drones

MDPI

Article

Closer Nap-of-the-Object Photogrammetry with Geographic Neural Radiance Fields

Haoyu Liu, Yizhi Zou, Lu Yang *, Huifu Chen, Lei Xia and Lubo Li

School of Automation Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China; 202452060424@std.uestc.edu.cn (H.L.)
* Correspondence: yanglu@uestc.edu.cn

Highlights

What are the main findings?

- A georeferenced NeRF-based UAV acquisition framework was developed to generate metric geometric priors for close-proximity waypoint planning.
- Experimental results achieved an average flight proximity of 5.2 cm on the Statue target and a PSNR improvement of up to 4.87 dB over the evaluated photogrammetry baseline.

What are the implications of the main findings?

- The framework supports close-range image acquisition for static cultural heritage and infrastructure targets where distant UAV imagery loses fine detail.
- The geometry-aware planning strategy can reduce redundant viewpoints and improve reconstruction quality under supervised close-proximity operation.

Abstract

High-precision 3D reconstruction of objects with complex surfaces, such as ancient architecture and detailed artworks, requires close-range image acquisition, which remains challenging for Unmanned Aerial Vehicle (UAV) systems. The operational proximity of current UAV workflows is often insufficient to capture fine geometric and textural details, limiting high-fidelity digitization. This paper presents a georeferenced NeRF-based UAV acquisition framework for automated waypoint planning and supervised close-proximity execution. The core of the framework is a path-planning module that operates on a metric geometric prior established through Geographic Neural Radiance Fields (Geo-NeRF), which denotes a georeferenced NeRF modeling pipeline rather than a new NeRF architecture or loss function. By generating waypoints directly on this neural representation and optimizing the flight path via a nearest-neighbor strategy, the proposed framework supports close-proximity image acquisition for static targets under controlled conditions. Empirical validation demonstrates improved close-range flight proximity, photographic accuracy, and 3D reconstruction fidelity compared with the evaluated baselines.

Keywords: Unmanned Aerial Vehicle (UAV); 3D reconstruction; Neural Radiance Fields (NeRF); path planning; Nap-of-the-Object

Check for updates

Academic Editor: Giordano Teza

Received: 11 May 2026
Revised: 19 June 2026
Accepted: 7 July 2026
Published: 9 July 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Three-dimensional (3D) reconstruction captures the shape and appearance of real-world objects and has become a fundamental technology in many domains [1]. In cultural heritage, it supports the high-fidelity digital preservation of historical sites and artifacts for

future access [2–7]. In industrial applications, it supports digital twin creation, simulation, quality control, and predictive maintenance for complex machinery and infrastructure [8,9]. Its use also extends to urban planning, geological surveying [10], and visual effects. These applications have motivated continued research on accurate and efficient data acquisition methodologies [11,12].

Among available sensing platforms, Unmanned Aerial Vehicles (UAVs) provide a flexible tool for 3D reconstruction and UAV photogrammetry [13–16]. Their mobility enables access to viewpoints that are difficult for ground-based sensors, which is valuable for bridges, power lines, historical monuments, and other large structures.

Despite the maturity of UAV-based 3D reconstruction, obtaining fine details from complex surfaces remains difficult. Reconstruction fidelity depends strongly on source-image resolution and camera-to-object distance [17]. Weathered sculptures, facade ornaments, and other fine structures require close-range imagery to reveal details that are lost in distant, lower-resolution views [18,19].

Close-proximity UAV acquisition, however, introduces safety and planning challenges. Small navigation errors, geometric-model errors, wind disturbance, or camera/gimbal offsets may become significant when the planned standoff distance is only a few centimeters. These constraints make it necessary to distinguish between automated waypoint planning and a general guarantee of collision-free autonomous flight.

Several path-planning strategies have been studied for UAV reconstruction. Regular paths, such as lawnmower grids or circular trajectories, are simple to deploy but do not adapt to local surface geometry [20]. They may either remain too far from the target for safety or miss important non-convex surface regions.

Model-based planning strategies address this issue by first reconstructing a coarse point cloud or mesh and then generating close-range viewpoints on this explicit model [21–25]. Traditional Structure from Motion (SfM) and Multi-View Stereo (MVS) pipelines, including COLMAP [26], are commonly used for this initial reconstruction. Their planning quality is limited by the completeness and stability of the coarse geometry. Sparse or distant images may produce noisy surfaces, holes, and inaccurate normals, which in turn force a larger safety margin and reduce close-range image quality [27,28].

Neural Radiance Fields (NeRF) provide a continuous volumetric scene representation with strong novel-view synthesis performance [29–31]. A NeRF representation can preserve details that are difficult to recover from sparse explicit reconstructions. However, standard NeRF training usually operates in an arbitrary local coordinate frame and does not directly provide a metric, UAV-executable geometry prior for close-range waypoint generation.

Active reconstruction methods and next-best-view approaches select informative viewpoints to improve reconstruction quality, often using uncertainty reduction or expected information gain [32,33]. These methods are complementary to this study. Their primary objective is view selection for improving a representation, whereas the objective here is to use a georeferenced NeRF-based representation as a metric geometric prior for surface-following UAV waypoint planning.

To address this problem, this paper presents a two-stage UAV acquisition framework for high-fidelity reconstruction of static targets with stable geometry. The framework first constructs a georeferenced NeRF-based model and then uses the extracted geometry to generate close-proximity waypoints for supervised autonomous execution. The main contributions are summarized as follows:

- A georeferenced NeRF-based modeling pipeline, referred to as Geographic Neural Radiance Fields (Geo-NeRF), is introduced for producing a metric geometric prior for UAV waypoint planning. The term Geo-NeRF denotes the georeferenced modeling workflow rather than a new NeRF network architecture or loss function.

- A surface-following waypoint generation strategy is developed by extracting a NeRF-derived geometric representation, estimating local surface normals, and placing close-proximity candidate viewpoints in a real-world coordinate frame.
- Real-world experiments on a detailed statue and a building facade evaluate the framework in terms of flight proximity, reconstruction quality (PSNR, SSIM, LPIPS), and geometric fidelity under controlled static-target conditions.

2. Materials and Methods

This work presents a georeferenced UAV acquisition framework designed to support high-fidelity 3D reconstruction of complex static targets. The core of the approach is a model-driven, two-stage data acquisition strategy that leverages an initial neural representation of the target to plan a subsequent close-proximity flight. This section provides an overview of the system architecture, key processing steps, and operational workflow.

2.1. System Overview

The goal of the system is to automate the planning of UAV image acquisition from an initial coarse scan to a close-range fine scan while keeping the execution under operator supervision. The UAV first performs a conservative preliminary scan to obtain a coarse geometric prior of the target. The resulting model is then used to plan close-proximity viewpoints for a second acquisition stage. The workflow is managed by a custom software application deployed on a DJI mobile controller, which communicates with the UAV and provides the operator with mission monitoring and emergency takeover capability.

