approach. The multivariate regression analysis shows the correlation between the product-related experience (c) (d) and characteristics of the users (a)(b). The correlation can reveal the satisfaction level of different groups on the SD-3DR approach.

4. Results

In this section, the 3D model of Zhenchenglou is reconstructed with CU-Recon. The results are compared with the model reconstructed in other commercial software. Moreover, the satisfaction of heritage conservation practitioners on the reconstruction quality and on the single drone-based 3D reconstruction are shown here.

4.1. 3D reconstruction

The reconstruction was conducted on an NVIDIA 3090ti GPU. 248 images were input into software CU-Recon, Metashape, Pix4D, DJI Terra, and Colmap in order to reconstruct the 3D point cloud of Zhenchenglou. The results are shown in Fig. 10. Software Colmap failed in this reconstruction task, it requires more memory when inputting the same number of photos. In this case, the limitation of software Colmap results in “out of memory”. For DJI Terra, the scope for reconstruction is required to be selected firstly. So, the reconstruction in DJI Terra only shows the structure of Zhenchenglou without surrounding landscape. Notably, DJI Terra faces limitations in realizing the reconstruction process without the presence of the Global Positioning System (GPS), as it relies on GPS for image matching purposes.

Using CU-Recon, we completely reconstructed all of the Zhenchenglou’s elements, even including structures, decorations, landscapes, and textures. The detailed surface information of the historical building and its surroundings are well represented. Moreover, the aerial images and corresponding depth maps from the four views are shown in Fig. 11. The smoothness of edges and sharp boundaries are more obvious in the depth maps. Compared to the other reconstructed models, the 3D geometric information in our software has a greater integrity. *CU-Recon* produces the clearest textures, which show the minimal distortions among all the solutions. The results in *DJI Terra* show incomplete reconstruction on the façade, because points are sparse. In the visualization results, it can be seen that *Metashape* and *Pix4D* cannot create a complete model of the studied historical building. The results of the two software lose more details in the main structure of the Zhenchenglou. The result of *Pix4D* is oversimplified, resulting in more noises. It is also noted that the reconstruction time in *CU-Recon* is far shorter than that of other software. Overall, our software exhibits outstanding performance and generalization ability when encountering unseen large-scale architectural heritage without any postprocessing.

4.2. Basic information about the participants

A total of 61 valid participants responded to the survey. Basic information about the participants were collected and arranged in Table 4. The mean age is 36.2. 59% of participants are male. All participants have experience in heritage protection. 23 participants work more than 10 years in the field of heritage protection, accounting for 38%. 74% of participants have more than 5-year work experience in this field. 23% of participants have work experience within 5 years. 32 of the participants have a master’s degree, but most of them are Ph.D. candidates, and will get their diplomas within 3 years. 39% of participants have obtained Ph.D. degree. For working countries, 32 of the participants are working in developed countries and others are working in developing countries.

For drone experience, the results are shown in Table 5. From the results, it can be known that the mean value of the question “Do you have one drone?” is 0.42. It means that 42% of the participants own one drone. But only 21% of participants have a drone operating license. The reason of this is because some countries do not introduce regulations for the drone operation. And although many countries have introduced relevant restrictive policies, enforcement is relatively lax. Therefore, the percentage of participants who own one drone operating license is so low. For the question “Have you operated the drone before?”, the mean value is 0.76, which means that 76% of participants have the experience to operate one drone. It is far higher than the average value of normal people, which is only 16% [60]. And this survey cares more about the experience of taking photos of buildings with drones, which is related to the topic using a single drone to record the architectural heritage. Compared to the 76% of participants who operated the drone, only 34% of participants have the experience of taking photos of buildings with drones.

