remote sensing

MDPI

Article

Fusion of Airborne and Ground-Based Multi-Source Data for High-Precision 3D Real-Scene Modeling of Historic Cultural District

Huineng Yan [1,2,3], Qi Yuan [4], Yaxin Wen [5,*], Yu Li [1,3], Zhigang Lu [1,2,6] and Rui Wang [1,2,6]

[1] School of Resources and Civil Engineering, Gannan University of Science and Technology, Ganzhou 341000, China; 9320230047@gnust.edu.cn (H.Y.); liyu@lntu.edu.cn (Y.L.); 9320060253@gnust.edu.cn (Z.L.); wangrui@gnust.edu.cn (R.W.)
[2] Key Laboratory of Intelligent and Green Development of Tungsten & Rare Earth Resources, Jiangxi Provincial Department of Education, Ganzhou 341000, China
[3] Ganzhou Key Laboratory of Brain-Inspired Computing and Intelligent Remote Sensing, Ganzhou 341000, China
[4] School of Surveying and Geoinformation Engineering, East China University of Technology, Nanchang 330013, China; yq@ecut.edu.cn
[5] School of Aerospace Intelligent Navigation and IoT, Aerospace Information Technology University, Jinan 250200, China
[6] Ganzhou Key Laboratory of Remote Sensing for Resource and Environment, Ganzhou 341000, China
* Correspondence: wenyaxin@aitech.edu.cn; Tel.: +86-183-9088-8459

Highlights

What are the main findings?

- We propose an automatic distortion region identification algorithm based on image grayscale variation parameters and construct a multi-platform collaborative framework fusing UAV oblique photogrammetry, smartphone close-range photography and RTK positioning.
- We effectively solve data gaps and texture deficiency problems in complex historic districts and realize high-precision 3D modeling with coordinate errors mainly controlled within 10–25 mm.

What are the implications of the main findings?

- The method overcomes the limitations of single UAV modeling, featuring low cost and high efficiency for refined 3D reconstruction of historic cultural districts.
- It provides a reliable technical paradigm for digital preservation, architectural restoration planning, and smart cultural tourism development of historic districts.

Academic Editors: Przemysław Klapa, Massimiliano Pepe, Pelagia Gawronek and Peter Kyseľ

Received: 15 April 2026
Revised: 17 June 2026
Accepted: 1 July 2026
Published: 3 July 2026

Abstract

Traditional Unmanned Aerial Vehicle (UAV) oblique photogrammetry for 3D real-scene modeling of historic cultural districts suffers from data gaps, insufficient texture, and poor accuracy in complex alleyway environments, hindering the widespread adoption of UAV technology. To address these challenges, this paper establishes a distortion region identification algorithm based on image grayscale variation range parameters. Then, through fusing UAV oblique photogrammetry, close-range smartphone photogrammetry, and Real-Time Kinematic (RTK) positioning technology, it ultimately constructs a 3D real-scene reconstruction technical framework. To validate the method's effectiveness and reliability, a field experiment was conducted in the Zaoerxiang Historic Cultural District of Zhanggong District, Ganzhou City, Jiangxi Province, China. The experimental results

demonstrate that the proposed algorithm can effectively identify distortions in the modeling results from UAV images. After fusing smartphone images from distorted regions and RTK measurements from ground control points (GCPs), the discrepancies in X, Y, and Z coordinates between the results and verification points mostly fall within 10 to 25 mm, while the differences from the measured lengths using a steel tape measure and a leveling rod were within the range of 10 to 20 mm. Furthermore, compared to approaches that rely solely on UAV images or on the fusion of UAV and all ground-based images for modeling, the method proposed in this paper restores building texture information in occluded areas and improves the accuracy of 3D real-scene modeling while simultaneously reducing data-processing and storage requirements and enhancing operational efficiency. It provides a referenceable technical framework for digital preservation, restoration planning, and smart cultural tourism of historic districts.

Keywords: UAV oblique photogrammetry; multi-source data fusion; historic cultural districts; 3D real-scene modeling; identification of distortion region

1. Introduction

Historic cultural districts bear the cultural heritage and collective memory of cities [1]. They serve as living repositories of urban historical continuity and act as pivotal nodes connecting the past and future, physical and virtual realms, and materiality and emotional resonance in the digital age [2–5]. As integral components of "digital twin cities", the 3D digitization of historic cultural districts has significant implications for intelligent cultural tourism, heritage preservation, refined conservation of cultural relics, and sustainable urban economic development [6–9].

Due to multiple constraints, including heritage preservation requirements, on-site operational limitations, the need for comprehensive observation, and the inherent characteristics of historic districts (such as rich architectural facade textures, high vegetation coverage, and dense spatial layouts), traditional 3D modeling techniques face significant challenges when constructing realistic models of historic cultural districts. 3D Terrestrial Laser Scanning (TLS) is limited by station placement and scanning field of view, which can easily result in occlusion and shadows in areas with complex structures, leading to gaps in point cloud data and loss of textural detail. Mobile Laser Scanning (MLS), on the other hand, struggles to fully cover pedestrian areas and architectural details because of narrow alleys and restricted access. Furthermore, both methods suffer from high equipment costs and lengthy fieldwork cycles [10,11].

In recent years, Unmanned Aerial Vehicle (UAV) oblique photogrammetry has effectively overcome these limitations through its non-contact, high-efficiency operation, flexible multi-view image acquisition capabilities, and centimeter-level spatial resolution. It has become the mainstream technical solution for high-resolution 3D reconstruction of historic cultural districts [12–15]. Particularly in historical cultural districts characterized by dense architectural textures and complex spatial layouts, this technology can efficiently capture comprehensive facade and roof information, significantly enhancing the geometric completeness and textural accuracy of photorealistic 3D models.

However, constrained by airspace regulations and dense built environments, UAV technology still faces numerous challenges in efficient data collection and high-resolution 3D modeling within historic cultural districts [16–18]. Currently, researchers have investigated bottlenecks in applying UAV oblique photogrammetry in 3D real-world modeling in historic cultural districts.

https://doi.org/10.3390/rs18132171

Erenoglu et al. [19] proposed a novel method based on multi-sensor data acquisition. This method aims to extract and distinguish material characteristics of cultural heritage from UAV photogrammetric data. Additionally, it can generate a 3D model for the Assus Ancient Theater located in Behramkale Village, Çanakkale, Turkey. This study integrated sensors that capture visible light, thermal radiation, and infrared radiation within the electromagnetic spectrum, leveraging the unique properties of each sensor to identify the materials used in the construction of the ancient theater. However, since the three sensor datasets were processed separately, true multi-source data fusion and joint adjustment were not achieved, which prevented the full exploitation of each sensor's advantages.

Li et al. [20] proposed the NRLI-UAV non-rigid registration method, which addresses the issue of point cloud deformation caused by the low-precision Positioning and Orientation System (POS) in low-cost UAV Light Detection and Ranging (LiDAR) systems. This deformation leads to the failure of rigid registration between images and LiDAR point clouds. The method corrects POS errors in the coarse registration by employing Structure from Motion (SfM) technology assisted by Global Navigation Satellite System (GNSS) and Inertial Measurement Unit (IMU) data and transforms 2D-3D registration into 3D-3D registration. Nevertheless, this method is highly dependent on the quality of the SfM depth map and the reliability of the GNSS/IMU-assisted signals.

Martínez-Carricondo et al. [21] utilized UAV orthophotography and oblique photogrammetry to construct an information model for the Cortijo del Fraile in Níjar, Almería Province, Spain. This model encompasses both interior and exterior spaces while accurately capturing architectural features. Its limitations lie in the fact that this modeling primarily serves the purposes of visualization, archival, and documentation. While effectively achieving digital presentation objectives, it fails to provide a high-precision 3D real-scene model for the detailed management of historical buildings.

Ulvi [22] used UAV photogrammetry to generate 3D models, digital surface models, and orthophotos for the cultural heritage site of the Eflatunpınar Hittite Water Monument. By integrating TLS with UAV technology, the accuracy of detail identification in the models was significantly improved. The study focused on comparing the differences between the two technologies in terms of data acquisition sensitivity, software costs, and application techniques, revealing significant distinctions in their characteristics. However, the research did not delve into the advantages and disadvantages of different technologies in deep fusion and refined 3D modeling.

Lin et al. [23] employed UAVs for data collection, utilizing "close-range and omnidirectional" oblique photogrammetry. By fusing image-matching and computer-vision technologies, they constructed a 3D real-scene model of the Santo Stefano Church in Volterra, Italy. This study focused on the application of digital technologies in architectural heritage restoration, achieving restoration through Virtual Reality (VR) technology. However, it did not explore methods for constructing refined 3D models of research subjects through multi-technology fusion.

Inspired by the active perception behavior of owls, Li et al. [24] proposed the Active Environment-aware Optimal Scanning (AEOS) framework to address the degradation in 3D perception and localization performance of UAVs caused by the narrow field of view and payload limitations of compact LiDAR sensors. By integrating a hybrid architecture that combines model predictive control and reinforcement learning, they achieved an inertial odometry system. They also constructed a point cloud simulation environment based on real-world LiDAR maps to support the transfer from simulation to the real-world. However, this framework currently focuses solely on optimizing scanning strategies and improving odometry accuracy for a single LiDAR sensor.

The aforementioned studies [19–24] primarily focus on research methodologies for constructing 3D real-scene models of historical cultural relics using UAV photogramme-

try. Their research objectives tend to focus on digital display or archiving information, such as building material properties, and most studies concentrate on individual structures. Research on fusing multi-source data to construct refined 3D real-scene models of historic cultural districts remain relatively scarce.

For historical, cultural, and architectural districts characterized by wide alleys, sparse buildings, and minimal vegetation cover, UAV photogrammetry can generate high-precision regional 3D real-scene models [25–28]. These models document spatial positioning, building materials, and other current conditions, thereby supporting the management and maintenance of historic structures. However, such ideal conditions are often rare. Constrained by the early economic development levels and construction technology, existing historic cultural districts commonly have narrow alleys, dense buildings, and significant vegetation obstruction [29–31]. In such scenarios, relying solely on UAV photogrammetry data for modeling can lead to defects such as data gaps and texture deficiencies [32,33], making it difficult to meet the demands for authentic representation and high-precision digital preservation of historic cultural districts.

To overcome the limitations of a single aerial platform, the advantages of multiple technologies can be leveraged through data fusion methods [34–37]. Current studies primarily employ the following methods: The first method involves adjusting the UAV flight path and attitude based on the modeling results from a single imaging route [23,38,39], followed by recapturing images of missed or low-quality areas for subsequent 3D modeling, which is extremely time-consuming, relying on accurate identification of data gaps and texture deficiencies within the model. Even when UAVs are redeployed to reshoot low-quality areas, occluded regions may still be missed due to atmospheric perspective limitations. The second method combines UAV photogrammetry with TLS equipment, using 3D laser scan data to supplement areas that are difficult to capture with UAV images [22,40]. However, it involves costly equipment and complex operations and is prone to errors in point cloud-image texture registration. The third method involves manually applying texture patches to neighborhoods with severe texture deficiencies based on the UAV 3D real-scene modeling result [41,42]. This method is labor-intensive and highly dependent on subjective judgment, making it difficult to ensure consistency in geometric textures [42,43].

In summary, current 3D real-scene modeling for historic cultural districts still suffers from issues such as low efficiency, high labor costs, insufficient automation, and reliance on specialized data-processing skills. There is an urgent need to develop low-cost, high-efficiency, high-precision 3D real-scene modeling methods that fuse multi-source data.

As a widely used low-cost imaging device, smartphones now possess the pixel accuracy required for 3D real-scene modeling [44–46]. However, their primary limitation lies in the difficulty of obtaining positional reference information for captured images, which prevents direct application in 3D modeling. By providing coordinate reference data for smartphone images, these images can compensate for the inherent perspective distortion in UAV-based data, enabling refined 3D modeling of historic cultural districts. To address these challenges, this paper proposes a multi-platform collaborative 3D modeling framework, which integrates UAV images, smartphone images, and ground Real-Time Kinematic (RTK) positioning data to efficiently complete the refined 3D modeling of historic cultural districts. Subsequently, a field study was carried out in the experimental area of Zaoerxiang, Ganzhou City, to verify the ability of the framework for high-precision 3D modeling in complex historic cultural districts, and the accuracy and reliability of the results are discussed.