The operational workflow of the proposed system, as illustrated in Figure 1, is divided into five numbered stages:

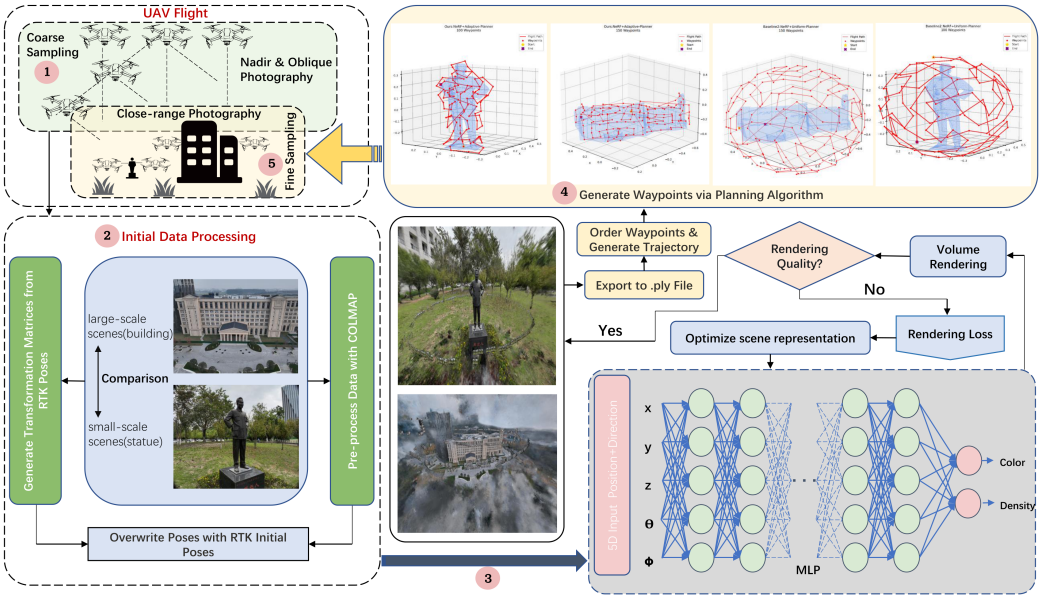


Figure 1. Revised overview of the proposed UAV acquisition and reconstruction workflow. The numbered stages indicate: (1) coarse image acquisition, (2) COLMAP pose estimation and RTK-based pose refinement, (3) initial georeferenced NeRF training, (4) geometry-aware waypoint generation and trajectory ordering, and (5) fine image acquisition for final NeRF reconstruction. The workflow consists of two main stages: data acquisition and data processing. In the data acquisition stage, the UAV first performs a Coarse Sampling flight to capture an initial image set and then performs a Fine Sampling flight guided by the planned waypoints. The data processing pipeline refines the camera poses, trains the initial model, extracts geometry for waypoint generation, and uses the fine images to train the final high-fidelity NeRF model.

2.1.1. Initial Data Acquisition and Coarse Modeling

The process begins with an initial conservative flight path that captures a broad view of the static target. Images from this stage are processed by COLMAP to estimate camera poses and generate the `transforms.json` file used by the NeRF training pipeline. The image records are then associated with RTK/GNSS measurements and converted into a local metric frame. In this paper, Geographic Neural Radiance Fields (Geo-NeRF) refers to this georeferenced NeRF modeling pipeline rather than to a modification of the NeRF network architecture or loss function. The resulting initial model serves as a metric geometric prior for waypoint generation.

2.1.2. Waypoint Generation on the Neural Representation

With the initial Geo-NeRF model established, the system transitions to waypoint generation. Rather than relying only on a coarse sparse reconstruction, the proposed algorithm extracts a geometric representation from the trained NeRF and generates candidate viewpoints from local surface geometry. The planned distance is the camera-center standoff to the reconstructed surface, not the obstacle-avoidance braking distance of the UAV platform. The specifics of this generation process are detailed in Section 2.3.

2.1.3. Close-Proximity Trajectory Planning

The unordered set of waypoints generated in the previous phase is organized into an executable visiting sequence. A nearest-neighbor heuristic is employed to connect the waypoints and reduce unnecessary travel distance. This stage determines the waypoint order; it does not replace a full online collision-avoidance or local re-planning module. Safety during the reported flights was handled through conservative coarse acquisition, preflight inspection of planned waypoints, stop-and-go execution, and operator supervision.

2.1.4. Supervised Waypoint Execution and Final Reconstruction

The finalized waypoint sequence is uploaded to the UAV through the custom application. During fine acquisition, the UAV moves between waypoints, stops, stabilizes, orients the camera, and captures one image at each waypoint. The experiments were conducted as supervised autonomous missions: the waypoint execution was automated, while a human operator monitored the flight and could stop or take over the mission if necessary. The fine images are then used to train the final high-fidelity NeRF, yielding a reconstruction with enhanced detail compared with the initial coarse-flight model.

2.2. Geo-NeRF for Initial Modeling

The generation of a reliable, georeferenced initial 3D model is the foundation of the proposed path planning strategy. This model must be sufficiently accurate to serve as a geometric reference for generating close-proximity waypoints in a real-world coordinate system. In this work, Geographic Neural Radiance Fields (Geo-NeRF) denotes a georeferenced NeRF processing pipeline. It uses the standard NeRF/NeRFStudio training framework and modifies the camera-pose preparation and exported geometry alignment, rather than introducing a new MLP architecture or a new loss function. This section first reviews the principles of standard NeRF and then details the georeferencing pipeline used for the initial model.

2.2.1. Principle of Neural Radiance Fields

A standard Neural Radiance Field represents a continuous 5D volumetric scene function, which is approximated by a Multi-Layer Perceptron (MLP), denoted as F_Θ. As shown

in Equation (1), this network maps a 3D spatial coordinate $\mathbf{x} = (x, y, z)$ and a 2D viewing direction $\mathbf{d} = (\theta, \phi)$ to an emitted color $\mathbf{c} = (r, g, b)$ and a volume density σ:

$$F_\Theta : (\mathbf{x}, \mathbf{d}) \rightarrow (\mathbf{c}, \sigma) \tag{1}$$

The expected color $C(\mathbf{r})$ of a camera ray $\mathbf{r}(t) = \mathbf{o} + t\mathbf{d}$ is computed by integrating the color and density along the ray from a near bound t_n to a far bound t_f, as described in Equation (2):

$$C(\mathbf{r}) = \int_{t_n}^{t_f} T(t)\sigma(\mathbf{r}(t))\mathbf{c}(\mathbf{r}(t), \mathbf{d})\, dt \tag{2}$$

where the transmittance $T(t) = \exp(-\int_{t_n}^{t} \sigma(\mathbf{r}(s))\, ds)$. The network is trained by minimizing the photometric error between the rendered pixel colors and the ground-truth pixel colors. However, this entire formulation operates within an arbitrary local frame with unknown scale, typically established by SfM algorithms. This makes it unsuitable for real-world robotic tasks that require georeferenced models with metric scale.

2.2.2. Geo-NeRF: Integrating RTK/GNSS and COLMAP Poses

The Geo-NeRF pipeline grounds the initial NeRF model in a local metric frame by combining COLMAP poses with RTK/GNSS records. COLMAP first estimates the relative camera poses from the coarse image set and produces a NeRFStudio-compatible `transforms.json` file. The RTK/GNSS records provide image-associated positions in the WGS84 datum. These positions are converted into a local East-North-Up (ENU) coordinate frame whose origin is defined by the first accepted camera record. The conversion is summarized in Equation (3):

$$\mathbf{p}_i^{\mathrm{ENU}} = \Pi_{\mathrm{ENU}}(\phi_i, \lambda_i, h_i; \phi_0, \lambda_0, h_0), \tag{3}$$

where (ϕ_i, λ_i, h_i) is the WGS84 latitude, longitude, and altitude of image i, (ϕ_0, λ_0, h_0) defines the local origin, and $\Pi_{\mathrm{ENU}}(\cdot)$ denotes the WGS84-to-ENU projection.