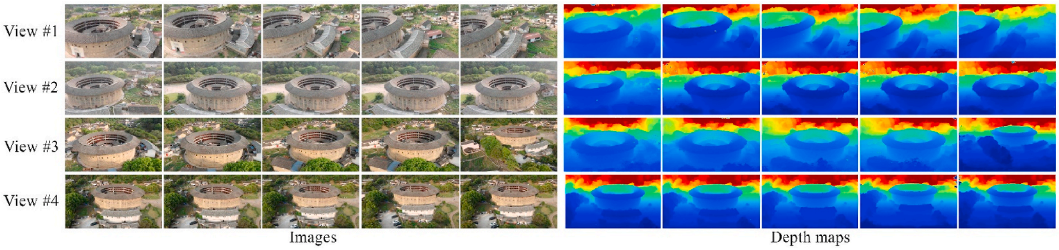


Fig. 11. The aerial images and corresponding depth maps from the four views.

Table 4
Personal data of participants.

Questions	N	%
Age		
Mean	36.2	
Median	33	
Std. Deviation	12.6	
Gender		
Male	36	59
Female	25	41
Working years in heritage protection		
1–5	16	26
6–10	22	36
>10	23	38
Education level		
Bachelor	5	8
Master	32	52
Ph.D.	24	39
Working Country		
Developed	32	52
Developing	29	48

Table 5
Drone experience of participants.

Questions	Mean
Do you have one drone?	0.42
Do you have a drone driving license?	0.21
Have you operated the drone before?	0.76
Did you take photos of buildings with drone?	0.34

4.3. Satisfaction level

The assessment of participant satisfaction on the 3D reconstructed model generated by *CU-Recon* involves four aspects, with mean satisfaction scores presented in Table 6. Regarding the clearness of the reconstructed model, the survey yielded a mean value of 3.6 with a standard deviation of 1.4. Notably, participants working in the field of heritage conservation expressed the lowest satisfaction level on clarity among the four evaluated aspects, with practitioners indicating a preference for rendered mesh models over 3D point cloud. For the accuracy of the 3D point cloud, the mean satisfaction score from the survey is 4.2. In terms of the integrity of the 3D point cloud, the mean satisfaction value is the highest among the four items, reaching 4.5, with the lowest standard deviation of 0.6, indicating a uniformly high satisfaction level with integrity. Additionally, participants expressed a mean satisfaction score of 4.4 for authenticity.

The assessment of participant satisfaction with the two scanning technologies is based on three aspects. The results are shown in Table 7. From the results, satisfaction levels on the three aspects showed the same pattern. Satisfaction levels were markedly higher for single drone photography compared to LiDAR-based scanning. In terms of instrument portability, the mean satisfaction score for single drone photography stands at 4.1, far higher than LiDAR-based scanning, which achieved a score of 3.3. In terms of operation, the disparity in satisfaction levels becomes more pronounced. Single drone photography received a mean satisfaction score of 4.5, while LiDAR-based scanning lagged behind with a score of 2.9. Additionally, when evaluating the cost, a pivotal factor in public engagement, participants expressed really high satisfaction, yielding a mean score of 4.6 for single drone photography. It is noted that the proportion of participants who were "very satisfied" is up to 63%, obviously higher than the other two aspects. Conversely, LiDAR-based scanning received a notably lower mean satisfaction score of 1.8 in this regard. Then, a one-way ANOVA was used in this study. The results showed that single drone photography in terms of portability and costs received significantly higher levels of satisfaction compared to LiDAR-based scanning ($p < 0.001$), while this difference is not significant in terms of operation ($p = 0.052$).

Table 6
Satisfaction levels on reconstruction quality of *CU-Recon*.

Questions	Mean	Std. Deviation
How satisfied are you with the four aspects of reconstructed models by CU-Recon?		
a) Clearness	3.6	1.4
b) Accuracy	4.2	0.6
c) Integrity	4.5	0.6
d) Authenticity	4.4	0.8

Table 7
Satisfaction levels on scanning technologies.