2. Methods

The proposed "air-ground-human" multi-platform collaborative 3D modeling method aims to integrate the respective advantages of UAV images, smartphone images, and RTK technology to overcome the inherent limitations of single aerial platforms, such as data gaps, insufficient accuracy, and the inability to meet detailed requirements. For areas where UAV images cannot be accurately modeled, smartphone imaging technology allows surveyors to implement ground-based densification in areas with lower modeling quality. Concurrently, RTK technology assigns photogrammetric control point coordinates to smartphone images, making them suitable for 3D real-scene modeling. The specific technical workflow is illustrated in Figure 1.

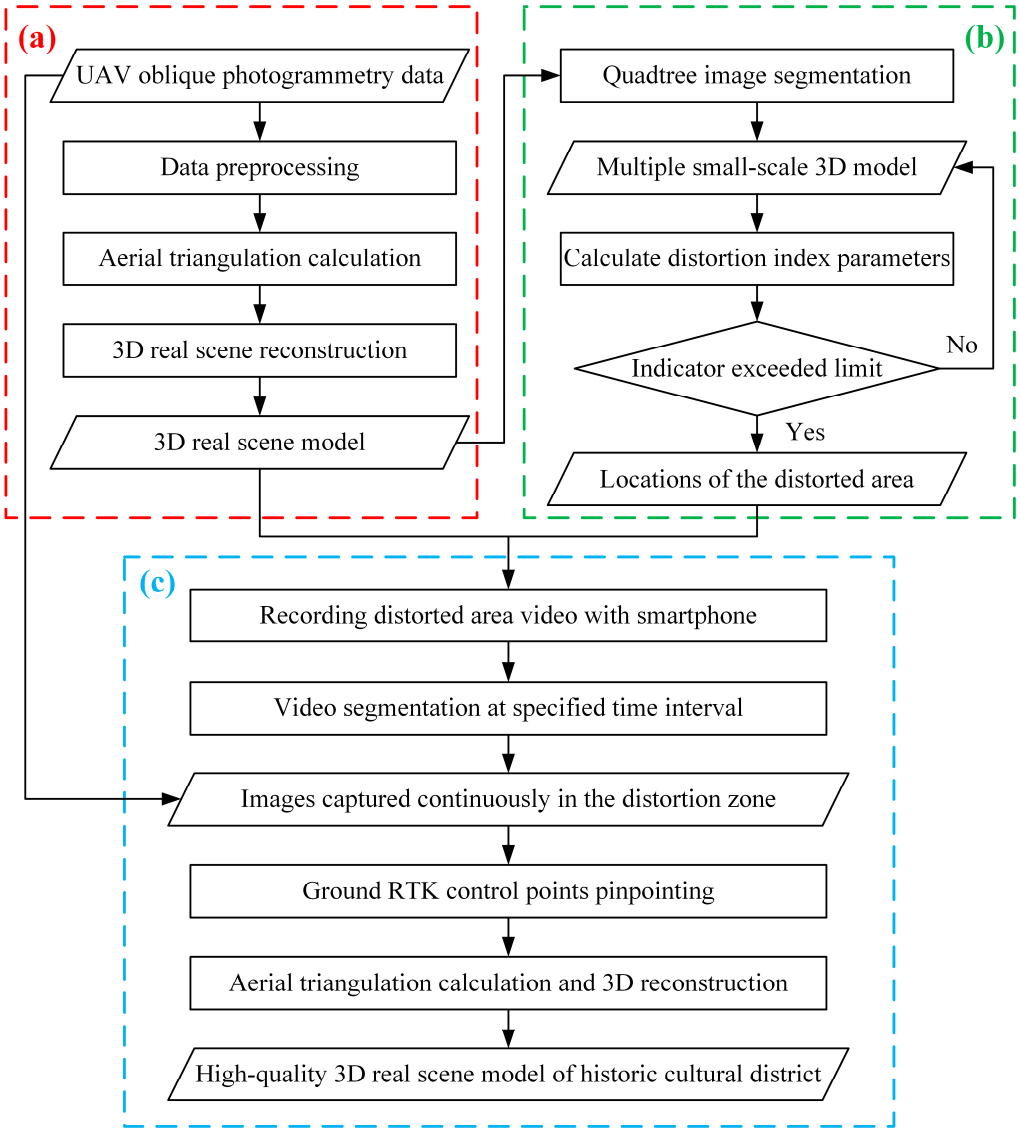

Figure 1. Data-processing flowchart for a multi-platform collaborative 3D real-scene modeling method. (**a**) 3D modeling of UAV images. (**b**) Identification of distorted regions. (**c**) 3D modeling through the fusion of ground-based mobile phone images and UAV images.

As shown in Figure 1, the UAV images, smartphone images, and RTK coordinate data undergo a three-part processing workflow, where (a) 3D modeling is based on the UAV oblique photogrammetry data, (b) analysis of 3D modeling results allows for identifying regions with distortion, and (c) fusion of UAV oblique photogrammetry data with smartphone close-range photography data generates a refined 3D model. For data collec-

tion, the UAV images capture most of the information about the historic cultural district from a low-altitude bird's-eye view. Surveyors positioned smartphones close to the distorted regions to obtain surface texture and ground-detail information, filling the gaps in the UAV's coverage. Ground Control Points (GCPs) provide high-precision coordinate data for smartphone images, enabling verification of the point accuracy in the 3D model. A schematic of the data acquisition process is shown in Figure S1 of the Supplementary Materials.

This study focuses on the specific context of historic cultural districts, which are characterized by narrow alleys, rich textures, and dense vegetation. Relying solely on UAV imagery for modeling can result in distorted and incomplete models. While supplementing the data with ground-level imagery captured by mobile phones can effectively address this issue, this approach has two major drawbacks: the high workload associated with point cloud generation and the significant computational demands of 3D modeling. To address this issue, we propose a method for the automatic identification of distorted areas based on the 3D modeling results of UAV images (as shown in Figure 1). Once distorted areas are identified, the method acquires corresponding smartphone imagery to supplement the data. Compared to traditional approaches, this method not only compensates for the distortions and gaps caused by occlusion in UAV imagery but also avoids the significant increase in workload or computational demands. Against the backdrop of increasing global attention to historic and cultural districts, this method facilitates the widespread adoption of UAV 3D modeling technology and significantly enhances work efficiency.

2.1. Processing of UAV Images

As a key method for 3D digitization, UAV oblique photogrammetry plays a vital role in the preservation and revitalization of historic cultural districts [47,48]. As shown in Figure 1a, during the UAV image processing phase, flight paths are first determined based on factors such as the area scope, block scale, building heights, obstacle conditions, UAV hardware configuration (including lenses, spare batteries, etc.), and airspace control regulations.

After acquiring UAV images, images with excessively small or large memory sizes are first filtered out (due to the high overlap rate in oblique photogrammetry, ground features between adjacent images theoretically exhibit minimal variation, resulting in limited memory fluctuation). Subsequently, the Wallis algorithm [49–51] is applied to adjust the mean and standard deviation of images, achieving brightness balance and color consistency. The principle of the Wallis algorithm lies in adjusting the mean and standard deviation of the color distribution of images to align them with the reference image, thereby achieving color balance across multiple images. Simultaneously, it enhances brightness in low-contrast areas while compressing brightness in high-contrast areas, thereby improving overall image contrast. Furthermore, through local statistics and interpolation processing, it smooths the image while preserving edge information to achieve noise suppression. The specific calculation equations are as follows:

$$f(x,y) = [g(x,y) - m_1] \times \frac{c_1 \times S_2}{c_1 \times S_1 + (1 - c_1) \times S_2} + c_2 m_2 + (1 - c_2) \times m_1 \quad (1)$$

$$g(x,y) = 0.299 \times R(x,y) + 0.587 \times G(x,y) + 0.114 \times \mathrm{B(x,y)} \quad (2)$$

where $f(x,y)$ denotes the grayscale value at pixel coordinate (x,y) in the output image after Wallis transformation, $g(x,y)$ represents the grayscale value of pixel coordinate (x,y) in the original image, m_1 denotes the mean grayscale of the original image, m_2 denotes the mean grayscale of the selected target image, S_1 denotes the standard deviation of the original image, S_2 denotes the standard deviation of the target image, c_1 rep-

resents the contrast expansion factor, c_2 denotes the brightness forcing factor (the specific explanation is provided in Section S2.1 of the Supplementary Materials), and $R(x,y)$, $G(x,y)$ and $B(x,y)$ represent the intensity values of the red, green, and blue channels at pixel coordinate (x,y) in the original image, respectively.

As shown in Equation (1), the control parameters c_1 and c_2 both range from 0 to 1. By mapping the mean and standard deviation, color consistency across multi-source images is achieved. $c_1 = 1$ indicates fully mapping the standard deviation of the original image to that of the target image, resulting in the strongest contrast adjustment; c_1 approaching 0 reduces the standard deviation of the output image, thus compressing contrast; and varying c_1 between 0 and 1 controls the intensity of contrast adjustment. $c_2 = 1$ means the mean of the output image fully converges to the target mean, while $c_2 = 0$ preserves the original mean, with c_2 varying between 0 and 1 to control the intensity of brightness adjustment.

Next, aerial triangulation calculations [52,53] are performed on the adjusted images. In aerial triangulation calculations, using the pre-calibrated initial values of the camera's internal orientation elements and the image's external orientation elements, scale-invariant feature transformation is utilized to automatically match multi-view images and generate control points. They are combined with field-measured control points to construct a regional network for adjustment using the beam method. Then, using the principle of least squares, the coordinates and orientation angles of all image capture locations, along with the object coordinates of densification points, are iteratively solved. Additional parameters are introduced for self-calibration to eliminate lens distortion and systematic errors. Finally, the reliability of the results is evaluated based on the mean error of unit weights and the external conformity accuracy of checkpoints, achieving high-precision reconstruction of the 3D geometric framework of the survey area.

In the specific implementation process, the beam of each individual image serves as an adjustment unit, and an error model is established through collinear equations. By utilizing image-matching and regional grid adjustments, the external vector parameters of the aerial photographs and the coordinates of unknown ground points can be determined. This process integrates aerial and ground images into a bundle adjustment method, establishing an error equation while incorporating GCPs with known coordinates as external constraints to ensure coordinate consistency and accuracy. The equations for constructing the error model are as follows:

$$v_{ij} = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} & a_{15} & a_{16} \\ a_{21} & a_{22} & a_{23} & a_{24} & a_{25} & a_{26} \end{bmatrix} \times \begin{bmatrix} \Delta X_S \\ \Delta Y_S \\ \Delta Z_S \\ \Delta\varphi \\ \Delta\omega \\ \Delta K \end{bmatrix} - \begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \end{bmatrix} \times \begin{bmatrix} \Delta X \\ \Delta Y \\ \Delta Z \end{bmatrix} - \begin{bmatrix} lx \\ ly \end{bmatrix} \tag{3}$$

$$\begin{cases} v_{X_p} = \hat{X}_p - X_{p0} \\ v_{Y_p} = \hat{Y}_p - Y_{p0}\ v \\ v_{Z_p} = \hat{Z}_p - Z_{p0} \end{cases} \tag{4}$$

where v_{ij} denotes the coordinate residual vector of the *i*-*t*h pixel in the *j*-*t*h image, while a_{11}, a_{12}, ……, and a_{26} represent the partial derivatives of the external orientation elements. ΔX_S, ΔY_S, ΔZ_S, $\Delta\varphi$, $\Delta\omega$, and ΔK represent the correction vectors for the external orientation elements, and ΔX, ΔY, and ΔZ denote the correction value vectors for the densified point coordinates. $\hat{X}_p$, $\hat{Y}_p$, and $\hat{Z}_p$ represent the estimated coordinate values of the GCPs; X_{p0}, Y_{p0}, and Z_{p0} denote the measured coordinate values of the GCPs.