The COLMAP reconstruction and the ENU trajectory are then brought into the same metric coordinate frame by the similarity transformation in Equation (4). Let $\mathbf{x}^{\mathrm{C}}$ denote a point or camera center in the COLMAP coordinate frame. The aligned coordinate $\mathbf{x}^{\mathrm{ENU}}$ is computed as

$$\mathbf{x}^{\mathrm{ENU}} = s\mathbf{R}_{\mathrm{a}}\mathbf{x}^{\mathrm{C}} + \mathbf{t}_{\mathrm{a}}, \tag{4}$$

where s is the scale factor, $\mathbf{R}_{\mathrm{a}}$ is the frame-alignment rotation, and $\mathbf{t}_{\mathrm{a}}$ is the translation. In implementation, the aligned camera transforms are written back to the `transforms.json` file used by NeRFStudio. The resulting camera-to-world matrix for image i is expressed as Equation (5):

$$\mathbf{P}_i = \begin{bmatrix} \mathbf{R}_i^{\mathrm{ENU}} & \mathbf{p}_i^{\mathrm{ENU}} \\ \mathbf{0}^{\mathrm{T}} & 1 \end{bmatrix} \tag{5}$$

where $\mathbf{R}_i^{\mathrm{ENU}}$ is the camera rotation after coordinate-frame alignment and $\mathbf{p}_i^{\mathrm{ENU}}$ is the RTK/GNSS-derived camera position in the local metric frame. After NeRF training, the exported point cloud or mesh is transformed with the same scale and coordinate alignment before it is used for waypoint generation. This procedure provides a metric geometric prior for the downstream planner while preserving the standard NeRF rendering and optimization formulation. Figure 2 summarizes the coordinate-processing pipeline from RTK/GNSS records and COLMAP poses to metric geometry for waypoint generation.

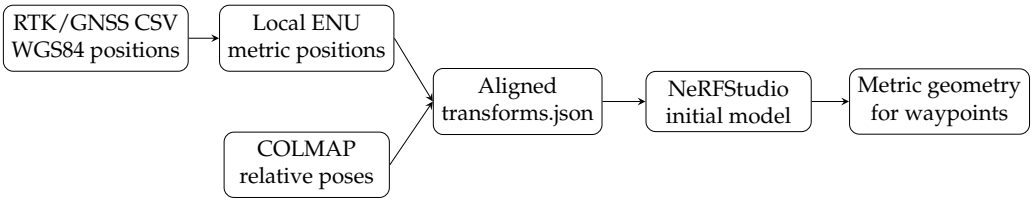


Figure 2. Schematic of the coordinate-processing pipeline used in Geo-NeRF. COLMAP provides the relative camera poses, while RTK/GNSS records provide metric WGS84 positions that are converted to a local ENU frame and used to align the NeRFStudio camera transforms.

2.3. Waypoint Generation on Neural Representation

Upon completion of the initial modeling phase, the system possesses a georeferenced point cloud derived from Geo-NeRF, serving as the geometric reference of the target. The subsequent step uses this representation to generate waypoints for the close-proximity data acquisition flight. The objective of this phase is twofold: first, to cover the visible surface of the target, and second, to reduce redundant waypoints before trajectory ordering. The waypoint-generation process consists of geometry cleaning, surface normal computation, initial waypoint seeding, and spatial density reduction.

2.3.1. Geometry Cleaning and Surface Reconstruction

Before waypoint generation, the exported geometry is preprocessed to reduce NeRF floaters and isolated artifacts. The target region is first cropped according to the coarse-flight reconstruction and the known inspection area. Small disconnected components and sparse outlier points are then removed before meshing. The cleaned point cloud is converted into a triangular mesh using a standard surface reconstruction algorithm. Finally, the mesh and the candidate waypoint set are visually inspected before flight execution, and candidate waypoints with abnormal normals or locations outside the target region are discarded during mission preparation. This preprocessing was used to prevent obvious background clutter or floating artifacts from being converted into close-range waypoints.

2.3.2. Surface Normal Estimation

To keep the camera viewing direction approximately perpendicular to the local surface at each capture point, the surface orientation is estimated from the reconstructed mesh. This mesh, composed of vertices and faces, provides the necessary topological information.

For each triangular face f in the mesh, defined by vertices $(\mathbf{v}_0, \mathbf{v}_1, \mathbf{v}_2)$, the face normal vector $\mathbf{n}_f$ is computed using the cross-product of two edge vectors, as shown in Equation (6):

$$\mathbf{n}_f = (\mathbf{v}_1 - \mathbf{v}_0) \times (\mathbf{v}_2 - \mathbf{v}_0) \tag{6}$$

The resulting vector $\mathbf{n}_f$ is normalized to a unit vector $\hat{\mathbf{n}}_f$. This vector represents the outward-facing direction perpendicular to the local surface patch, providing the precise orientation for subsequent waypoint placement.

2.3.3. Initial Waypoint Seeding

Following the computation of normal vectors, a dense set of candidate waypoints is generated. For each triangular face f, one waypoint $\mathbf{w}_f$ is generated by translating the face centroid $\mathbf{c}_f$ along its unit normal vector $\hat{\mathbf{n}}_f$ by a predefined standoff distance d_{standoff}. Here, d_{standoff} denotes the nominal planned distance between the camera optical center and the reconstructed surface. It is not a manufacturer obstacle-avoidance braking distance or a guaranteed clearance margin for the UAV body. In the reported experiments, this nominal camera-center standoff was set to 5 cm for the fine acquisition stage.

The centroid $\mathbf{c}_f$ is calculated as the arithmetic mean of the face vertices in Equation (7):

$$\mathbf{c}_f = \frac{\mathbf{v}_0 + \mathbf{v}_1 + \mathbf{v}_2}{3} \tag{7}$$

The corresponding waypoint is then defined by Equation (8):

$$\mathbf{w}_f = \mathbf{c}_f + \hat{\mathbf{n}}_f \, d_{\text{standoff}} \tag{8}$$

This method yields an initial set of waypoints equal in number to the selected mesh faces. While it provides dense coverage of the reconstructed surface, it also introduces redundancy, particularly in regions of high mesh density. Executing a flight path that visits every generated waypoint would be operationally inefficient.

2.3.4. Waypoint Density Reduction via Spatial Clustering

A spatial optimization strategy is introduced to mitigate waypoint redundancy and ensure computational feasibility for path planning. This step aims to downsample the waypoints while preserving the spatial distribution and coverage integrity established in the previous stage.

A spatial clustering approach facilitated by a KD-Tree is used for this reduction. The KD-Tree is a space-partitioning data structure that enables efficient nearest-neighbor searches in 3D space. The algorithm operates as follows: for each waypoint $\mathbf{w}_i$ in the dense set, the KD-Tree is queried to identify its k nearest neighbors (where k is a user-defined parameter, e.g., `group_size` = 10). This group of k neighboring waypoints is then collapsed into a single representative waypoint by computing their geometric centroid. To prevent redundant processing, once a waypoint is assigned to a group, it is marked as processed and excluded from future queries.

This clustering process downsamples the initial dense set. By adjusting the `group_size` parameter, the final density of the waypoints can be controlled, producing a practical trade-off between image coverage and flight length. A smaller group size preserves more viewpoints but increases the path length, whereas a larger group size shortens the path at the risk of undersampling fine surface variations. In the reported experiments, the dense candidate set was reduced approximately from 800 to 100 waypoints for the Statue and from 3500 to 150 waypoints for the Building. The output of this stage is a reduced set of waypoints for the trajectory planning phase discussed in the following section.

2.4. Trajectory Construction

Upon generating the reduced set of waypoints, the final planning step is to determine an efficient visiting sequence. This task is framed as a Traveling Salesperson Problem (TSP), where the waypoints represent the locations to be visited and the travel cost is defined by the Euclidean distance in 3D space.