Questions	LiDAR-based scanning		Single drone photography	
	Mean	Std. Deviation	Mean	Std. Deviation
a) Portability	3.3	1.4	4.1	0.5
b) Operation	2.9	1.4	4.5	0.7
c) Cost	1.8	1.9	4.6	0.8

4.4. Regression results

In this section, the determinants of satisfaction on the reconstruction quality of *CU-Recon* are figured out by multivariate regression. The dependent variable is the satisfaction level, which is the mean value of the four items. The independent variables are personal data and drone experience, including working years, working countries, “Do you have one drone?”, “Do you have a drone driving license?”, “Did you operate the drone before?”, “Did you take photos for buildings by drone?”. The outcomes are provided in Table 8. Notably, variables, like the gender, and education, were excluded from the analysis due to their insignificance in each model, indicating no significant relationship with satisfaction regarding the reconstruction quality. The model showed a notable variance in satisfaction levels, with a predictive power R^2 of 0.482.

Table 8 shows that “owning a drone”, “operated a drone”, and “took photos of building with drone” are related to higher satisfaction level on the reconstruction quality. “Owning a drone” increases satisfaction by 0.19. For “operated a drone”, it is a statistically significant predictor, which can enhance satisfaction by 0.28. Moreover, the satisfaction can be increased by “took photos for building by drone” at 0.21. Besides these three indicators, judging from the standardized coefficients, 1–5 working year is the most statistically significant predictor (−0.45). With the working year increasing, the satisfaction goes down. This may be because practitioners with more working experience are too familiar with the traditional methods to accept new digital methods. Surprisingly, the working country development level also influences their satisfaction level. Working in a developed country increases satisfaction by 0.23, which show more influence than working in a developing country (0.16).

5. Discussion

In this section, the significance, limitations, and future work of this study are shown.

5.1. Significance of the single drone 3D reconstruction approach

Engaging the public in heritage conservation endeavors can significantly enhance the sustainability of such initiatives. This notion aligns with the principles set forth in the Washington Charter, which emphatically emphasizes the pivotal role of residents’ participation and involvement in the success of conservation programs [61]. Heritage, as perceived and embraced by the public, can often better encapsulate the cultural significance of built assets than designations determined solely by expert evaluations. Previous studies have primarily explored the potential roles of public participation in the decision-making processes of heritage conservation [62]. In the decision-making processes, the effectiveness of public engagement hinges on the ability to influence decision-making [63]. In contrast, primarily tokenism entails a superficial involvement of community participants who serve as information providers and, at best, participate in consultations without wielding substantial decision-making influence [64].

However, achieving genuine empowerment of the public remains a challenging endeavor. Thus, it is imperative to depart from the conventional government-led heritage protection framework, which relies on government authority for decision-making and government funding for implementation. A new approach, grounded in mass participation and fostering spontaneous engagement, is needed to enable bottom-up heritage protection, particularly for the preservation of often overlooked architectural heritage [65].

In comparison to maintenance and restoration efforts, documentation represents the most fundamental approach to participating in heritage conservation. Nonetheless, traditional recording techniques demand significant financial resources and professional expertise. The 3D reconstruction tool *CU-Recon* proposed in this article, achieved a high satisfaction following expert evaluation, and are proved to meet the requirements for the reconstruction quality in architectural heritage protection. Importantly, these advancements have substantially reduced the financial costs and technical barriers associated with public participation, which has been proved in the survey. Notably, *CU-Recon* enables 3D reconstruction of architectural heritage through single drone photography, empowering the public to spontaneously document architectural heritage rooted in collective memory. This innovative method opens up one route that was previously inaccessible, allowing heritage once forgotten by the populace, and even deemed of limited heritage value by the government, to regain vitality through digitization.