Equation (3) can be simplified as its matrix form: $V_1 = A \times \Delta_1 - B \times \Delta_2 - L_1$, where the weighting matrix P_1 is determined based on the estimated variance components

after recalibration; simplify Equation (4) into matrix form: $V_2 = E \times \Delta_3 - L_2$, where the weighting matrix P_2 is determined based on the measurement accuracy of control points. At this point, the error equation and weighting matrices can be combined to yield

$$\begin{bmatrix} V_1 \\ V_2 \end{bmatrix} = \begin{bmatrix} A & -B & 0 \\ 0 & 0 & E \end{bmatrix} \times \begin{bmatrix} \Delta_1 \\ \Delta_2 \\ \Delta_3 \end{bmatrix} - \begin{bmatrix} L_1 \\ L_2 \end{bmatrix} \quad (5)$$

$$P = \begin{bmatrix} P_1 \\ P_2 \end{bmatrix} \quad (6)$$

According to the principle of least squares, the error equation corresponding to Equation (5) and the weight matrix corresponding to Equation (6) must minimize the sum of squared residuals ($V^T PV = \min$). By taking the derivative and setting it to zero, the regularized equation can be derived; its solution represents the optimal estimate of the parameters. Given the large number of observation points in UAV images, the cyclic block method can be employed for the solution, leveraging the sparse matrix structure to reduce computational complexity.

Based on the precise image pose determined by aerial triangulation calculations, the computational domain is first defined by clipping 3D objects of interest to eliminate redundant aerial images and invalid background elements. Subsequently, depending on hardware performance and data volume, the survey area is divided into regular tiles with defined overlap bands to enable parallel processing of large-scale data and effectively manage memory usage. At the resolution level, the target reconstruction resolution and texture atlas specifications are adaptively set according to the ground sampling distance to control the level of detail in the output model and the amount of video memory used. At the same time, a simplification ratio is preset to provide parameter constraints for subsequent multi-level detail output, ultimately forming a comprehensive reconstruction scheme that spans from spatial extent to computational granularity.

Subsequently, based on the principles of multi-view stereovision, scene depth is recovered on a pixel-by-pixel basis using kernel-based constraints and a Multi-View Stereo (MVS) algorithm [54,55], generating a high-density 3D point cloud. After statistical filtering and outlier removal, the discrete point cloud is converted into a continuous manifold surface using the Delaunay triangulation method [56,57], and normal consistency is optimized via Laplacian smoothing. During the texture mapping stage, the optimal source images are selected based on weighting criteria for visibility, angle of incidence, and resolution. Texture seam paths are determined by minimizing graph-cutting energy, and the results are packaged into texture atlases via UV parameterization. Texture feathering is applied in tile overlap regions to eliminate color discrepancies. Finally, a multi-level detail model is generated using a quadratic error-based edge folding algorithm, achieving a topologically closed, geometrically accurate, and visually consistent reconstruction of a realistic 3D surface.

2.2. Identification of Distortion Regions

UAV oblique photogrammetry has become the preferred technology for regional 3D modeling [58–60]. However, as discussed in the Introduction Section, due to the unique characteristics of historic cultural districts, such as the narrow alley effect, vegetation occlusion, and complex textures, 3D real-scene modeling based solely on UAV images often inevitably suffers from texture loss and insufficient accuracy. In such cases, relying solely on the data-processing techniques described in Section 2.1 is insufficient to achieve the desired results. To overcome this bottleneck, there is an urgent need to develop algorithms that can accurately identify distorted regions within UAV modeling results, thus guiding the reshooting of relevant images.

Research on image-based 3D modeling and distortion identification demonstrates that multiple approaches can address distortion issues in localized regions of existing 3D models [61–63]. However, existing approaches generally suffer from limited applicability and excessive implementation complexity, making them ill-suited for real-world production challenges such as diverse model types, ambiguous distortion patterns, and complex noise sources. Furthermore, fusing multiple methods further exacerbates complexity, which hinders engineering implementation. Therefore, this paper proposes a distortion region identification method based on image grayscale value changes, optimized for the characteristics of historic cultural districts.

During 3D model reconstruction, distortion may occur due to multiple factors, such as data acquisition errors, flaws in the reconstruction algorithm, or model simplification processes. In the final output 3D model, these distortions manifest as spatially clustered black voids or white reflective bodies. Consequently, distorted regions exhibit significant differences in grayscale values compared to surrounding areas, whereas correctly modeled regions influenced by terrain feature continuity and topography do not show pronounced spatial grayscale variations. To reduce computational load and enhance the contrast of anomalous distortion regions, this study employs Equation (2) to convert RGB textures into grayscale values.

Next, the degree of difference in texture grayscale values between different regions is calculated to identify distorted regions. To minimize errors and computational load, a quadtree segmentation method is employed to progressively subdivide the study area into smaller units. The mean and variance of texture grayscale values are then computed within each quadtree unit. When the mean and variance of a region significantly deviate from those of the surrounding areas, it indicates a risk of 3D modeling distortion. The core criterion for judgment is whether the deviation exceeds a preset threshold range, which can be calculated as follows:

$$D_{R_i} = \begin{cases} 1 \; if \; V_{R_i} \notin \left[M\left(V_{R_1},\cdots,V_{R_n}\right) - n_1 \times V\left(V_{R_1},\cdots,V_{R_n}\right), M\left(V_{R_1},\cdots,V_{R_n}\right) + n_2 \times V\left(V_{R_1},\cdots,V_{R_n}\right)\right] \\ 1 \; if \; M_{R_i} \notin \left[M\left(M_{R_1},\cdots,M_{R_n}\right) - n_3 \times V\left(M_{R_1},\cdots,M_{R_n}\right), M\left(M_{R_1},\cdots,M_{R_n}\right) + n_4 \times V\left(M_{R_1},\cdots,M_{R_n}\right)\right] \\ 0 \; if \; V_{R_i} \in \left[M\left(V_{R_1},\cdots,V_{R_n}\right) - n_1 \times V\left(V_{R_1},\cdots,V_{R_n}\right), M\left(V_{R_1},\cdots,V_{R_n}\right) + n_2 \times V\left(V_{R_1},\cdots,V_{R_n}\right)\right] \\ 0 \; if \; M_{R_i} \in \left[M\left(M_{R_1},\cdots,M_{R_n}\right) - n_3 \times V\left(M_{R_1},\cdots,M_{R_n}\right), M\left(M_{R_1},\cdots,M_{R_n}\right) + n_4 \times V\left(M_{R_1},\cdots,M_{R_n}\right)\right] \end{cases} \quad (7)$$

where D_{R_i} represents the indicator value for identifying region R_i as a distorted region, a value of 1 indicates a distorted region, while 0 indicates a non-distorted region; V_{R_i} denotes the variance of texture grayscale values within region R_i; $M\left(V_{R_1},\cdots,V_{R_n}\right)$ represents the mean of the variance of texture grayscale values across regions R_1, R_2, …, R_n; n_1 and n_2 represent scaling factor for variance; $V\left(V_{R_1},\cdots,V_{R_n}\right)$ denotes the variance of the texture grayscale values across regions R_1, R_2, …, R_n; M_{R_i} denotes the mean of the texture grayscale values in region R_i; $M\left(M_{R_1},\cdots,M_{R_n}\right)$ denotes the mean of texture grayscale values mean across regions R_1, R_2, …, R_n; n_3 and n_4 denote scaling factor for mean; and $V\left(M_{R_1},\cdots,M_{R_n}\right)$ denotes the variance of the mean texture grayscale values across regions R_1, R_2, …, R_n.

The key to accurately identifying distorted regions lies in correctly determining the coefficients n_1, n_2, n_3, and n_4. Excessively large coefficients may lead to the missed detection of distorted regions, whereas overly small coefficients may misclassify undistorted regions as distorted. This paper selects distortion samples within the region, then tests the identification performance of different coefficients for distorted regions to determine the optimal coefficients. To objectively evaluate the effectiveness of the distortion region, two assessment indexes are established: the identification rate and accuracy for the distortion region, which can be calculated as follows.

$$I_1(n_1, n_2, n_3, n_4) = N_1(n_1, n_2, n_3, n_4)/N_2(n_1, n_2, n_3, n_4) \quad (8)$$

$$I_2(n_1, n_2, n_3, n_4) = N_1(n_1, n_2, n_3, n_4)/N_3(n_1, n_2, n_3, n_4) \quad (9)$$

where $I_1(n_1, n_2, n_3, n_4)$ denotes the distorted region identification ratio, $N_1(n_1, n_2, n_3, n_4)$ represents the number of correctly identified distorted regions, $N_2(n_1, n_2, n_3, n_4)$ denotes the total number of distorted regions, $I_2(n_1, n_2, n_3, n_4)$ represents the distorted region identification accuracy, and $N_3(n_1, n_2, n_3, n_4)$ denotes the total number of identified distorted regions.

Based on the corresponding I_1 and I_2 values for different coefficients n_1, n_2, n_3, and n_4, the optimal coefficients are obtained when both I_1 and I_2 approach their maximum values.

To determine the optimal coefficients n_1, n_2, n_3, and n_4, first select a set of coefficients n_1, n_2, n_3, and n_4, and then use an iterative process to calculate the corresponding I_1 and I_2 values for each coefficient n_1, n_2, n_3, and n_4. I_1 and I_2 both approach their maximum values, which indicates that the distortion region identification is most effective. At this point, based on the I_1 and I_2 values corresponding to different sets of coefficients n_1, n_2, n_3, and n_4, the set of coefficients n_1, n_2, n_3, and n_4 for which both I_1 and I_2 approach their maximum value is selected as the optimal coefficient set.

After determining the values of n_1, n_2, n_3, and n_4, Equation (7) is applied to achieve precise identification of the distorted region. Subsequently, by combining images of the distorted regions obtained through other methods (smartphone images are used in this paper), data fusion techniques are applied to enhance the modeling accuracy of these distorted regions.

2.3. Fusion Processing of Multi-Source Data

Distorted regions primarily occur in locations where UAV image coverage is insufficient or where obstacles exist within the airspace [64,65]. For historical cultural districts, such distortions mainly manifest as low-altitude wall sections, recessed doors and windows, and obstructions from trees. Based on the distortion region identification method proposed in Section 2.2, the distorted regions within the 3D modeling results of UAV images can be precisely located. Given that smartphone close-range photography and low-altitude UAV images complement each other in terms of perspective, and smartphones can replace equipment such as laser scanners, this approach significantly reduces hardware and labor costs while decreasing reliance on specialized data-processing skills. By using smartphone images to complete missing geometric and textural information, it enables the 3D real-scene models with just one flight operation supplemented by ground-based photography.

For the distorted regions identified by the method described in Section 2.2, close-range photogrammetry using handheld smartphones on the ground is employed to supplement the spatial configuration and surface texture details. GCPs are used to jointly estimate the internal and external orientation parameters of the smartphone images. RTK technology determines the coordinates of the GCPs, and the smartphone images are calibrated using the pinpoint method. To facilitate data processing, minimize coordinate system conversion calculations between aerial and ground data, and reduce cumulative measurement errors, the RTK positioning service used for GCPs is identical to that employed for the UAV. The associative relationships among different aerial and ground data are illustrated in Figure 2.

As shown in Figure 2, UAV images are primarily used to establish an initial 3D real-scene model, which serves as foundational reference data for identifying distorted regions via the method described in Section 2.2. Subsequently, close-range photogrammetric images of the distorted regions captured by smartphones are combined with GCPs to solve for internal and external orientation parameters. At this stage, the UAV 3D real-scene

model, GCPs' coordinates, smartphone images, and their internal/external orientation parameters are input into the data-processing system. Through joint adjustment and block-based parallel processing, the geometric structure and texture information of the distorted regions are restored. The error equations and weight matrices within the data fusion processing are analogous to Equations (5) and (6).