2.4.1. The Nearest Neighbor Heuristic for TSP

Solving the TSP optimally is an NP-hard problem. For a large number of waypoints, exact algorithms are computationally infeasible, particularly for systems intended for practical, in-field deployment. Therefore, to balance path efficiency with computational cost, the implementation employs a greedy heuristic: the Nearest Neighbor (NN) algorithm. The NN algorithm first computes a pairwise distance matrix for all reduced waypoints. Starting from the waypoint closest to the UAV's current position or a designated home

point, the algorithm repeatedly appends the nearest unvisited waypoint until all waypoints have been visited. The output is an ordered sequence of indices defined in Equation (9):

$$P = \{p_1, p_2, \ldots, p_n\}, \tag{9}$$

where n is the total number of waypoints. This sequence defines the high-level waypoint order.

2.4.2. Rationale and Practical Implementation

The primary rationale for selecting the NN heuristic is computational efficiency. With a time complexity of approximately $O(n^2)$, it generates a solution with low latency for waypoint sets containing hundreds of points. The NN heuristic does not guarantee a globally optimal route, and exact TSP solvers such as Concorde or post-processing strategies such as 2-opt can further reduce path length. In this study, NN was used because the objective was to obtain a practical visiting order for experimental acquisition rather than to solve the global routing problem optimally. This choice is discussed further in Section 5.

In practical implementation, the ordered waypoint sequence serves as a command list for the UAV flight controller through the custom application. The fine-acquisition mission follows a stop-and-go protocol. At each waypoint $\mathbf{w}_{p_i}$, the UAV navigates to the planned position, stops and stabilizes, orients the camera/gimbal toward the corresponding surface point, and captures one high-resolution image. The experiments were conducted with obstacle avoidance disabled because the planned camera-center standoff was below the configurable braking distances of the platform. Safety was therefore handled through pre-flight inspection of the generated waypoints, low-speed waypoint execution, visual supervision, and the operator's ability to stop or manually take over the flight. With the generation of this executable waypoint sequence, the planning phase is concluded and the system is prepared for supervised close-proximity data acquisition.

3. Experimental Setup

The efficacy and performance of the proposed UAV acquisition framework were evaluated on real-world objects with different geometric characteristics. This section describes the hardware platform, target objects, acquisition protocol, baseline methods, and evaluation metrics.

3.1. Implementation Details

3.1.1. Hardware Platform

The experimental setup comprises a commercial-grade UAV equipped with an integrated sensing payload and positioning system, a ground controller for the custom software, and a workstation for offline data processing.

UAV and Sensing Payload: The aerial platform is the DJI Matrice 4E quadcopter (DJI, Shenzhen, China). This platform was selected for two features used in the experiments: (1) its integrated Real-Time Kinematic (RTK) module, and (2) its support for the DJI Mobile SDK, which enables the deployment of the custom waypoint-execution application.

The UAV is equipped with an integrated multi-camera payload. For the initial data acquisition and subsequent reconstruction, the wide-angle camera was used because its broad field of view is suitable for capturing a comprehensive scene representation. The specifications of this primary sensor are as follows:

- Sensor: 4/3 CMOS;
- Effective Pixels: 20 MP;
- Equivalent Focal Length: 24 mm;
- Field of View (FOV): 84°;

- Aperture: f/2.8 to f/11.

While the platform also includes medium and long telephoto lenses, they were not used in the experiments described in this paper to ensure a consistent data acquisition methodology.

Positioning and Georeferencing System: To support the Geo-NeRF pipeline, the UAV's onboard RTK system was used in conjunction with the Qianxun Network RTK service (Qianxun, Shanghai, China). This network provides real-time differential correction data to the UAV through an internet link. The RTK/GNSS records were used to place the COLMAP/NeRFStudio camera transforms in a local metric coordinate frame. Full hardware-level lever-arm calibration and time-synchronization error analysis were not separately measured in this study; this is treated as a limitation in Section 5.

Ground Control and Computational Resources: The custom-developed flight control application was operated on the DJI RC Plus remote controller (DJI, Shenzhen, China), serving as the interface for mission initiation, waypoint execution, monitoring, and safety supervision.

The computationally intensive task of training the Neural Radiance Fields was performed on a dedicated workstation with the following specifications:

- GPU: NVIDIA GeForce RTX 4090 (NVIDIA Corporation, Santa Clara, CA, USA);
- RAM: 96 GB (3×32 GB);
- CPU: Intel Xeon Gold 6154 (Intel Corporation, Santa Clara, CA, USA).

This configuration ensures that the models can be trained efficiently, enabling a practical workflow from data capture to final 3D model generation. The reconstruction pipeline used COLMAP 3.11.1 for SfM/MVS processing and Nerfstudio 1.1.5 for NeRF training.

Acquisition Protocol: The experiments used a two-stage acquisition protocol. During the coarse stage, the UAV captured overview images at an approximate distance of 7 m from the Statue and 20 m from the Building. During the fine stage, the nominal planned camera-center standoff was 5 cm. Each fine waypoint was executed in a stop-and-go manner: the UAV stopped at the waypoint, stabilized, and captured one image. The missions were monitored by an operator using the remote controller.

3.1.2. Target Objects

To evaluate the performance of the system, target objects with distinct characteristics were selected to represent cultural heritage digitization and large-scale inspection scenarios. The primary subject is a life-sized dark stone statue, hereafter referred to as the Statue, as shown in Figure 3.

The Statue stands approximately 2.9 m in height (including the plinth) and is situated in a complex outdoor environment with foliage and background clutter. This object was selected to test the system's capabilities in three key aspects:

- Geometric Complexity: The statue features a high degree of non-convex geometry. Fine-scale details, such as facial features, clothing folds, and plinth inscriptions, are difficult to resolve using traditional standoff photogrammetry. These features provide a benchmark for evaluating the fidelity improvements enabled by close-proximity flight.
- Challenging Photometric Properties: The surface consists of a dark stone-like material. This introduces low-texture and low-light regions in occluded areas, testing the robustness of the Geo-NeRF pipeline when processing high-resolution, close-range imagery.
- Uncontrolled Environment: The outdoor setting introduces real-world variables, including fluctuating ambient lighting and background distractions. This scenario tests the method outside of controlled laboratory conditions.

In addition to the Statue, experiments were conducted on a second object: the facade of a large institutional building (Building), as shown in Figure 4. This target evaluates

the system's scalability on large-scale, predominantly planar structures. The Building is characterized by expansive surfaces and repetitive architectural features, such as windows, pillars, and cornices. These characteristics present challenges distinct from the statue's non-convex geometry, specifically regarding the maintenance of geometric consistency over large areas and the reconstruction of sharp edges.

By using the Statue as the primary test case and the Building as a large-scale facade case, the experiments evaluate both non-convex local geometry and broad planar geometry.



Figure 3. The Statue target object used for experimental evaluation. The visible Chinese inscription on the plinth identifies the commemorated scholar and is not used in the reconstruction analysis.



Figure 4. The Building target object, representing a large-scale planar structure.

3.1.3. Compared Methods

To evaluate the performance of the proposed system, a comparative analysis was conducted against two baseline methods. These baselines were selected to isolate two factors: (1) the use of a georeferenced NeRF-derived geometric prior for planning, and (2) the use of geometry-aware waypoint placement instead of generic viewpoint sampling. For all experiments, the final high-fidelity 3D model was reconstructed using the same standard NeRF training procedure to compare the quality of the acquired image data.