5.2. Limitations and future work

During the survey, we also had a lot of communications with the practitioners in the field of architectural heritage conservation. They expressed that although the 3D cloud point reconstruction can provide an accurate and complete model, they prefer the rendered mesh model. Therefore, the work at the next stage is to develop an approach to create mesh model based on our 3D reconstruction model. Moreover, while we have made significant strides in 3D reconstruction technology and software that facilitate public engagement in recording architectural heritage, a key challenge remains, as digital information captured by the public now resides solely on their individual devices, limiting its accessibility and impact [26]. To address this issue, there is a pressing need to establish a dedicated digital platform where the public can upload, share, and access these valuable heritage data. Moreover, the web-based

Table 8
Determinants of satisfaction on the reconstruction quality of *CU-Recon*.

	Coeff.	St. Coeff.
Constant	4.175	
Personal data		
Working years		
1–5	−0.453*	−0.163
6–10	−0.152	−0.093
>10	−0.245***	−0.136
Working countries		
Developed country	0.229*	0.128
Developing country	0.163*	0.08
Drone experience		
Do you have one drone?	0.186**	0.113
Do you have a drone driving license?		
Did you operate the drone before?	0.284***	0.17
Did you take photos for buildings by drone?	0.205***	0.128
R square	0.482	

$*p < 0.05$. $**p < 0.01$. $***p < 0.001$.

platform is supposed to provide the online 3D interaction service with low technical barrier and cost. Besides public participation, this platform can serve a dual purpose, also offer a space for publicizing, conducting business related to heritage, and heritage restoration. We have preliminarily leveraged Geographic Information Systems (GIS) and digital twin technology to create an integrated platform to showcase the digitized Tulou and will continue to optimize the user interface and build up a world heritage map [66]. The platform is being developed and improved, as shown in Fig. 12. Based on the models derived from the software tool *CU-Recon*. This platform will offer an innovative technical avenue for public engagement in heritage protection, further advancing the goal of bottom-up, sustainable architectural heritage preservation.

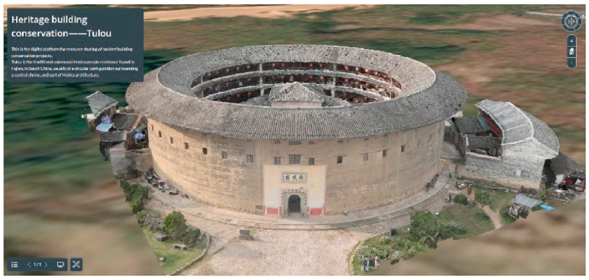
6. Conclusions

This study developed a single drone-based novel 3D reconstruction approach for public engagement into architectural heritage conservation. Firstly, the LCM-based MVS network was developed, enabling the public to employ only one drone for image capture and 3D reconstruction of buildings. Next, a photogrammetry software tool *CU-Recon* was provided for executing the codes of the LCM-based MVS network. *CU-Recon* offers a simple operational interface in order to streamline the 3D reconstruction process for users.

For the LCM-based MVS network, its performance was proved by experiments. The network shows 0.313 mm overall score on the standard MVS benchmarks. The experimental results on the BlendedMVS dataset show that our MVS network achieves the lowest estimation error in the depth estimation. Particularly, our method is well-suited for scenes at the building scale, demonstrating competitive performance in 3D reconstruction.

To verify the effectiveness of the SD-3DR approach, the study selected architectural heritage Hakka Tulou, Zhenchenglou, for a real-world test. The 3D model of Zhenchenglou was reconstructed with CU-Recon and other four commercial 3D reconstruction software. Practitioners in the field of heritage conservation was invited for a satisfaction survey to quantitatively examine the reconstruction quality through CU-Recon and the effectiveness of the SD-3DR approach from a perspective of heritage conservation. For the reconstruction quality evaluation, satisfaction scores were obtained across four aspects: clearness (3.6), accuracy (4.2), integrity (4.5), and authenticity (4.4). For the effectiveness of the SD-3DR approach, participants expressed a mean satisfaction score of 4.175 out of 5. The survey results also show that single drone photography for architectural heritage conservation is preferred over LiDAR-based scanning. It is noted that drone operation experience significantly influences the satisfaction on CU-Recon.