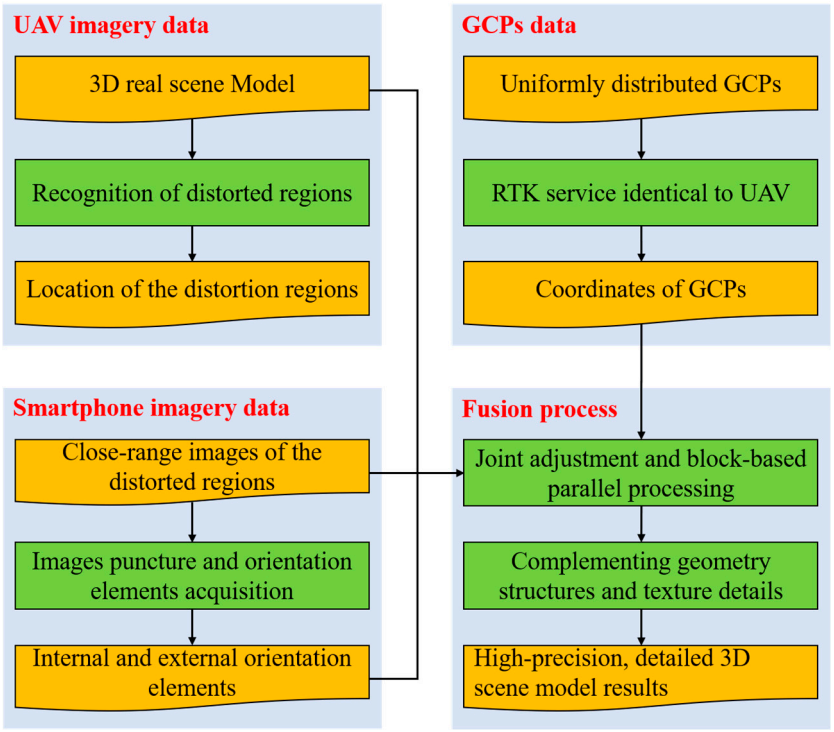

Figure 2. Flowchart of the relationship between UAV images, GCPs, and smartphone photogrammetry data.

To ensure seamless fusion of UAV images, smartphone images, and GCPs, GCPs are uniformly distributed throughout the study area based on the spatial distribution characteristics of the historical cultural district. By fusing these three types of data, the advantages of UAV oblique images, handheld smartphone photography, and ground RTK surveying can be fully leveraged. This approach shows promise in overcoming the limitations of UAV images in detail representation with only a minimal increase in time and cost, offering complementary perspectives, cost-effectiveness, and high-precision fusion. Ultimately, it enables the construction of high-precision 3D real-scene models of historic cultural districts.

3. Experiment

3.1. The Study Area and Data Acquisition

The Zaoerxiang historical cultural district (hereinafter referred to as Zaoerxiang) in Zhanggong District, Ganzhou City, Jiangxi Province, China, is selected as the test subject. Zaoerxiang is located in the northeastern core area of Ganzhou ancient city (25.860°N, 114.947°E), which is surrounded by the Gan River. This curved cobblestone alley evolved from Song Yin Street and earned its name Zaoerxiang during the Ming and Qing dynasties because of the concentration of official residences. Zaoerxiang blends Hakka, Huizhou, and Western architectural styles. Once, its riverside wharf was bustling with ships.

Zaoerxiang winds in an S-shape, stretching 227.3 m in length and having a narrowest width of 3.2 m. On its two sides, there are 47 historical structures, including two-to three-story shops, clan halls, and guild halls, with intricate and elaborate side details. These

buildings combine timber-frame, brick-and-stone, and Western styles, and the vegetation coverage rate is 35% (see Figure 3d). The absence of open takeoff and landing areas within the district, combined with suboptimal RTK aerial photography visibility conditions, posed challenges for UAV operations and GCP deployment. Its narrow alleyways, dense vegetation, and significant aerial visual obstructions collectively contribute to the characteristic features of a typical historic cultural district, making it highly suitable for validating the effectiveness of the methodology proposed in this paper.

Following the methodological workflow outlined in this paper, oblique photogrammetric images of Zaoerxiang are first acquired using a UAV. The experimental UAV employed is the DJI Matrice 300 RTK (Shenzhen DJI Technology Co., Ltd., Shenzhen, China; see Figure 3a), which is equipped with six-direction obstacle avoidance, extended endurance, and long-range video transmission capabilities. The imaging system employs the RIEBO DG3M half-frame five-lens oblique camera (Chengdu RIEBO Technology Co., Ltd., Chengdu, China; see Figure 3b), which features dual medium-telephoto focal lengths of 30 mm (ortho) and 45 mm (oblique). The UAV-mounted RTK service is provided by Shanghai Huace Navigation Technology Ltd. (Shanghai, China), which operates within the CGCS2000 coordinate system.

Based on the block scale, building heights, obstruction conditions, and spatial distribution characteristics of the study area, and considering factors such as UAV takeoff and landing sites (see Figure 3c), regional scope, and equipment endurance, the aerial survey route is designed as a bow-shaped flight path. The flight altitude is set at 100 m, with a forward/backward overlap of 80% and a lateral overlap of 70%. To ensure comprehensive coverage of the survey area by each frame from the five-lens camera, the flight zone was extended outward by 50 m, forming 14 flight paths (see Figure S2 in the Supplementary Materials, yellow dashed lines). Detailed views of the equipment, UAV takeoff/landing sites, regional satellite images, and GCP distribution are shown in Figure 3.

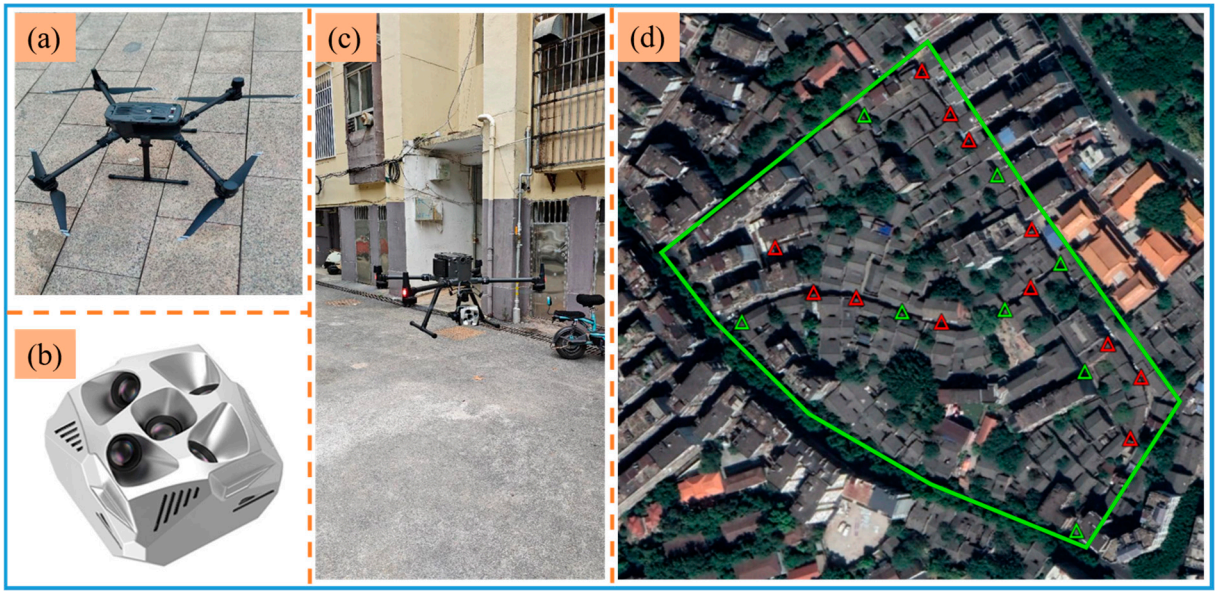

Figure 3. Detailed views of equipment, UAV takeoff/landing sites, and regional satellite images. (**a**) Close-up of DJI Matrice 300 RTK UAV. (**b**) Close-up of RIEBO DG3M half-frame five-lens oblique photography camera. (**c**) UAV takeoff photo. (**d**) Satellite images of the area (obtained from Google Earth), with solid green lines denoting the main scope of Zaoerxiang, and triangles marking GCPs (where red triangles represent points used for 3D modeling and green triangles represent points used for accuracy verification).

As shown in Figure3d, the entire study area (red solid line) covers approximately 133, 741 m², characterized by dense vegetation that creates significant visual obstructions in the airspace. 20 GCPs are established within the study area. These points are distributed relatively uniformly across the region, primarily along main roads and alleys, to facilitate subsequent spatial positioning using smartphone images. Based on the flight path configuration, the actual flight duration for this UAV data collection is 7 min and 28 s, capturing a total of 2265 images that occupied 14 GB of storage space.

To prevent damage to the landscape of the historic cultural district during GCP deployment, ground markers are affixed to printed paper templates using adhesive tape (see Figure S3a in the Supplementary Materials, featuring both diagonally symmetrical rectangles and centrally symmetrical triangles). Special personnel are assigned during the experiment to prevent marker damage. After deploying GCPs, their coordinates are measured using the "Southern Galaxy 1 RTK" surveying equipment (see Figure S3b in the Supplementary Materials, with a horizontal accuracy of 8 mm + 1 × 10^{-6}D (where D is baseline length) and an elevation accuracy of 15 mm + 1 × 10^{-6}D.) and the same RTK service as the UAV (see Figure S3c in the Supplementary Materials). Ground images are captured using a handheld Huawei Mate 80 smartphone, recording an 11 min and 23 s video along a predetermined route (yellow solid line in Figure S3d of the Supplementary Materials) at 30 frames per second. Due to excessive image overlaps within the video and considering human movement speed, two frames per second are selected for extraction, ultimately generating a dataset of 1366 images.

3.2. Airborne and Ground-Based Data Fusion Processing

First, based on the image data acquired by the UAV and the precise positioning service provided by the airborne RTK system, an initial 3D real-scene model of Zaoerxiang is established following the data-processing workflow described in Section 2.1 (as shown in Figure 4a). Due to the large memory footprint of the 3D modeling results, the output data is divided into 25 sub-files for storage, which is based on spatial location.

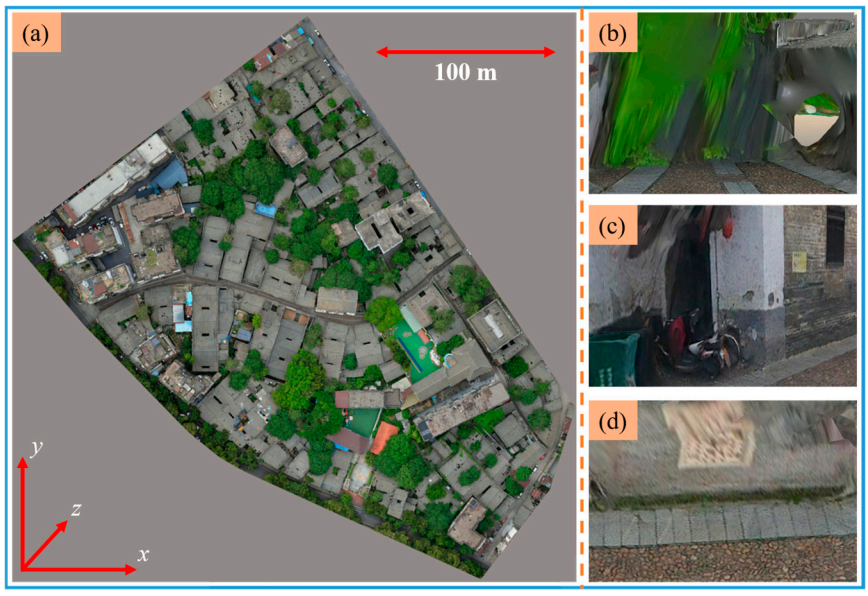

Figure 4. 3D real-scene model based on UAV images and samples of distorted regions. (**a**) 3D real-scene model from UAV images. Subfigures (**b**–**d**) show samples of distorted regions caused by tree obstruction, narrow alleyways, and rich textures, respectively.

Through multi-angle visualization analysis of the modeling quality of the initial UAV 3D real-scene model, it is found that the model effectively captures most terrain features

in areas without significant occlusion, with relatively rich texture details in wide corridor regions. However, in tree-occluded areas and narrow corridors, the model exhibits noticeable voids and loss of texture information, which is consistent with the theoretical predictions discussed earlier. One sub-file is selected from 25 sub-files and subdivided into 32 × 32 regions using a quadtree subsampling method, as specifically shown in Figures S4–S6 in the Supplementary Materials. Through visual inspection, all distorted regions (45 regions in total) within the sub-file are marked as test samples, and three typical examples are shown in Figure 4b–d. This sub-file serves as the experimental subject to determine the optimal coefficients n_1, n_2, n_3, and n_4, for identifying distorted regions in subsequent analyses.

As previously mentioned, 45 distorted regions are identified within the selected sub-file through visual inspection. The performance of distortion detection is evaluated using the selected samples as a benchmark. If all distorted samples can be accurately identified without misclassifying undistorted regions as distorted, this indicates that the identification method possesses good coverage and accuracy rates. The formula for the distortion region identification is shown in Equation (7), with the focus on determining the optimal values for parameters n_1, n_2, n_3, and n_4.