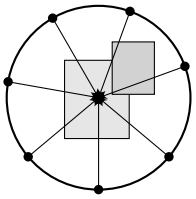
- Baseline 1: Traditional SfM+MVS Guided Planning (SfM+MVS-Planner). This method uses the same initial images as the proposed method. Instead of training Geo-NeRF

for the initial geometric prior, it processes the images with COLMAP and MVS to generate a dense point cloud and triangular mesh. The resulting mesh is then passed to the same waypoint generation and trajectory-ordering modules described in Sections 2.3 and 2.4. This baseline compares the effect of replacing the NeRF-derived prior with a conventional photogrammetric prior.

- Baseline 2: Geo-NeRF with Regular Viewpoint Sampling (Geo-NeRF+Regular-Planner). This baseline uses the same Geo-NeRF-derived mesh as the proposed method, but replaces geometry-aware waypoint seeding with regular viewpoint sampling. A bounding sphere is computed around the mesh and expanded by the nominal standoff distance. Viewpoints are distributed on the sphere and oriented toward the object center; occluded or invalid viewpoints are discarded. The remaining viewpoints are ordered using the same nearest-neighbor heuristic. This baseline represents a common object-centered coverage strategy and is useful for testing whether the proposed surface-normal waypoint placement adds value beyond using Geo-NeRF alone.
- Proposed: Geo-NeRF-Guided Irregular Planning (Geo-NeRF+Irregular-Planner). This is the integrated system detailed in Section 2.1. It uses the Geo-NeRF-derived geometry to generate surface-normal waypoints and then orders them with the nearest-neighbor heuristic.

Figure 5 illustrates the difference between the regular viewpoint baseline and the proposed surface-normal waypoint placement. The figure is a schematic explanation of the sampling strategies, not an additional experimental result.

(**a**) Regular object-centered sampling



(**b**) Surface-normal waypoint placement

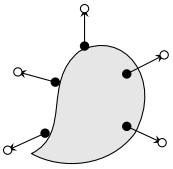


Figure 5. Schematic comparison of viewpoint placement strategies. Regular sampling places cameras on a global bounding surface and points them toward the object center, whereas the proposed method places waypoints by offsetting local surface centroids along estimated normals.

By comparing the proposed system against these two baselines, the experiments separate the effect of the initial geometric prior from the effect of surface-aware waypoint placement. For facade-like structures, a facade-grid planner could be a stronger scene-specific regular baseline than a bounding-sphere planner; such scene-specific regular planning is considered an important direction for broader future comparisons.

3.1.4. Evaluation Metrics

To evaluate the system's performance, standard image quality metrics were used for the 3D reconstruction and a geometric metric was used for the flight trajectory.

The photorealistic quality of the reconstruction is assessed by rendering novel views from the trained NeRF models and comparing them against independently captured validation images. These validation images were not used for training any of the compared methods, and the same validation camera poses were used when evaluating all methods. The following metrics are used:

- Peak Signal-to-Noise Ratio (PSNR): Measures the pixel-level reconstruction accuracy. Higher values indicate lower error.

- Structural Similarity Index Measure (SSIM): Evaluates structural similarity, aligning more closely with human visual perception than PSNR. Higher values indicate better structural preservation.
- Learned Perceptual Image Patch Similarity (LPIPS): Computes the distance between deep feature representations (using VGG) of the rendered and ground-truth images. Lower values indicate higher perceptual similarity.

In addition to reconstruction quality, the effectiveness of the path planning algorithm is quantified using a trajectory-based metric:

- Average Flight Proximity (AFP): This metric quantifies the average realized distance between the executed trajectory and the target surface. It is calculated as the mean Euclidean distance from the recorded trajectory points to the target object's surface, which is represented by the highest-fidelity reconstructed mesh. A lower AFP value indicates closer image acquisition and therefore higher potential ground sampling resolution.

4. Results

4.1. Quantitative Evaluation

The quantitative results, summarized in Table 1, demonstrate the improved performance of the proposed method across all metrics for both the geometrically complex Statue target and the large-scale planar Building target.

Impact of Geo-NeRF Geometric Prior. A direct comparison between the proposed method and Baseline 1 (SfM+MVS-Planner) highlights the advantages of using Geo-NeRF as the initial model for path planning. The proposed system achieves a PSNR improvement of 4.19 dB and 4.87 dB for the Statue and Building targets, respectively, over the traditional photogrammetry pipeline. This performance gap is attributed to the limitations of SfM/MVS in generating complete and clean models from sparse initial imagery. Inaccuracies and holes in the COLMAP-generated mesh can lead to waypoint placement errors and less consistent close-range image acquisition.

Impact of Geometry-Aware Planning Strategy. The comparison with Baseline 2 (NeRF+Regular-Planner) evaluates the effect of surface-normal waypoint placement. Although both methods utilize a NeRF-based initial model, the irregular planner yields a notable performance gain, improving PSNR by 2.55 dB for the Statue and 2.44 dB for the Building. Perceptual error (LPIPS) is also reduced by 21.7% and 20.4%, respectively. This indicates that the sampling strategy is a critical factor. The regular viewpoint sampling of Baseline 2 provides broad coverage, but it is less adapted to local geometric complexity, such as facial features or architectural cornices. In contrast, the proposed method generates waypoints directly from local surface geometry.

Flight Proximity Analysis. The flight performance metrics provide a direct explanation for the observed reconstruction differences. The proposed system achieved an AFP of 5.2 cm for the Statue and 5.5 cm for the Building, which is close to the nominal 5 cm camera-center standoff. All methods used the same nominal standoff setting, but AFP was measured against the final reference surface. Therefore, if a baseline plans waypoints on a displaced or incomplete planning mesh, its realized distance to the reference surface can be larger than the nominal value. Baseline 1 produced AFP values of 9.8 cm and 12.1 cm because its SfM/MVS planning geometry contained more missing or noisy regions. Baseline 2 produced larger average standoffs of 18.5 cm and 25.3 cm because its viewpoints were placed on a global bounding sphere rather than offset from local surface patches. These larger acquisition distances reduce the effective image resolution on the target surface.

In summary, the quantitative results show that the Geo-NeRF-derived prior improves planning over the SfM/MVS prior in the tested cases, and that geometry-aware waypoint placement improves reconstruction quality compared with regular object-centered sampling.

Table 1. Quantitative comparison on both the Statue and Building targets. The proposed method achieves the best results among the evaluated methods. Best results are highlighted in bold. Upward arrows indicate that higher values are better, and downward arrows indicate that lower values are better.

Method	Statue Target				Building Target			
	PSNR (↑)	SSIM (↑)	LPIPS (↓)	AFP (cm) (↓)	PSNR (↑)	SSIM (↑)	LPIPS (↓)	AFP (cm) (↓)
SfM+MVS (Baseline 1)	22.54	0.851	0.234	9.8	24.12	0.899	0.187	12.1
Geo-NeRF+Regular (B. 2)	24.18	0.883	0.198	18.5	26.55	0.921	0.142	25.3
Proposed Geo-NeRF+Irregular	**26.73**	**0.912**	**0.155**	**5.2**	**28.99**	**0.945**	**0.113**	**5.5**

4.2. Qualitative Comparison

The visual results, presented in Figure 6, support the quantitative findings and provide insight into the performance of each method on the challenging Statue target.