Overall, based on the software tool *CU-Recon*, the SD-3DR approach significantly contributes to improving public engagement in protecting remote and lesser-known heritage buildings. It will change the situation of government-led architectural heritage conservation. The approach provides valuable insights for practitioners in creating bottom-up heritage conservation routes.



Heritage building conservation——Tulou

This is the digital platform for resource sharing of ancient building conservation projects.
Tulou is the traditional communal Hakka people residence found in Fujian, in South China, usually of a circular configuration surrounding a central shrine, and part of Hakka architecture.

Fig. 12. The 3D model on the WebGIS platform.

CRediT authorship contribution statement

Qingxiang Li: Writing – original draft, Methodology, Investigation, Conceptualization. **Guidong Yang:** Methodology, Investigation. **Chuanxiang Gao:** Investigation. **Yijun Huang:** Investigation. **Jihan Zhang:** Investigation. **Dongyue Huang:** Investigation. **Benyun Zhao:** Investigation. **Xi Chen:** Writing – review & editing, Supervision. **Ben M. Chen:** Writing – review & editing, Supervision, Funding acquisition.

Declaration of competing interest

I confirm that we have mentioned all organizations that funded our research in the Acknowledgements section of my submission, including grant numbers where appropriate. We declare that we have no commercial or associative interest that represents a conflict of interest with other people or organizations that can inappropriately influence our work entitled, "Single Drone-Based 3D Reconstruction Approach to Improve Public Engagement in Conservation of Heritage Buildings: A Case of Hakka Tulou".

Data availability

Data will be made available on request.

Acknowledgement

This work was sponsored by the InnoHK of the Government of the Hong Kong Special Administrative Region via the Hong Kong Centre for Logistics Robotics. Appreciation is also given to the CUHK Direct Grant for Research No. 4055168 and the Research Grants Council of the Hong Kong Special Administrative Region, China (UGC/FDS16/E04/21). Authors would like to appreciate the support from Prof. Wenxing Hong.