After selecting the distorted samples, the algorithm shown in Figure 1b is used to identify the regions affected by distortion. The specific implementation steps are as follows: firstly, obtain the 3D point cloud data for each sub-region (1024 in total) within the selected sub-file, which includes the X, Y, and Z coordinates and grayscale values of each pixel; secondly, calculate the mean and variance of the grayscale values for all pixels within each sub-region; subsequently, different parameter combinations (n_1, n_2, n_3, and n_4) are iteratively input to identify distorted regions based on Equation (7); after identification, Equations (8) and (9) are used to calculate the distortion identification rate I_1 and identification accuracy I_2, respectively.

Finally, based on the criterion that both I_1 and I_2 reach their maximum values, the optimal parameter combination is determined by analyzing the changes in the identification coverage rate (I_1) and precision rate (I_2) corresponding to different distortion parameter combinations (n_1, n_2, n_3, and n_4). This principle requires that all distorted samples be fully identified while minimizing both misclassifications and missed classifications. Through iterative traversing of multiple sets of parameters, the optimal parameter combination most suitable for distortion identification in this region is ultimately determined. The changes in identification performance metrics corresponding to different parameters are shown in Figures 5 and S11 of the Supplementary Materials.

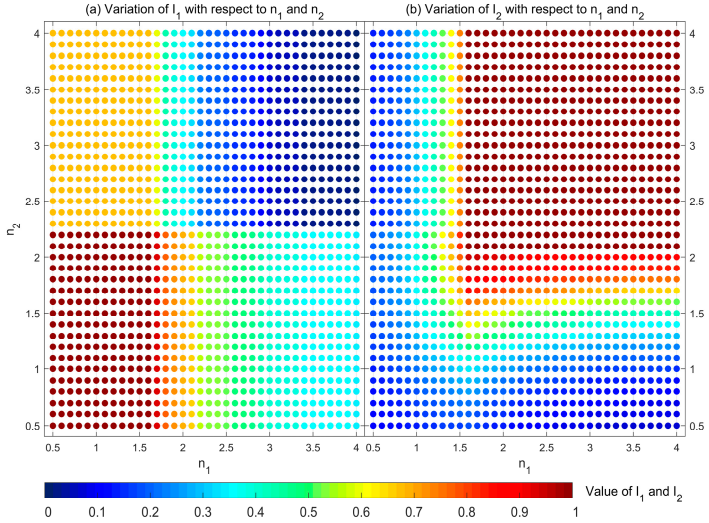

Figure 5. Variation in the values of parameters I_1 and I_2 for different n_1 and n_2 parameters. (**a**) I_1 values. (**b**) I_2 values. A higher value of I_1 indicates greater coverage in identifying distorted re-

gions, increasing the probability that actual distorted regions are identified. A higher value of I_2 indicates greater accuracy in identifying distorted regions, increasing the probability that detected distorted regions are actual distorted regions.

After determining the optimal n_1, n_2, n_3, and n_4, the distortion region identification method proposed in Section 2.2 is applied to process the entire UAV 3D real-scene modeling results. To validate the accuracy of the identification results, a sub-file is randomly selected for verification. By comparing the visual identification results with the algorithm-based identification results, the calculated I_1 and I_2 are 0.98 and 0.92, respectively. This indicates that the proposed method can identify most distorted regions. Although occasional omissions of distorted regions occur, there are minor misclassifications of normal regions as distorted. At this stage, smartphone image data from corresponding areas is employed as supplementary data. Following the method described in Section 2.3, this data is jointly processed with the UAV image to ultimately achieve high-precision 3D modeling results for the study area.

To verify the applicability and robustness of the distortion region identification algorithm under various scene conditions, this study selected the Jinye Cultural Square district in Hanfang Town, Ganxian District, Ganzhou City (hereinafter referred to as Jinye), an area characterized by sparse buildings and rich textural features, as a supplementary experimental site. Jinye presents a stark contrast to Zaoerxiang, the original study area, which is characterized by dense buildings and monotonous textural features. Zaoerxiang is a typical high-density urban historic district with a compact building layout and highly homogeneous facade textures. In contrast, Jinye is an open town square with sparse building distribution, diverse land features, and rich textural information. These two areas represent two typical scenarios, urban historic districts and town public squares, and exhibit significant differences in building density, spatial structure, texture complexity, and land feature composition. This effectively validates the algorithm's universality across diverse scenarios.

Jinye is located in the heart of the market town of Hanfang, Ganxian District (25.490°N, 115.100°E), and it is a distinctive cultural site that has evolved from the local century-old tradition of flue-cured tobacco trade. Grounded in the centuries-old traditional tobacco-trading history of Hanfang Town, this area is a public space that serves multiple public functions, including hosting local folk gatherings, staging cultural performances, and facilitating community interactions. This is a prime example of the organic integration of distinctive historical heritage with modern rural development in southern Ganzhou City. The overall layout of the site is orderly and well-organized, encompassing diverse elements including cultural landscape walls, a folk performance stage, distinctive structures, ancient trees and greenery, and antique-style recreational facilities. It combines man-made landscape structures, extensive hard-surface paving, and natural vegetation, creating varied topography and a diverse spatial structure. Within the area, structures are distributed in a staggered pattern, with dense vegetation covering some sections and intricate details adorning the building facades. This poses challenges, including obstructed views, uneven lighting, and interference from complex ground features. This highly representative scene is suitable for verifying the universality and robustness of the distortion region identification algorithm proposed in this study.

The experimental procedure aligns with that used in Zaoerxiang. First, a 3D real-world model was constructed based on oblique aerial imagery captured by a single UAV. The imaging equipment, flight parameters, and data-processing methods are identical to those used in the earlier Zaoerxiang experiment. The modeling results and three typical distortion regions are presented in Figure S7 in the Supplementary Materials. Subsequently, following the method described above, sub-files of color images, grayscale im-

ages, and grayscale images with a quadtree grid were generated sequentially and are presented in Figures S8–S10, respectively. Fifty-six distortion sample points were selected through visual inspection. Using the aforementioned distortion region identification algorithm, the variations in the distortion region identification performance parameters I_1 and I_2 corresponding to different combinations of threshold parameters are plotted in Figures S12 and S13, respectively.

Figures S12 and S13 demonstrate that the trends of the distortion region recognition performance parameters for Jinye under different threshold parameters are highly consistent with those for Zaoerxiang, which validates the applicability of the distortion recognition algorithm proposed in this study under various scene conditions. After determining the optimal threshold parameters (n1 and n2, n3 and n4), the values of the distortion region recognition performance metrics I_1 and I_2 obtained from the Jinye experiments reached 0.95 and 0.91 (Figure S12) and 0.96 and 0.95 (Figure S13). These results indicate that the algorithm proposed in this study can still accurately identify the vast majority of distorted regions in scenarios characterized by sparse buildings and rich textures, and its recognition accuracy shows good consistency with the results from the Zaoerxiang experiment. These findings confirm that the method proposed in this study has strong adaptability and stable recognition performance when dealing with variations in building density, changes in texture complexity, and transitions between scene types. For comparative analysis and discussion, Figure 6 presents the overall 3D modeling results of the study area, along with the modeling results corresponding to the distorted region samples shown in Figure 4b–d.

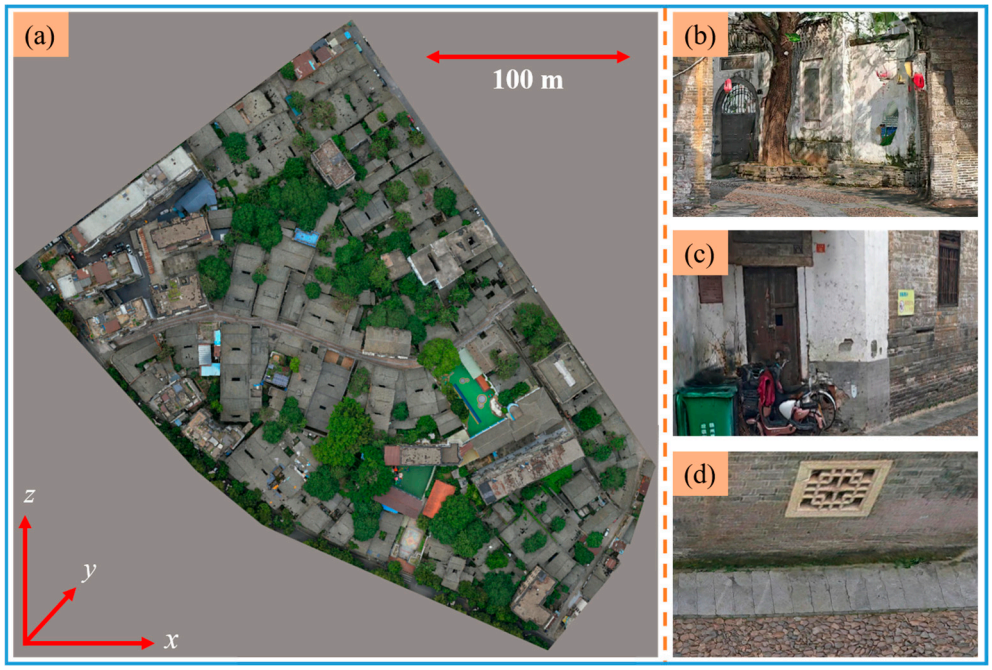

Figure 6. 3D real-scene modeling results fusing UAV images and ground-based smartphone images. Subfigure (**a**) shows the overall 3D modeling results of the study area, and subfigures (**b**–**d**) show the 3D modeling effects after adding ground-based images to the corresponding distorted regions in Figure 4b–d.

To verify the advantages of the method described in this paper over traditional methods (which combine all ground-based smartphone images and UAV images for 3D modeling) in terms of modeling quality and computational complexity, we performed 3D modeling on all collected ground-based smartphone images after pinpointing with the control points. The 3D modeling results of the study area and the distorted region samples

obtained using the traditional method are shown in Figure 7. The time required for pinpointing, computational time, and memory usage of the 3D modeling method using only UAV imagery, the method proposed in this paper, and the traditional method are summarized in Table 1.

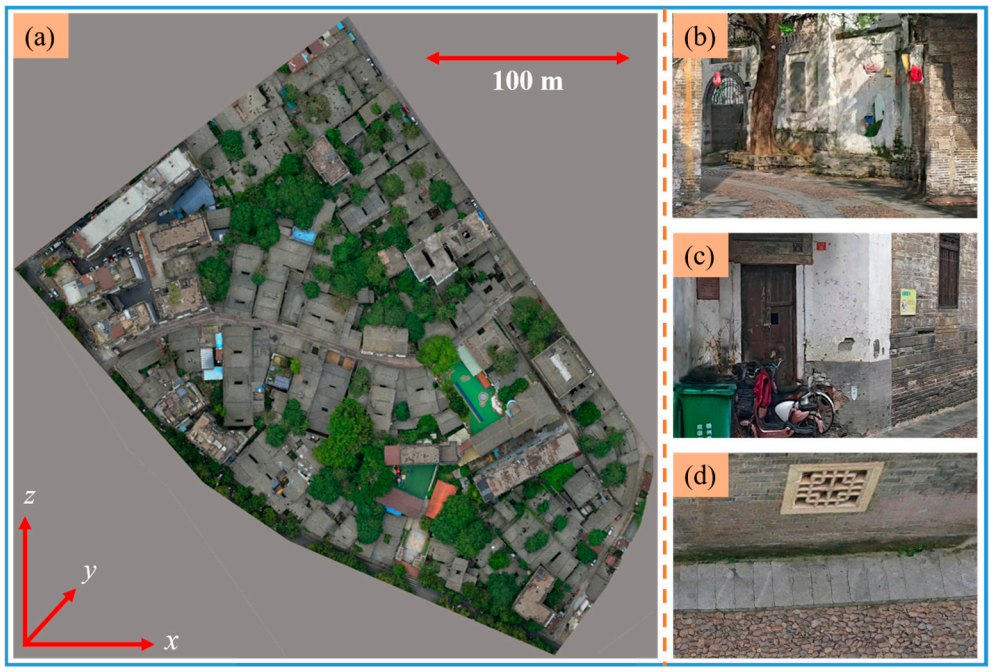

Figure 7. 3D real-scene modeling results fusing UAV images and all ground-based smartphone images. Subfigure (**a**) shows the overall 3D modeling results of the study area; subfigures (**b**–**d**) show the 3D modeling effects after adding ground-based images to the corresponding distorted regions in Figure 4b–d.