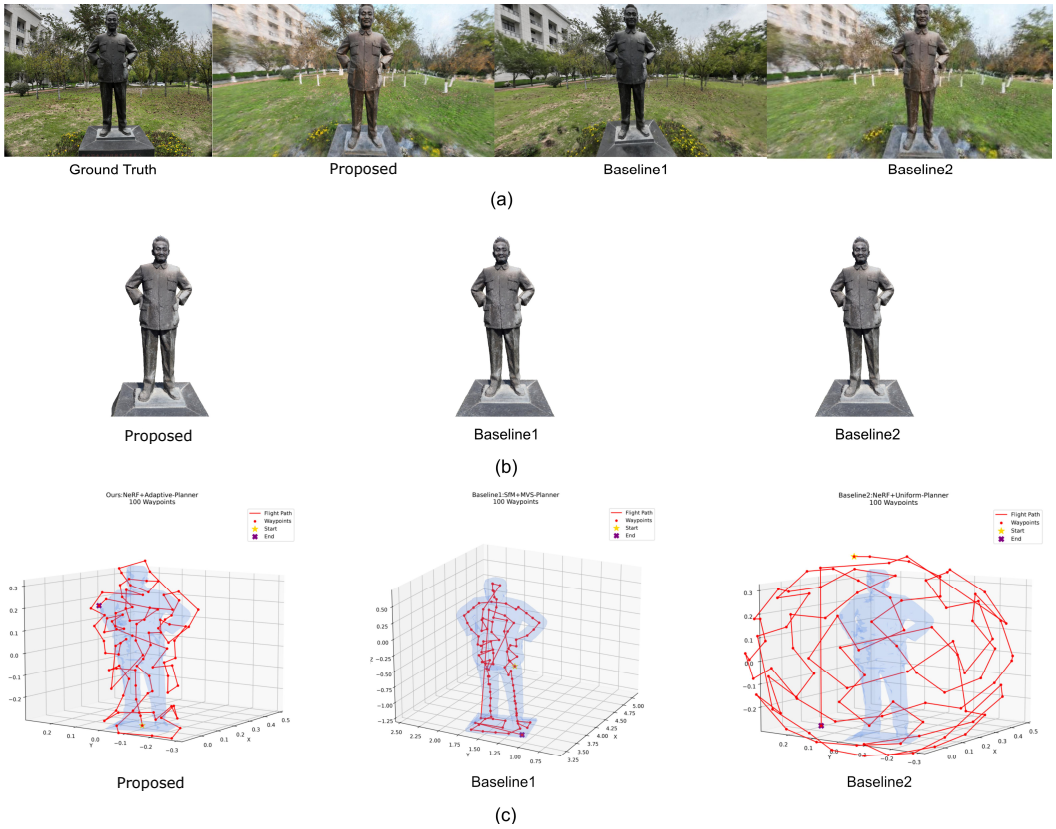


Figure 6. Qualitative comparison on the Statue dataset. (**a**) Novel view synthesis results compared against ground truth. (**b**) Close-up views of the reconstructed geometry. The proposed method (**left**) preserves more fine details than the baselines. (**c**) Visualized flight trajectories. The proposed trajectory (**left**) more closely follows the surface geometry.

Novel View Synthesis Quality (Figure 6a). The top row compares the synthesized views against the ground truth. The proposed method produces a sharper rendering that closely matches the reference photograph. In contrast, Baseline 1 exhibits blurring and ghosting artifacts, particularly around the statue's silhouette, indicating misalignment in the underlying representation. Baseline 2 improves over Baseline 1 but captures fewer subtle surface textures and lighting details than the proposed method.

Geometric Fidelity (Figure 6b). An examination of the extracted geometry reveals the source of these visual differences. The model generated by the proposed method preserves fine geometric details, such as the buttons on the jacket, rendering a textured surface consistent with the real object. The geometry from Baseline 1 is noisy and overly smoothed.

Baseline 2 produces a cleaner mesh than Baseline 1 but lacks some high-frequency surface details captured by the irregular approach.

Flight Trajectory Analysis (Figure 6c). The visualized flight paths explain the performance gaps. The proposed trajectory (left) conforms more closely to the intricate geometry of the statue. The path from Baseline 1 (middle) is visibly more conservative and distant due to the inaccuracies of its initial model. The path from Baseline 2 (right) forms a generic semi-spherical shell and is less adapted to concave regions, such as the space between the arms and torso.

To facilitate inspection of local geometric details, Figure 7 provides enlarged close-up crops from Figure 6b.



Figure 7. Enlarged close-up crops from Figure 6b, focusing on the face and upper-body details of the Statue. The crops are generated from the same qualitative comparison figure and are included only to improve visual inspection of local geometric details.

Analysis on the Building Target. The large-scale Building target provides a second test case with predominantly planar geometry (Figure 8).

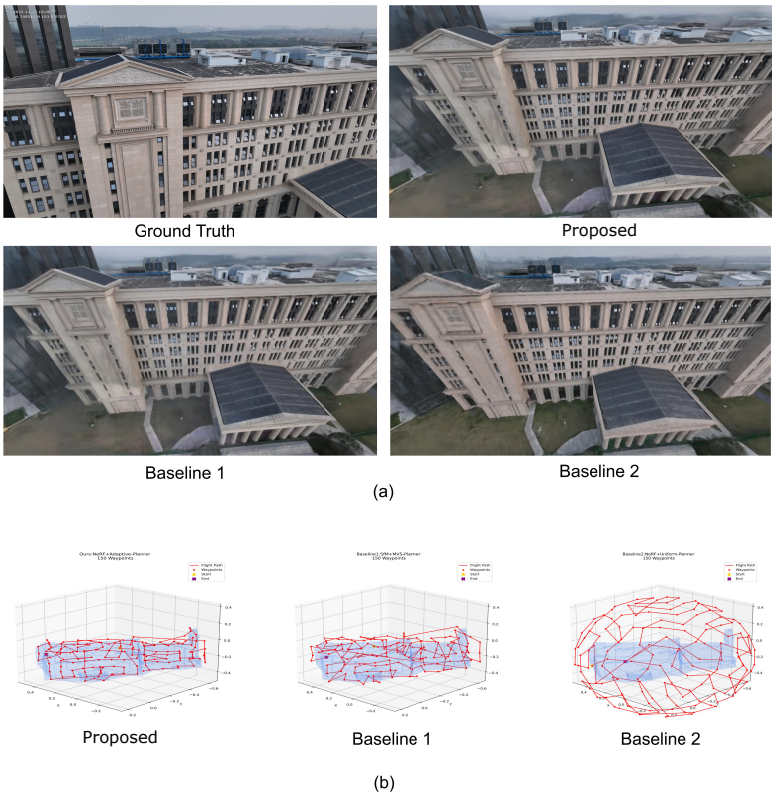


Figure 8. Qualitative comparison on the Building dataset. (**a**) Novel view synthesis results showing architectural details. (**b**) Visualized flight trajectories. The proposed method (**left**) adapts to the planar facade structure.

Novel View Synthesis Quality (Figure 8a). The proposed method generates novel views with high sharpness. Architectural elements, such as window frames and decorative cornices, are rendered without the distortion present in Baseline 1. The rendering from Baseline 2 appears smoother and misses some fine stone textures preserved by the proposed method.

Flight Trajectory Analysis (Figure 8b). The irregular planner (left) generates a path that follows the building's planar surfaces and protrusions. Baseline 1 (middle) exhibits an erratic path with inconsistent standoff distances. Baseline 2 (right) generates a spherical pattern that is suboptimal for the planar structure, resulting in redundant travel through empty space and weaker coverage of corners.

In summary, the qualitative evidence supports the conclusion that geometry-aware close-proximity acquisition contributes to improved visual fidelity and geometric detail in the tested scenarios.

4.3. Ablation Study

To evaluate the contribution of each key component within the proposed system, a series of ablation studies was conducted by systematically simplifying the full pipeline and analyzing the resulting performance impact. The results, summarized in Table 2, show the role of each module in reconstruction quality and operational efficiency.

Table 2. Ablation Study on both the Statue and Building targets. The results validate the necessity of each system component for achieving high reconstruction quality and operational efficiency. Upward arrows indicate that higher values are better, and downward arrows indicate that lower values are better. N/A indicates that the trajectory length exceeded the executable range and therefore was not actually executed.