References

[1] M. Shan, Y.-F. Chen, Z. Zhai, J. Du, Investigating the critical issues in the conservation of heritage building: The case of China, J. Build. Eng. 51 (2022) 104319.
[2] D. Borosnyoi-Crawley, Non-destructive strength estimation of vintage clay bricks based on rebound hardness in architectural heritage buildings, J. Build. Eng. 80 (2023) 108055.
[3] D. Mısırlısoy, Günçe Kjsc, society, Adaptive reuse strategies for heritage buildings, A holistic approach 26 (2016) 91–98.
[4] M.I. Adegoriola, J.H. Lai, E.H. Chan, A.J. Darko, Heritage Building Maintenance Management (HBMM): A Bibliometric-Qualitative Analysis of Literature, J. Build. Eng. vol. 42 (2021) 102416.
[5] A.O. Shehata, N.A. Megahed, M.M. Shahda, A.M.J.B. Hassan, Environment, (3Ts) Green Conservation Framework: A Hierarchical-Based Sustainability Approach, vol. 224, 2022 109523.
[6] Y. Zhao, D. Ponzini, R.J.H.I. Zhang, The Policy Networks of Heritage-Led Development in Chinese Historic Cities: the Case of Xi'an's Big Wild Goose Pagoda Area, vol. 96, 2020 102106.
[7] L. Long, Q. Li, Z. Gan, J. Mu, M. Overend, D. Zhang, Life cycle assessment of stone buildings in the taihang mountains of hebei province: evolution towards cleaner production and operation, J. Clean prod. vol. 399 (2023) 136625.
[8] K. Yuan, Q. Li, W. Ni, X. Lü, G. Della Vecchia, H. Wang, et al., Analysis of the Structural and Environmental Impacts of Hydrophilic ZSM-5 Molecular Sieve on Loess, vol. 366, 2023 130248.
[9] K. Yuan, Q. Li, W. Ni, L. Zhao, H. Wang, Graphene stabilized loess: Mechanical properties, microstructural evolution and life cycle assessment 389 (2023) 136081.
[10] P. Lewińska, M. Róg, A. Żądło, S.J.M. Szombara, To save from oblivion: comparative analysis of remote sensing means of documenting forgotten architectural treasures–Zagórz Monastery complex, Poland 189 (2022) 110447.
[11] Disney Entertainment Company, Film: Coco, 2017.
[12] I.E. Aigwi, A.N. Nwadike, A.T.H. Le, F.E. Rotimi, T. Sorrell, R. Jafarzadeh, et al., Prioritising optimal underutilised historical buildings for adaptive reuse: a performance-based MCDA framework validation in Auckland, New Zealand 11 (2022) 181–204.
[13] R.R. Nadkarni, Puthuvayi BjjoBE, A comprehensive literature review of Multi-Criteria Decision Making methods in heritage buildings 32 (2020) 101814.
[14] J. Gonçalves, R. Mateus, J.D. Silvestre, A.P. Roders, L.J.S.C. Bragança, Society, Attitudes matter: measuring the intention-behaviour gap in built heritage conservation 70 (2021) 102913.
[15] F. Remondino, A.J.A.G. Rizzi, Reality-based 3D documentation of natural and cultural heritage sites—techniques, problems, and examples 2 (2010) 85–100.
[16] Q. Li, L. Zhu, Y. Sun, L. Lu, Y.J.E. Yang, Performance prediction of Building Integrated Photovoltaics under no-shading, shading and masking conditions using a multi-physics model 213 (2020) 118795.
[17] Q. Li, C. Monticelli, A. Kutlu, Zanelli AjjoCP, Feasibility of Textile Envelope Integrated Flexible Photovoltaic in Europe: Carbon Footprint Assessment and Life Cycle Cost Analysis, 2023 139716.
[18] J. Moyano, Justo-Estebaranz Á, J.E. Nieto-Julián, A.O. Barrera, M. Fernández-Alconchel, Evaluation of records using terrestrial laser scanner in architectural heritage for information modeling in hbim construction: the case study of the la anunciación church (seville), J. Build. Eng. vol. 62 (2022) 105190.
[19] G. Pérez, A. Escola, J.R. Rosell-Polo, J. Coma, R. Arasanz, B. Marrero, et al., 3D Characterization of a Boston Ivy Double-Skin Green Building Facade Using a LiDAR System, vol. 206, 2021 108320.
[20] M. Abate, A.C.J. Evangelista, V.W.J.B. Tam, Comparative response spectrum analysis on 15-and 50-story reinforced concrete buildings having shear walls with and without openings as per EN1998-1 seismic code 13 (2023) 1303.
[21] Anna M. Manferdini, M. Russo, Multi-scalar 3D digitization of Cultural Heritage using a low-cost integrated approach, Digital Heritage International Congress (DigitalHeritage) 1 (2013) 153–160, 2013.
[22] F. Jalaei, F. Jalaei, S.J.S.C. Mohammadi, Society, An integrated BIM-LEED application to automate sustainable design assessment framework at the conceptual stage of building projects 53 (2020) 101979.
[23] Q. Li, C. Monticelli, A. Kutlu, Zanelli AjjoCP, Feasibility of textile envelope integrated flexible photovoltaic in Europe: carbon footprint assessment and life cycle cost analysis 430 (2023) 139716.
[24] Q. Li, C. Monticelli, A.J.R.E. Zanelli, Life cycle assessment of organic solar cells and perovskite solar cells with graphene transparent electrodes 195 (2022) 906–917.
[25] T. Wang, V. Gan, Automated joint 3D reconstruction and visual inspection for buildings using computer vision and transfer learning, Auto. Construct. 149 (2023) 104810.
[26] A. Peponi, P. Morgado, P.J.S.C. Kumble, Society, Life cycle thinking and machine learning for urban metabolism assessment and prediction 80 (2022) 103754.