Table 1. Comparison of pinpointing time, computational time, and memory usage among different 3D modeling methods.

Statistics	UAV Images	The Proposed Method	The Traditional Method
Pinpoint time	0 min	31 min	4 h 36 min
Calculation time	5 h 6 min	6 h 33 min	8 h 45 min
Memory usage	81.63 GB	106.24 GB	148.25 GB

After establishing the 3D real-scene model, the RTK measurement coordinates of selected GCPs are compared with the 3D modeling results before and after smartphone image fusion. This includes 12 internal modeling GCPs and eight external verification GCPs, where ΔX, ΔY, and ΔZ represent the absolute values of the errors in the three directions, respectively, and $\Delta D = \sqrt{\Delta X^2 + \Delta Y^2 + \Delta Z^2}$. Statistical results are detailed in Tables 2 and S1 of the Supplementary Materials, where the root mean square error (RMSE) and mean absolute error (MAE) are used as statistical metrics.

Table 2. Statistical comparison of RTK-measured coordinates and 3D modeling results using UAV images, the proposed method, and the traditional method (which fuses all ground-based smartphone images and UAV images for 3D modeling) at 8 external verification GCPs.

GCPs	UAV Images (mm)				The Proposed Method (mm)				The Traditional Method (mm)			
	ΔX	ΔY	ΔZ	ΔD	ΔX	ΔY	ΔZ	ΔD	ΔX	ΔY	ΔZ	ΔD
G06	23.8	22.7	18.1	37.5	13.4	20.5	8.7	26.0	12.9	19.9	10.4	25.9
G08	28.8	22.6	19.6	41.5	21.8	19.7	14.9	32.9	20.2	18.4	15.7	31.5

G09	19.7	27.9	11.4	36.0	13.1	22.9	8.5	27.7	12.3	22.1	7.7	26.4
G14	24.9	10.5	27.4	38.5	23.2	9.9	22.6	33.9	22.6	8	21.6	32.3
G15	12.7	27.5	23.8	38.5	11.0	24.6	19.4	33.2	12.9	25.8	18.2	34.1
G16	28.0	14.2	23.4	39.2	18.4	11.7	12.1	24.9	20.2	11.4	11.4	25.8
G17	20.4	10.5	28.0	36.2	15.2	6.3	22.1	27.6	13.9	5.6	22.4	27
G18	22.7	25.2	17.1	38.0	11.8	19.0	11.2	25.0	12.6	17.1	10.8	23.8
RMSE	23.1	21.3	21.7	38.2	16.6	17.9	15.9	29.1	16.4	17.3	15.7	28.6
MAE	22.6	20.1	21.1	38.2	16	16.8	14.9	28.9	16	16	14.8	28.4

To further validate the advantages of the method described in this paper in terms of modeling performance, several typical objects with clearly defined boundaries that are easy to measure within Zaoerxiang (e.g., doors, windows, wall corners, and steps) were selected. Length measurements were taken using a steel tape measure, and vertical distance measurements were taken using an indium–ceramic spirit level; readings are rounded to the nearest millimeter. The field measurement scene is shown in Figure S8 in the Supplementary Materials. To mitigate the impact of random measurement errors, the average of three readings was taken for each measurement as the final result. The measurement results were compared with those obtained from UAV modeling, UAV-fused all ground smartphone images, and the modeling results of the method proposed in this paper. Results from 12 measurement objects are summarized in Table 3, while statistics for the remaining 36 measurement objects are provided in Table S2 in the Supplementary Materials.

Table 3. Statistical comparison of field measurement results for 12 targets versus measurements obtained using UAV modeling, UAV-fusion of all ground-based smartphone images, and the method proposed in this paper.

Number	Measurement Object (Unit: mm)	Measured Value	UAV Images	Proposed Method	Traditional Method	Δ UAV Images	Δ Proposed Method	Δ Traditional Method
1	Width of stone pedestal	578	611.8	559.5	558.5	33.8	−18.5	−19.5
2	Height of stone pedestal	432	394.0	412.2	416.1	−38	−19.8	−15.9
3	Wall height	1025	979.5	1038.2	1036.9	−45.5	13.2	11.9
4	Wall width	717	762.4	701.3	701.7	45.4	−15.7	−15.3
5	Circle diameter	1451	1497.2	1465.5	1461.9	46.2	14.5	10.9
6	Length of stone slab	1530	1583.8	1541.4	1540.9	53.8	11.4	10.9
7	Width of stone slab	1937	1881.1	1952.2	1955.8	−55.9	15.2	18.8
8	Width of door frame	1285	1235.2	1272.4	1272.3	−49.8	−12.6	−12.7
9	Height of door frame	2238	2200.5	2251.1	2254.5	−37.5	13.1	16.5
10	Width of windowsill	1007	956.9	999.1	997.6	−50.1	−7.9	−9.4
11	Height of windowsill	1371	1425.5	1358.8	1355.8	54.5	−12.2	−15.2
12	Length of step	3746	3790.4	3762.8	3761.5	44.4	16.8	15.5

4. Discussion

Based on the experimental results, a comparative analysis was conducted between the 3D modeling results derived solely from UAV images and those obtained by fusing ground-based smartphone images.

4.1. 3D Modeling Results of UAV Images and Distortion Identification

As shown in Figure 4a, the 3D real-scene model constructed from UAV images effectively captures the overall landscape of the study area, where alleys, buildings, vegetation contours, and sports fields are all well-represented. However, due to tree obstructions, narrow alleyways, and overly complex texture information on the buildings themselves, the UAV fails to collect sufficient effective data during aerial surveys in certain areas. This

often results in the loss of surface texture and feature information in the reconstructed model. As shown in Figure 4b–d, the modeling results exhibit noticeable distortions due to tree obstruction, narrow alleys, and rich textures, respectively. These distortions primarily stem from the UAV's shooting distance and methodology. Close-range photogrammetry offers superior building detail and texture resolution, making ground-based smartphone imaging devices an ideal solution for targeted data supplementation.

After dividing the model results into 25 sub-files, one sub-file is selected for the distortion region identification based on the method introduced in Section 2.2. As shown in Figure 5a, when parameters n_1 and n_2 take smaller values, the I_1 value is larger. Specifically, when n_1 ranges from 0.5 to 1.6 and n_2 ranges from 0.5 to 2.2, I_1 equals 1, indicating accurate identification of numerous distorted regions. Simultaneously, as n_1 and n_2 increase, I_1 gradually decreases and eventually approaches 0, signifying a reduction in the number of accurately identified distorted regions. Combining the distortion region determination formula (Equation (7)) with the I_1 parameter calculation formula (Equation (8)), it is evident that when n_1 and n_2 are small, the range of normal regions defined by the mean gray value narrows. Consequently, most regions exceeding this normal range are classified as distorted. Therefore, when n_1 and n_2 are small, I_1 is large, and the proportion of correctly identified distorted regions is high. Conversely, when n_1 and n_2 are large, only a small number of regions are classified as distorted, causing I_1 to gradually decrease as I_1 increases, and the proportion of correctly identified distorted regions to decrease accordingly.

As shown in Figure 5b, the trend of I_2 values with respect to n_1 and n_2 parameters are opposite to that of I_1 values. When n_1 and n_2 are small, I_1 remains low, falling within the range of 0.2 to 0.4. This indicates a high number of misclassified regions among the identified distortion regions, resulting in low identification accuracy. As n_1 and n_2 increase, I_2 gradually rise. When n_1 is between 1.6 and 4 and n_2 is between 2.1 and 4, I_2 reaches 1, indicating that most identified distortion regions correspond to actual distortion regions. Combining the distortion region determination formula (Equation (7)) with the I_2 parameter calculation formula (Equation (9)), it can be observed that when n_1 and n_2 are small, the numerical range for determining regions as normal narrows. This causes some normal regions to be misclassified as distorted, thereby reducing identification accuracy. Conversely, when n_1 and n_2 are large, the numerical range for determining regions as normal expands. At this point, the identified distorted regions correspond to regions where the average gray value significantly deviates from that of other regions, matching the actual distorted regions and thereby improving the identification accuracy.

The experimental results shown in Figures S11 and S13 of the Supplementary Materials exhibit a numerical trend like that in Figure 5: as n_3 and n_4 increase, the value of I_1 gradually decreases, while the value of I_2 gradually increases. For Figure S11, when n_3 varies between 0.5 and 1.7, and n_4 varies between 0.5 and 2.5, I_1 equals 1; when n_3 varies between 1.7 and 4 and n_4 varies between 1.8 and 4, I_2 equals 1. Results in Figures 5 and S7 of the Supplementary Materials indicate that both the variance and mean of image grayscale values can serve as indicators for identifying distorted regions. By setting precise coefficients n_1, n_2, n_3, and n_4, anomalous regions exceeding defined numerical ranges are detected. After balancing distortion region detection coverage and accuracy, the final parameters selected for the study region are as follows: the grayscale variance coefficients n_1 and n_2 set to 1.6 and 2.1, respectively, and the grayscale mean coefficients n_3 and n_4 set to 1.7 and 1.8, respectively.

4.2. 3D Modeling Results from Multi-Source Data Fusion

Based on the 3D modeling results derived from UAV images and the spatial locations of identified distortion regions, ground photography is conducted within the historic cul-

tural district using a smartphone to supplement information on building facades in alleyways and images obscured by trees. Subsequently, the data is fused with the UAV images using the method described in Section 2.3.

As shown in Figure 6a, the 3D modeling results enhanced with smartphone images exhibit minimal differences from the UAV-based model when viewed from a bird's-eye perspective. They clearly reveal elements such as alleyways, buildings, vegetation contours, and sports fields, with boundary lines appearing sharper at edges and corners. Regarding the rendering of building texture details, Figure 6b–d collectively demonstrate that ground-level smartphone images effectively supplement the UAV-based modeling results. Distortions caused by tree obstructions, narrow alleys, and rich textures are effectively suppressed. The model results clearly fill gaps in areas inadequately captured by UAV images, ensuring the complete preservation of detailed information within the historic cultural district. In Figure 7, the results of the modeling that fuses UAV images with all ground-based images are similar to those in Figure 6, demonstrating good restoration of both overall effects and local details. This consistency validates the reliability of the method proposed in this paper.

Regarding tree obstruction, narrow alleys, and texture-rich distortion phenomena, statistical fusion of ground-based smartphone images and UAV images for 3D modeling reveals significantly improved texture information: tree obstructions reduce gap areas by approximately 201 m^2, spatial positioning of 23 distorted regions caused by narrow alleyways is corrected, and 65 door/window objects with rich texture details are successfully modeled with precision. Furthermore, the multi-source data fusion 3D modeling technique identifies 15 buildings at high risk of collapse and 29 locations with severely damaged roof tile structures within the study area, totaling over 367 m^2 of missing tiles. Detailed survey findings have been submitted to the historical district restoration authorities, providing practical value for cultural heritage preservation efforts. These results were also reflected in the fusion of UAV images with all ground-based images, as smartphone images effectively supplemented the observational data for the study area.

Compared to traditional methods that fuse all ground-based images, the method proposed in this paper offers advantages in terms of operational efficiency and resource consumption. As shown in Table 1, the time required for GCPs pinpointing was reduced from 4 h 36 min to 31 min (a reduction of approximately 88.8%), the total computation time was reduced from 8 h 45 min to 6 h 33 min (approximately 25.1% decrease), and memory usage was reduced from 148.25 GB to 106.24 GB (about 28.3% decline). This is primarily attributed to the method's optimization of multi-source data selection, which retains only key ground images that make a substantial contribution to the completion of occluded areas and the improvement of accuracy, thereby effectively eliminating redundant data. Compared to the standalone UAV images solution, although the method described in this paper increases computation time by approximately 28% and memory usage by 30%, the incremental cost is manageable. It effectively restores building textures in occluded areas and improves modeling accuracy while ensuring operational efficiency, providing a feasible solution that balances efficiency and accuracy for large-scale 3D real-scene modeling of historic districts.