Configuration	Statue Target		Building Target	
	PSNR (↑)	Traj. Length (m) (↓)	PSNR (↑)	Traj. Length (m) (↓)
Proposed (Full System)	**26.73**	**18.6**	**28.99**	**900.5**
Proposed w/o Waypoint Reduction	26.61	>500 (N/A)	28.91	>2000 (N/A)
Proposed w/o TSP Solver	**26.73**	28.3	**28.99**	1454
Proposed w/o Geo-NeRF (using SfM)	22.54	25.7	24.12	1305.6

On the Necessity of Waypoint Reduction. The "Proposed w/o Waypoint Reduction" configuration plans a path through all initially seeded waypoints. While this provides dense theoretical coverage and only slightly changes PSNR, the resulting trajectory length becomes operationally infeasible. As shown in Table 2, the path length for the Statue grew to an impractical size (marked as N/A because the trajectory length exceeded the executable range and was not actually executed, >500 m), and for the larger Building, it surpassed 2000 m. This result indicates that spatial clustering is necessary for producing a feasible number of waypoints.

On the Necessity of the TSP Solver. In the "Proposed w/o TSP Solver" experiment, the final waypoint set was ordered using a simple coordinate-based sorting method. As the visited waypoints remained identical, the reconstruction quality (PSNR) was effectively unchanged. However, the trajectory length increased by 52% (from 18.6 m to 28.3 m) for the Statue and by 61% (from 900.5 m to 1454 m) for the Building. This indicates that TSP-based sequencing is important for reducing travel distance, particularly for large-scale inspection tasks.

On the Necessity of the Geo-NeRF Prior. Finally, the "Proposed w/o Geo-NeRF" configuration, which is identical to the SfM+MVS-Planner baseline, shows a decrease in reconstruction quality. The PSNR dropped by 4.19 dB for the Statue and 4.87 dB for the

Building. This indicates that the quality of the initial geometric prior is a limiting factor for close-proximity planning.

Each component therefore contributes to the final result: the Geo-NeRF model provides the geometric prior, waypoint reduction controls the number of capture locations, and TSP-based sequencing reduces the flight path length.

5. Discussion

This paper presents a georeferenced NeRF-based UAV acquisition framework for high-fidelity 3D reconstruction of static targets. The main contribution is a two-stage planning strategy in which a coarse flight is used to build a metric Geo-NeRF prior, and the extracted geometry is then used to generate surface-aware close-proximity waypoints. The experiments on the Statue and Building targets show that this strategy can reduce the average acquisition distance and improve reconstruction metrics compared with the evaluated SfM/MVS and regular-sampling baselines.

The proposed framework is related to active reconstruction and next-best-view methods, which select views according to uncertainty reduction or expected information gain [32,33]. The focus here is different: the method uses a georeferenced neural representation as a metric geometric prior for UAV waypoint generation. These two directions are complementary. Future work could combine the present surface-following planner with active view-selection criteria to decide which candidate waypoints are most informative.

Several limitations should be noted. First, the current implementation is best suited to static or quasi-static targets with stable geometry. Thin vegetation, small branches, transparent or highly reflective materials, and objects that move during acquisition can introduce errors in the reconstructed geometry and may therefore produce unreliable waypoint locations. Flexible objects affected by wind are particularly challenging because the waypoint may no longer correspond to the surface position at flight time.

Second, the reported close-proximity flights were conducted as supervised autonomous missions, not as fully autonomous collision-avoidance flights. The nearest-neighbor module only orders waypoints and does not perform online obstacle avoidance or local re-planning. Because the nominal camera-center standoff was smaller than the configurable obstacle-avoidance braking distance of the UAV platform, obstacle avoidance was not used as the primary safety mechanism in the fine acquisition stage. Future versions should include visibility-aware edge checking, occupancy- or signed-distance-field-based collision checking, online local replanning, and hardware-specific clearance constraints before deployment in less controlled environments.

Third, the RTK/GNSS data were used for metric coordinate alignment, but a separate hardware-level analysis of camera lever arm, timestamp synchronization, and residual pose error was not performed. This may introduce small discrepancies between the planned camera-center position and the actual optical center. A more complete calibration procedure would improve the reproducibility and safety margin of the method.

Finally, the current trajectory implementation uses stop-and-go waypoint execution and the nearest-neighbor heuristic. This is simple and robust for experimental acquisition, but it is not optimal in terms of flight time, smoothness, or battery usage. Because detailed battery and flight-duration logs were not recorded, this study reports trajectory length as the operational-efficiency proxy. Future work should evaluate 2-opt or exact TSP solvers such as Concorde, continuous B-spline trajectory smoothing, and time-parameterized UAV motion constraints.

In conclusion, the proposed framework is a step toward georeferenced, close-proximity UAV acquisition for high-fidelity 3D digitization, with potential applications in cultural heritage preservation and infrastructure inspection.

Author Contributions: Conceptualization, H.L. and L.Y.; methodology, H.L.; software, H.L.; validation, H.L.; formal analysis, H.L.; investigation, H.L.; resources, L.Y. and L.X.; data curation, H.L. and L.L.; writing—original draft preparation, H.L.; writing—review and editing, L.Y.; visualization, H.L.; supervision, L.Y. and Y.Z.; project administration, L.Y. and H.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available upon request from the corresponding author.

Acknowledgments: The authors acknowledge the support from the School of Automation Engineering, University of Electronic Science and Technology of China, for providing the experimental environment and computing resources.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

Geo-NeRF	Geographic Neural Radiance Fields
UAV	Unmanned Aerial Vehicle
AFP	Average Flight Proximity
PSNR	Peak Signal-to-Noise Ratio
SSIM	Structural Similarity Index Measure
LPIPS	Learned Perceptual Image Patch Similarity
SfM	Structure from Motion
MVS	Multi-View Stereo
RTK	Real-Time Kinematic
MLP	Multi-Layer Perceptron

References

1. Vacca, G.; Vecchi, E. Integrated Geomatic Approaches for the 3D Documentation and Analysis of the Church of Saint Andrew in Orani, Sardinia. *Remote Sens.* **2025**, *17*, 3376. [CrossRef]
2. Zheng, J.; Liu, K. 3D UAV Trajectory Planning with Obstacle Avoidance for UAV-Enabled Time-Constrained Data Collection Systems. *IEEE Trans. Veh. Technol.* **2025**, *74*, 1460–1474. [CrossRef]
3. Zachos, A.; Anagnostopoulos, C.N. Using Terrestrial Laser Scanning, Unmanned Aerial Vehicles and Mixed Reality Methodologies for Digital Survey, 3D Modelling and Historical Recreation of Religious Heritage Monuments. *arXiv* **2023**, arXiv:2401.01380. [CrossRef]
4. Ope, M.K.; Islam, A.; Ameen, M.R.; Miah, A.S.M.; Islam, M.R.; Shin, J. Oitijjo-3D: Generative AI Framework for Rapid 3D Heritage Reconstruction from Street View Imagery. *arXiv* **2025**, arXiv:2511.00362. [CrossRef]
5. Mendoza, M.A.D.; De La Hoz Franco, E.; Gómez, J.E.G. Technologies for the Preservation of Cultural Heritage—A Systematic Review of the Literature. *Sustainability* **2023**, *15*, 1059. [CrossRef]
6. Skublewska-Paszkowska, M.; Milosz, M.; Powroznik, P.; Lukasik, E. 3D technologies for intangible cultural heritage preservation—Literature review for selected databases. *Herit. Sci.* **2022**, *10*, 3. [CrossRef] [PubMed]
7. Kantaros, A.; Ganetsos, T.; Petrescu, F.I.T. Three-Dimensional Printing and 3D Scanning: Emerging Technologies Exhibiting High Potential in the Field of Cultural Heritage. *Appl. Sci.* **2023**, *13*, 4777. [CrossRef]
8. Sun, Y.; Cheng, Y.; Cao, Y.; Zhang, Y.; Shen, W. Multi-View Reconstruction with Global Context for 3D Anomaly Detection. *arXiv* **2025**, arXiv:2507.21555. [CrossRef]
9. Bae, C.; Choi, E.; Lee, S. Technologies, Applications, and Challenges of Digital Twin Across Industries: A Systematic Review of the State-of-the-Art Literature. *IEEE Access* **2025**, *13*, 152843–152869. [CrossRef]
10. Wan, P.; Han, X.; Zhai, R.; Gan, X. Automated Recognition of Rock Mass Discontinuities on Vegetated High Slopes Using UAV Photogrammetry and an Improved Superpoint Transformer. *Remote Sens.* **2026**, *18*, 357. [CrossRef]