[27] Q.X. Li, A. Zanelli, A review on fabrication and applications of textile envelope integrated flexible photovoltaic systems, Renew Sust Energ Rev 139 (2021).
[28] Q. Li, A.J.R. Zanelli, S.E. Reviews, A review on fabrication and applications of textile envelope integrated flexible photovoltaic systems 139 (2021) 110678.
[29] W.B. Putra, G. Faisal, N.I. Dewi, Y. Firzal, Unmanned Aerial Vehicle (UAV) Photogrammetry for Heritage Building Documentation, 2023.
[30] A. Candelario-Garrido, J. García-Sanz-Calcedo, A.M.R. Rodríguez, Society, A quantitative analysis on the feasibility of 4D planning graphic systems versus conventional systems in building projects, Sustain. Cities Soc. 35 (2017) 378–384.
[31] D. Chang, A. Božič, T. Zhang, Q. Yan, Y. Chen, S. Süsstrunk, et al., RC-MVSNet: unsupervised multi-view stereo with neural rendering, in: European Conference on Computer Vision, Springer, 2022, pp. 665–680.
[32] J. Liu, J. Gao, S. Ji, C. Zeng, S. Zhang, J. Gong, et al., Deep learning based multi-view stereo matching and 3D scene reconstruction from oblique aerial images, J. Photogrammetry Remote Sensing 204 (2023) 42–60.
[33] X. Wang, E. Dong, J. Tong, Z. Sun, W. Li, F. Duan, Recurrent multi-view stereo depth inference with pyramid of images, in: 2022 IEEE International Conference on Mechatronics and Automation (ICMA), IEEE, 2022, pp. 259–263.
[34] J. Li, Z. Bai, W. Cheng, H. Liu, Feature pyramid multi-view stereo network based on self-attention mechanism, in: Proceedings of the 2022 5th International Conference on Image and Graphics Processing, 2022, pp. 226–233.
[35] X. Gu, Z. Fan, S. Zhu, Z. Dai, F. Tan, P. Tan, Cascade cost volume for high-resolution multi-view stereo and stereo matching, Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (2020) 2495–2504.
[36] Y. Barwise, P. Kumar, A. Tiwari, F. Rafi-Butt, A. McNabola, S. Cole, et al., The co-development of HedgeDATE, a public engagement and decision support tool for air pollution exposure mitigation by green infrastructure 75 (2021) 103299.
[37] T.-Y. Lin, P. Goyal, R. Girshick, K. He, P. Dollár, Focal loss for dense object detection, Proceedings of the IEEE international conference on computer vision (2017) 2980–2988.
[38] X. Li, W. Wang, L. Wu, S. Chen, X. Hu, J. Li, et al., Generalized Focal Loss: Learning Qualified and Distributed Bounding Boxes for Dense Object Detection, vol. 33, 2020, pp. 21002–21012.
[39] H. Aanæs, R.R. Jensen, G. Vogiatzis, E. Tola, ABJIJoCV. Dahl, Large-scale Data for Multiple-View Stereopsis, vol. 120, 2016, pp. 153–168.
[40] S. Galliani, K. Lasinger, K. Schindler, Massively parallel multiview stereopsis by surface normal diffusion, Proceedings of the IEEE International Conference on Computer Vision (2015) 873–881.
[41] J.L. Schonberger, J.-M. Frahm, Structure-from-motion revisited, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 4104–4113.
[42] Y. Ding, W. Yuan, Q. Zhu, H. Zhang, X. Liu, Y. Wang, et al., Transmvsnet: global context-aware multi-view stereo network with transformers, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022, pp. 8585–8594.
[43] G. Xu, X. Wang, X. Ding, X. Yang, Iterative geometry encoding volume for stereo matching, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023, p. 21919, 28.