A statistical analysis of the differences between the model results for 12 internally modeled GCPs and the RTK measurements (see Table S1 in the Supplementary Materials) indicates that, in the UAV images modeling results, most X, Y, and Z coordinate differences ranged from 10 to 20 mm, with point position errors ranging from 23 to 35 mm (RMSE of 28.2 mm, MAE of 28 mm). In contrast, the results of the fusion method described in this paper show that most X, Y, and Z coordinate differences range from 3 to 20 mm, with point position errors ranging from 14 to 28 mm (RMSE of 20.8 mm, MAE of 20.3 mm), representing an improvement in accuracy of approximately 8 mm compared to the UAV

method. For reference, the X, Y, and Z coordinate differences in traditional full-fusion methods typically range from 3 to 23 mm, with point position errors ranging from 13 to 28 mm (RMSE of 20.6 mm, MAE of 20.1 mm), which is roughly on par with the accuracy of the method described in this paper.

For the eight external validation GCPs, the differences between the model results and the RTK measurements are detailed in Table 2. As shown on the left side of Table 2, the differences in X, Y, and Z coordinates between the UAV modeling results and the GCPs mostly range from 13 to 29 mm, with point errors ranging from 36 to 42 mm (RMSE of 38.2 mm). The middle section of Table 2 shows that the differences between the results of the proposed fusion method and the validation points in the X, Y, and Z directions are primarily concentrated between 6 and 24 mm, with point errors ranging from 25 to 34 mm (RMSE of 29.1 mm), representing an improvement of approximately 9 mm in accuracy compared to the UAV method. The right side of Table 2 also presents the statistical results of the traditional full-fusion method. It shows that the differences in the X, Y, and Z coordinate range from 6 to 26 mm, with point errors ranging from 24 to 34 mm (RMSE of 28.6 mm), and its accuracy is comparable to that of the method described in this paper.

Field measurement validation indicates (Tables 3 and S2 in the Supplementary Materials) that the modeling results based solely on UAV images exhibit significant deviations from actual measurements. For most of the 48 target groups, the absolute error ranges from 30 to 60 mm, with substantial positive and negative fluctuations, indicating the presence of both systematic and random errors. By introducing key ground-based images for fusion optimization, the measurement deviations in the proposed method converged significantly. For more than 80% of the detected targets, the error was controlled within ±20 mm, and the degree of dispersion was significantly reduced. Compared with traditional full-fusion methods, the measurement accuracy of the proposed method is essentially equivalent (the deviation distributions of the two are similar; each has its own advantages and disadvantages, and the RMSE difference is minimal). However, the data-processing volume, point cloud generation time, and memory consumption are all significantly reduced. This indicates that the key image-selective fusion strategy proposed in this paper effectively eliminates redundant ground images, achieving a good balance between accuracy and efficiency while maintaining the geometric measurement accuracy of complex historic district scenarios, thereby demonstrating greater practical applicability in engineering.

Overall, this "air-ground-human" multi-platform collaborative model achieves significant improvements over traditional UAV-based 3D modeling methods. This solution utilizes UAV images as the foundational layer for 3D modeling, identifies distorted regions where UAV-based modeling is suboptimal, and fuses ground-based smartphone photography data to generate a 3D model with higher precision and richer textural information, which is well-suited for promoting high-precision 3D modeling applications in historic cultural districts.

5. Conclusions

Historical cultural districts serve as the core carriers of urban cultural heritage, making the establishment of refined 3D models crucial. While traditional UAV oblique photogrammetry can provide high-precision data, it faces significant limitations in narrow alleys, high-density building areas, and obstacle-dense historic districts. This paper proposes a technical framework based on "air-ground-human" multi-platform collaborative data acquisition and fused air-ground adjustment. It can effectively address issues such as narrow alleyway obstructions, texture loss, and insufficient accuracy in 3D modeling of complex historical cultural districts. Moreover, it can achieve high-precision 3D real-

scene modeling for such regions, significantly enhancing application efficiency and accuracy of UAV oblique photogrammetry.

Using the case study of 3D real-scene modeling in Zaoerxiang, experimental results demonstrate that the fused model significantly enhances the integrity of texture information and the accuracy of the point cloud. It provides a technical framework that can serve as a reference for the digital preservation of similar historical districts. When using the method described in this paper for 3D modeling of historical cultural districts, adjustments must be made to accommodate the following situations. If the study area shows significant textural variations or spatial heterogeneity, it is recommended to implement zoned modeling based on architectural texture characteristics and the modeling scope. If field data collection cannot be completed in a single session, GCPs should be properly preserved, and subsequent data collection should be completed during periods with similar weather and lighting conditions to ensure geometric and radiometric consistency across multiple data sets.

Future studies will focus on adaptive identification and precise localization of distorted regions within deep learning frameworks, enhancing the adaptability of methods for determining optimal threshold parameters, minimizing reliance on external data, and combining with multi-temporal image fusion to achieve refined 3D temporal reconstruction. Ultimately, they will provide quantifiable, replicable technical pathways and empirical evidence for authentic preservation, scientifically informed restoration sequencing, and minimally invasive interventions in historic cultural districts.

In addition, research on the 3D modeling of historical cultural districts will focus on four key areas: artificial intelligence-driven adaptive data collection, joint physical-data distortion identification, semantic-level multimodal fusion, and path planning with regulatory compliance. Through this research, it is anticipated that existing multi-technology integration will be surpassed, leading to genuine methodological breakthroughs.

Supplementary Materials: The following supporting information can be downloaded at https://www.mdpi.com/article/10.3390/rs18132171/s1; Figure S1: Schematic diagram of field data collection for multi-platform collaborative 3D real-scene modeling. Figure S2: Zaoerxiang imagery base map and UAV flight path (yellow dashed line). Figure S3: Ground markers, surveying equipment, GCPs coordinate measurement photos, and smartphone photography routes. (a) Diagonally symmetrical rectangular ground markers and centrally symmetrical triangular ground markers. (b) Close-up of "Southern Galaxy 1 RTK" surveying equipment. (c) Fieldwork photos of measuring GCP coordinates. (d) Regional satellite imagery (sourced from Google Earth), with the smartphone photography route marked by solid yellow lines. Figure S4: The RGB color information of the selected sub-file in the 3D modeling results of the UAV imagery for Zaoerxiang. Due to storage limitations for result presentation, this result is a 1:100 downsampled version of the model result. Figure S5: The greyscale color information of the selected sub-file of the 3D modeling results of the UAV imagery for Zaoerxiang. Due to storage limitations for result presentation, this result is a 1:100 downsampled version of the model result. Figure S6: A 32 × 32 sub-block generated by downsampling the grayscale image of the selected UAV-derived 3D modeling sub-region in Zaoerxiang, using a quadtree method. Due to storage limitations for result presentation, this result is a 1:100 downsampled version of the model result. Figure S7: 3D real-scene model based on UAV images and samples of distorted regions. (a) 3D real-scene model from UAV images. Subfigures (b–d) show samples of distorted regions caused by tree obstruction, narrow alleyways, and rich textures, respectively. Figure S8: RGB color information of the selected sub-file in the 3D modeling results of the UAV imagery. Due to storage limitations for result presentation, this result is a 1:100 downsampled version of the model result. Figure S9: Greyscale color information of the selected sub-file in the 3D modeling results of the UAV imagery. Due to storage limitations for result presentation, this result is a 1:100 downsampled version of the model result. Figure S10: A 32 × 32 sub-block generated

by downsampling the grayscale image of the selected UAV-derived 3D modeling sub-region of Jinye using a quadtree method. Due to storage limitations for result presentation, this result is a 1:100 downsampled version of the model result. Figure S11: Variation in the values of parameters I_1 and I_2 for different n_3 and n_4 parameters. Figure S12: Variation in the values of parameters I_1 and I_2 for different n_1 and n_2 parameters. Figure S13: Variation in the values of parameters I_1 and I_2 for different n_3 and n_4 parameters. Table S1: Statistical comparison of RTK-measured coordinates and 3D modeling results using UAV images, the proposed method, and the traditional method (which fuses all ground-based smartphone images and UAV images for 3D modeling) at 12 internal modeling GCPs. Figure S14: Scene of on-site measurement of architectural details in the Zaoerxiang historic district. Table S2: Statistical comparison of field measurement results for 36 targets versus measurements obtained using UAV modeling, UAV-fusion of all ground-based smartphone images, and the method proposed in this paper.

Author Contributions: Conceptualization, methodology, writing—original draft, formal analysis, supervision, software, H.Y.; data curation, visualization, Q.Y.; data curation, visualization, writing—review and editing, Y.W.; data curation, methodology, writing—review and editing, Y.L.; conceptualization, data curation, supervision, Z.L.; methodology, visualization, writing—review and editing, R.W. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Natural Science Foundation of China (grant No. 52364018), the Science and Technology Research Project of Jiangxi Provincial Department of Education (grant No. GJJ2403702 and GJJ2403703), the Jiangxi Provincial Natural Science Foundation (grant No. 20252BAC240257), and the Shandong Province Natural Science Foundation (grant No. ZR2025QC1055).

Data Availability Statement: Data will be made available on request.

Acknowledgments: We extend our gratitude to Chengdu RIEBO Technology Co., Ltd. (Chengdu, China) for providing the DG3M half-frame oblique photography camera. We also thank Bentley Systems, Inc. (Trenton, Pennsylvania, U.S.) for granting access to the Context Capture software (version 10.20.1.5562). All line-drawing figures in this paper were created using MATLAB (version R2022a, the MathWorks, Inc, Natick, Massachusetts, U.S.).

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

References

1. Jiao, M.; Lu, L. Spatiotemporal Distribution of Toponymic Cultural Heritage in Jiangsu Province and Its Historical and Geographical Influencing Factors. *Herit. Sci.* **2024**, *12*, 377. https://doi.org/10.1186/s40494-024-01492-y.
2. Santagata, W. Cultural Districts, Property Rights and Sustainable Economic Growth. *Int. J. Urban Reg. Res.* **2002**, *26*, 9–23. https://doi.org/10.1111/1468-2427.00360.
3. Le Blanc, A. Cultural Districts, a New Strategy for Regional Development? The South-East Cultural District in Sicily. *Reg. Stud.* **2010**, *44*, 905–917. https://doi.org/10.1080/00343400903427936.
4. Noonan, D.S. How US Cultural Districts Reshape Neighbourhoods. *Cult. Trends* **2013**, *22*, 203–212. https://doi.org/10.1080/09548963.2013.817652.
5. Zhang, Y.; Han, Y. Vitality Evaluation of Historical and Cultural Districts Based on the Values Dimension: Districts in Beijing City, China. *Herit. Sci.* **2022**, *10*, 137. https://doi.org/10.1186/s40494-022-00776-5.
6. Rua, H.; Alvito, P. Living the Past: 3D Models, Virtual Reality and Game Engines as Tools for Supporting Archaeology and the Reconstruction of Cultural Heritage–the Case-Study of the Roman Villa of Casal de Freiria. *J. Archaeol. Sci.* **2011**, *38*, 3296–3308. https://doi.org/10.1016/j.jas.2011.07.015.
7. Nuccio, M.; Ponzini, D. What Does a Cultural District Actually Do? Critically Reappraising 15 Years of Cultural District Policy in Italy. *Eur. Urban Reg. Stud.* **2017**, *24*, 405–424. https://doi.org/10.1177/0969776416643749.
8. Pepe, M.; Costantino, D.; Alfio, V.S.; Restuccia, A.G.; Papalino, N.M. Scan to BIM for the Digital Management and Representation in 3D GIS Environment of Cultural Heritage Site. *J. Cult. Herit.* **2021**, *50*, 115–125. https://doi.org/10.1016/j.culher.2021.05.006.