11. Chang, J.; Ye, C.; Wu, Y.; Chen, Y.; Zhang, Y.; Luo, Z.; Li, C.; Zhi, Y.; Han, X. ReconViaGen: Towards Accurate Multi-view 3D Object Reconstruction via Generation. *arXiv* **2025**, arXiv:2510.23306. [CrossRef]
12. Bai, Y.; Wong, L.; Twan, T. Survey on Fundamental Deep Learning 3D Reconstruction Techniques. *arXiv* **2024**, arXiv:2407.08137. [CrossRef]
13. Gómez-Raya, G.C.; Veres-Vitályos, Á.; Lemic, F.; Royo, P.; Montagud, M.; Fernández, S.; Abadal, S.; Costa-Pérez, X. Experimental Assessment of Neural 3D Reconstruction for Small UAV-based Applications. *arXiv* **2025**, arXiv:2506.19491. [CrossRef]
14. Maboudi, M.; Homaei, M.; Song, S.; Malihi, S.; Saadatseresht, M.; Gerke, M. A Review on Viewpoints and Path Planning for UAV-Based 3-D Reconstruction. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2023**, *16*, 5026–5048. [CrossRef]
15. Zhang, Y.; Rui, X.; Song, W. A UAV-Based Multi-Scenario RGB-Thermal Dataset and Fusion Model for Enhanced Forest Fire Detection. *Remote Sens.* **2025**, *17*, 2593. [CrossRef]
16. Lee, T.; Kong, Y. Developing 3D River Channel Modeling with UAV-Based Point Cloud Data. *Remote Sens.* **2026**, *18*, 495. [CrossRef]
17. Farah, M.; Alruwaili, A. Optimizing Ground Sampling Distance for Drone-Based GIS Mapping: A Case Study in Riyadh, Saudi Arabia. In Proceedings of the 2024 9th International Conference on Robotics and Automation Engineering (ICRAE), Singapore, 15–17 November 2024; pp. 1–5. [CrossRef]
18. Han, S.; Lu, Z.; Miao, S.; Chang, Q.; Xu, N.; Long, R. Detailed 3D Modeling and Component Monomerization Extraction of Buildings Using Close-Range Photogrammetry. In Proceedings of the 2025 10th International Conference on Image, Vision and Computing (ICIVC), Chengdu, China, 16–18 July 2025; pp. 385–390. [CrossRef]
19. Panagiotopoulou, A.; Grammatikopoulos, L.; El Saer, A.; Petsa, E.; Charou, E.; Ragia, L.; Karras, G. Super-Resolution Techniques in Photogrammetric 3D Reconstruction from Close-Range UAV Imagery. *Heritage* **2023**, *6*, 2701–2715. [CrossRef]
20. Shetty, S.J.; Ravichandran, R.; Tony, L.A.; Abhinay, N.S.; Das, K.; Ghose, D. Chapter 16—Implementation of survivor detection strategies using drones. In *Unmanned Aerial Systems*; Koubaa, A., Azar, A.T., Eds.; Advances in Nonlinear Dynamics and Chaos (ANDC); Academic Press: Cambridge, MA, USA, 2021; pp. 417–438. [CrossRef]
21. Zhang, S.; Liu, C.; Haala, N. Three-Dimensional Path Planning of Uavs Imaging for Complete Photogrammetric Reconstruction. *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.* **2020**, *V-1-2020*, 325–331. [CrossRef]
22. Peng, C.; Isler, V. Adaptive View Planning for Aerial 3D Reconstruction. In Proceedings of the 2019 International Conference on Robotics and Automation (ICRA), Montreal, QC, Canada, 20–24 May 2019; pp. 2981–2987. [CrossRef]
23. Yan, F.; Xia, E.; Li, Z.; Zhou, Z. Sampling-Based Path Planning for High-Quality Aerial 3D Reconstruction of Urban Scenes. *Remote Sens.* **2021**, *13*, 989. [CrossRef]
24. Roberts, M.; Shah, S.; Dey, D.; Truong, A.; Sinha, S.; Kapoor, A.; Hanrahan, P.; Joshi, N. Submodular Trajectory Optimization for Aerial 3D Scanning. In Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017; pp. 5334–5343. [CrossRef]
25. Koch, T.; Körner, M.; Fraundorfer, F. Automatic and Semantically-Aware 3D UAV Flight Planning for Image-Based 3D Reconstruction. *Remote Sens.* **2019**, *11*, 1550. [CrossRef]
26. Schönberger, J.L.; Frahm, J.M. Structure-from-Motion Revisited. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 4104–4113. [CrossRef]
27. Wu, Z.; Marais, P.; Rüther, H. A UAV-based sparse viewpoint planning framework for detailed 3D modelling of cultural heritage monuments. *ISPRS J. Photogramm. Remote Sens.* **2024**, *218*, 555–571. [CrossRef]
28. Smith, N.; Moehrle, N.; Goesele, M.; Heidrich, W. Aerial path planning for urban scene reconstruction: A continuous optimization method and benchmark. *ACM Trans. Graph.* **2018**, *37*, 183. [CrossRef]
29. Mildenhall, B.; Srinivasan, P.P.; Tancik, M.; Barron, J.T.; Ramamoorthi, R.; Ng, R. NeRF: Representing scenes as neural radiance fields for view synthesis. *Commun. ACM* **2021**, *65*, 99–106. [CrossRef]
30. Li, Z.; Wang, Q.; Yue, H.; Nie, X. Toward Automatic 3D Model Reconstruction of Building Curtain Walls from UAV Images Based on NeRF and Deep Learning. *Remote Sens.* **2025**, *17*, 3368. [CrossRef]
31. Li, L.; Zhang, Y.; Jiang, Z.; Wang, Z.; Zhang, L.; Gao, H. Unmanned Aerial Vehicle-Neural Radiance Field (UAV-NeRF): Learning Multiview Drone Three-Dimensional Reconstruction with Neural Radiance Field. *Remote Sens.* **2024**, *16*, 4168. [CrossRef]
32. Krainin, M.; Curless, B.; Fox, D. Autonomous generation of complete 3D object models using next best view manipulation planning. In Proceedings of the 2011 IEEE International Conference on Robotics and Automation, Shanghai, China, 9–13 May 2011; pp. 5031–5037. [CrossRef]
33. Marza, P.; Matignon, L.; Simonin, O.; Batra, D.; Wolf, C.; Chaplot, D.S. AutoNeRF: Training Implicit Scene Representations with Autonomous Agents. In Proceedings of the 2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Abu Dhabi, United Arab Emirates, 14–18 October 2024; pp. 13442–13449. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.