[44] Z. Zhang, H. Gao, Y. Hu, R. Wang, N2MVSNet: non-local neighbors aware multi-view stereo network, in: ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, 2023, pp. 1–5.
[45] Q. Yan, Q. Wang, K. Zhao, B. Li, X. Chu, F. Deng, Rethinking disparity: a depth range free multi-view stereo based on disparity, in: Proceedings of the AAAI Conference on Artificial Intelligence, 2023, pp. 3091–3099.
[46] X. Ma, Y. Gong, Q. Wang, J. Huang, L. Chen, F. Yu, Epp-mvsnet: epipolar-assembling based depth prediction for multi-view stereo, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021, pp. 5732–5740.
[47] H. Huang, Fujian's Tulou: A Treasure of Chinese Traditional Civilian Residence, Springer Nature, 2019.
[48] Q. Li, X. Sun, C. Chen, X.J.E. Yang, Buildings, Characterizing the household energy consumption in heritage Nanjing Tulou buildings, China, A comparative field survey study 49 (2012) 317–326.
[49] M.J.S. Ueda, A preliminary environmental assessment for the preservation and restoration of Fujian Hakka Tulou complexes 4 (2012) 2803–2817.
[50] L. Hua, C. Chen, H. Fang, X. Wang, 3D documentation on Chinese Hakka Tulou and Internet-based virtual experience for cultural tourism: a case study of Yongding County, Fujian, J. Cult. Heritage 29 (2018) 173–179.
[51] DJI. Dji.
[52] J.L. Giese, J. Cote, Defining consumer satisfaction 1 (2000) 1–22.
[53] L. Vico, Authenticity and Realism: Virtual vs Physical Restoration, McDonald Institute, 2018.
[54] V.M. de Almeida, S. Wefers, O.J. Murphy, P.C. Heritage, An Interdisciplinary Discussion of the Terminologies Used in Cultural Heritage Research, vol. 1, 2017.
[55] S. Hermon, F. Niccolucci, Digital Authenticity and the London Charter, McDonald Institute, 2018.
[56] F. Baltar, IJIr Brunet, Social Research 2.0: Virtual Snowball Sampling Method Using Facebook, vol. 22, 2012, pp. 57–74.
[57] P. Biernacki, DJSm Waldorf, research, Snowball Sampling: Problems and Techniques of Chain Referral Sampling, vol. 10, 1981, pp. 141–163.
[58] J.C. De Winter, Dodou Djpa, evaluation research, Five-point Likert items: t test versus, Mann-Whitney-Wilcoxon 15 (2010) 1–12.
[59] S. Jansen, The Impact of the Have–Want Discrepancy on Residential Satisfaction, vol. 40, 2014, pp. 26–38.
[60] H. Eißfeldt, M.J.T.R.P. Biella, The Public Acceptance of Drones–Challenges for Advanced Aerial Mobility (AAM), vol. 66, 2022, pp. 80–88.
[61] ICOMOS, The ICOMOS Charter for the Conservation of Historic Towns and Urban Areas, 1987.
[62] C. Tweed, M.J.L. Sutherland, u planning, Built cultural heritage and sustainable urban development 83 (2007) 62–69.
[63] E.H. Yung, E.H.J.H.I. Chan, Problem issues of public participation in built-heritage conservation: two controversial cases in Hong Kong 35 (2011) 457–466.
[64] J. Abbott, Sharing the City: Community Participation in Urban Management, routledge, 2013.
[65] S. Lidelöw, T. Örn, A. Luciani, A. Rizzo, society, Energy-efficiency Measures for Heritage Buildings: A Literature Review, vol. 45, 2019, pp. 231–242.
[66] L. Zhao, J. Zhang, H. Jing, J. Wu, Y. Huang, A Blockchain-Based Cryptographic Interaction Method of Digital Museum Collections, vol. 59, 2023, pp. 69–82.