9. Long, L.; Gan, Z.; Yang, G.; Li, Q. A 3D Reconstruction Pipeline for Generating Textured Models of Large-Scale Architectural Heritage. *J. Build. Eng.* **2025**, *113*, 114064. https://doi.org/10.1016/j.jobe.2025.114064.
10. D'hont, B.; Calders, K.; Bartholomeus, H.; Lau, A.; Terryn, L.; Verhelst, T.; Verbeeck, H. Evaluating Airborne, Mobile and Terrestrial Laser Scanning for Urban Tree Inventories: A Case Study in Ghent, Belgium. *Urban For. Urban Green.* **2024**, *99*, 128428. https://doi.org/10.1016/j.ufug.2024.128428.
11. Salah, R.; Géczy, N.; Ajtayné Károlyfi, K. Predictive Hybrid Scan-to-BIM Method Improves Heritage Building Documentation Completeness and Accuracy. *Sci. Rep.* **2026**, *16*, 7622. https://doi.org/10.1038/s41598-026-38200-8.
12. Fernández-Hernandez, J.; González-Aguilera, D.; Rodríguez-Gonzálvez, P.; Mancera-Taboada, J. Image-Based Modelling from Unmanned Aerial Vehicle (UAV) Photogrammetry: An Effective, Low-Cost Tool for Archaeological Applications. *Archaeometry* **2015**, *57*, 128–145. https://doi.org/10.1111/arcm.12078.
13. Yi, S.; Liu, X.; Li, J.; Chen, L. UAVformer: A Composite Transformer Network for Urban Scene Segmentation of UAV Images. *Pattern Recognit.* **2023**, *133*, 109019. https://doi.org/10.1016/j.patcog.2022.109019.
14. Zhang, C.; Zou, Y.; Wang, F.; Dimyadi, J. Automated UAV Image-to-BIM Registration for Planar and Curved Building Façades Using Structure-from-Motion and 3D Surface Unwrapping. *Autom. Constr.* **2025**, *174*, 106148. https://doi.org/10.1016/j.autcon.2025.106148.
15. Lin, J.; Zhang, J.; Gu, M.; Yu, X.; Zhou, J.; Zheng, J.; Bai, X. RARFLoc: Robust Absolute and Relative Fused Visual Localization for UAVs. *Inf. Fusion* **2025**, *127*, 103905. https://doi.org/10.1016/j.inffus.2025.103905.
16. Jiang, S.; Jiang, W.; Wang, L. Unmanned Aerial Vehicle-Based Photogrammetric 3D Mapping: A Survey of Techniques, Applications, and Challenges. *IEEE Geosci. Remote Sens. Mag.* **2021**, *10*, 135–171. https://doi.org/10.1109/MGRS.2021.3122248.
17. 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. https://doi.org/10.1109/JSTARS.2023.3276427.
18. Adil, M.; Song, H.; Jan, M.A.; Khan, M.K.; He, X.; Farouk, A.; Jin, Z. UAV-Assisted IoT Applications, QoS Requirements and Challenges with Future Research Directions. *ACM Comput. Surv.* **2024**, *56*, 251. https://doi.org/10.1145/3657287.
19. Erenoglu, R.C.; Akcay, O.; Erenoglu, O. An UAS-Assisted Multi-Sensor Approach for 3D Modeling and Reconstruction of Cultural Heritage Site. *J. Cult. Herit.* **2017**, *26*, 79–90. https://doi.org/10.1016/j.culher.2017.02.007.
20. Li, J.; Yang, B.; Chen, C.; Habib, A. NRLI-UAV: Non-Rigid Registration of Sequential Raw Laser Scans and Images for Low-Cost UAV LiDAR Point Cloud Quality Improvement. *ISPRS J. Photogramm. Remote Sens.* **2019**, *158*, 123–145. https://doi.org/10.1016/j.isprsjprs.2019.10.009.
21. Martínez-Carricondo, P.; Carvajal-Ramírez, F.; Yero-Paneque, L.; Agüera-Vega, F. Combination of Nadiral and Oblique UAV Photogrammetry and HBIM for the Virtual Reconstruction of Cultural Heritage. Case Study of Cortijo Del Fraile in Níjar, Almería (Spain). *Build. Res. Inf.* **2020**, *48*, 140–159. https://doi.org/10.1080/09613218.2019.1626213.
22. Ulvi, A. Documentation, Three-Dimensional (3D) Modelling and Visualization of Cultural Heritage by Using Unmanned Aerial Vehicle (UAV) Photogrammetry and Terrestrial Laser Scanners. *Int. J. Remote Sens.* **2021**, *42*, 1994–2021. https://doi.org/10.1080/01431161.2020.1834164.
23. Lin, G.; Li, G.; Giordano, A.; Sang, K.; Stendardo, L.; Yang, X. Three-Dimensional Documentation and Reconversion of Architectural Heritage by UAV and HBIM: A Study of Santo Stefano Church in Italy. *Drones* **2024**, *8*, 250. https://doi.org/10.3390/drones8060250.
24. Li, J.; Xu, X.; Liu, Z.; Yuan, S.; Cao, M.; Xie, L. Aeos: Active Environment-Aware Optimal Scanning Control for Uav Lidar-Inertial Odometry in Complex Scenes. *ISPRS J. Photogramm. Remote Sens.* **2026**, *232*, 476–491. https://doi.org/10.1016/j.isprsjprs.2026.01.006.
25. Bakirman, T.; Bayram, B.; Akpinar, B.; Karabulut, M.F.; Bayrak, O.C.; Yigitoglu, A.; Seker, D.Z. Implementation of Ultra-Light UAV Systems for Cultural Heritage Documentation. *J. Cult. Herit.* **2020**, *44*, 174–184. https://doi.org/10.1016/j.culher.2020.01.006.
26. Yiğit, A.Y.; Uysal, M. Automatic Crack Detection and Structural Inspection of Cultural Heritage Buildings Using UAV Photogrammetry and Digital Twin Technology. *J. Build. Eng.* **2024**, *94*, 109952. https://doi.org/10.1016/j.jobe.2024.109952.
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. https://doi.org/10.1016/j.isprsjprs.2024.10.028.
28. Xu, J.; Zhang, S.; Jing, H.; Hancock, C.; Qiao, P.; Shen, N.; Blay, K.B. Improving Real-Scene 3D Model Quality of Unmanned Aerial Vehicle Oblique-Photogrammetry with a Ground Camera. *Remote Sens.* **2024**, *16*, 3933. https://doi.org/10.3390/rs16213933.
29. Tang, L.; Nikolopoulou, M.; Zhang, N. Bioclimatic Design of Historic Villages in Central-Western Regions of China. *Energy Build.* **2014**, *70*, 271–278. https://doi.org/10.1016/j.enbuild.2013.11.067.

30. Li, N.; Guo, Z.; Geng, W.; Li, L.; Li, Z. Design Strategies for Renovation of Public Space in Beijing's Traditional Communities Based on Measured Microclimate and Thermal Comfort. *Sustain. Cities Soc.* **2023**, *99*, 104927. https://doi.org/10.1016/j.scs.2023.104927.
31. Li, X.; Zhou, X.; Weng, F.; Ding, F.; Wu, Y.; Yi, Z. Evolution of Cultural Landscape Heritage Layers and Value Assessment in Urban Countryside Historic Districts: The Case of Jiufeng Sheshan, Shanghai, China. *Herit. Sci.* **2024**, *12*, 96. https://doi.org/10.1186/s40494-024-01204-6.
32. Sieberth, T.; Wackrow, R.; Chandler, J.H. Automatic Detection of Blurred Images in UAV Image Sets. *ISPRS J. Photogramm. Remote Sens.* **2016**, *122*, 1–16. https://doi.org/10.1016/j.isprsjprs.2016.09.010.
33. López, P.; Zubasti, P.; García, J.; Molina, J.M. A UAV-Based Framework for Visual Detection and Geospatial Mapping of Real Road Surface Defects. Drones **2026**, 10, 119, https://doi.org/10.3390/drones10020119.
34. Liu, N.; Dai, W.; Santerre, R.; Hu, J.; Shi, Q.; Yang, C. High Spatio-Temporal Resolution Deformation Time Series with the Fusion of InSAR and GNSS Data Using Spatio-Temporal Random Effect Model. *IEEE Trans. Geosci. Remote Sens.* **2018**, *57*, 364–380. https://doi.org/10.1109/TGRS.2018.2854736.
35. Yan, H.; Dai, W.; Xie, L.; Xu, W. Fusion of GNSS and InSAR Time Series Using the Improved STRE Model: Applications to the San Francisco Bay Area and Southern California. *J. Geod.* **2022**, *96*, 47. https://doi.org/10.1007/s00190-022-01636-7.
36. Hassani, S.; Dackermann, U.; Mousavi, M.; Li, J. A Systematic Review of Data Fusion Techniques for Optimized Structural Health Monitoring. *Inf. Fusion* **2024**, *103*, 102136. https://doi.org/10.1016/j.inffus.2023.102136.
37. Yan, H.; Lu, Z.; Li, Y.; Wang, R.; Li, F.; Tang, Q. InSAR Atmospheric Error Correction Method Based on GNSS Spatial Interpolation with Application to the Southern North Island of New Zealand. *IEEE Trans. Geosci. Remote Sens.* **2025**, *64*, 4100815. https://doi.org/10.1109/TGRS.2025.3638790.
38. Zheng, X.; Wang, F.; Li, Z. A Multi-UAV Cooperative Route Planning Methodology for 3D Fine-Resolution Building Model Reconstruction. *ISPRS J. Photogramm. Remote Sens.* **2018**, *146*, 483–494. https://doi.org/10.1016/j.isprsjprs.2018.11.004.
39. Chen, P.; Tan, Y.; Yi, W. Adaptive Planning of Multi-UAV Refined Inspection Path for Complex and Irregular Building Clusters. *Autom. Constr.* **2026**, *183*, 106787. https://doi.org/10.1016/j.autcon.2026.106787.
40. Tysiac, P.; Sieńska, A.; Tarnowska, M.; Kedziorski, P.; Jagoda, M. Combination of Terrestrial Laser Scanning and UAV Photogrammetry for 3D Modelling and Degradation Assessment of Heritage Building Based on a Lighting Analysis: Case Study—St. Adalbert Church in Gdansk, Poland. *Herit. Sci.* **2023**, *11*, 53. https://doi.org/10.1186/s40494-023-00897-5.
41. Liu, Z.; Wu, W.; Wang, D.; Cui, B.; Gu, X. Automatic Extraction and 3D Modeling of Real Road Scenes Using UAV Imagery and Deep Learning Semantic Segmentation. *Int. J. Digit. Earth* **2024**, *17*, 2365970. https://doi.org/10.1080/17538947.2024.2365970.
42. Mehrishal, S.; Kim, J.; Shao, Y.; Song, J.J. Artificial Intelligence-Aided Semi-Automatic Joint Trace Detection from Textured Three-Dimensional Models of Rock Mass. *J. Rock Mech. Geotech. Eng.* **2025**, *17*, 1973–1985. https://doi.org/10.1016/j.jrmge.2024.09.031.
43. Shi, Y.; Wang, W.; Zhang, J.; Li, D.; Liu, F.; Li, W. UAV Remote Sensing and Deep Learning for Assessing and Optimizing Architectural Texture in Traditional Villages. *npj Herit. Sci.* **2025**, *13*, 325. https://doi.org/10.1038/s40494-025-01804-w.
44. Rose, G.; Raghuram, P.; Watson, S.; Wigley, E. Platform Urbanism, Smartphone Applications and Valuing Data in a Smart City. *Trans. Inst. Br. Geogr.* **2021**, *46*, 59–72. https://doi.org/10.1111/tran.12400.
45. Zhang, X.; Rui, J.; Xia, G.; Yang, J.; Cai, C.; Zhao, W. Revealing Disparities and Driving Factors in Leisure Activity Segregation of Residents and Tourists: A Data-Driven Analysis of Smart Phone Data. *Appl. Geogr.* **2025**, *176*, 103513. https://doi.org/10.1016/j.apgeog.2025.103513.
46. Niu, Z.; Xi, K.; Liao, Y.; Tao, P.; Ke, T. A Practical Framework for Estimating Façade Opening Rates of Rural Buildings Using Real-Scene 3D Models Derived from Unmanned Aerial Vehicle Photogrammetry. *Remote Sens.* **2025**, *17*, 1596. https://doi.org/10.3390/rs17091596.
47. Li, Q.; Yang, G.; Gao, C.; Huang, Y.; Zhang, J.; Huang, D.; Zhao, B.; Chen, X.; Chen, B.M. Single Drone-Based 3D Reconstruction Approach to Improve Public Engagement in Conservation of Heritage Buildings: A Case of Hakka Tulou. *J. Build. Eng.* **2024**, *87*, 108954. https://doi.org/10.1016/j.jobe.2024.108954.
48. Miky, Y.; Alshawabkeh, Y.; Baik, A. Using Deep Learning for Enrichment of Heritage BIM: Al Radwan House in Historic Jeddah as a Case Study. *Herit. Sci.* **2024**, *12*, 255. https://doi.org/10.1186/s40494-024-01382-3.
49. Wallis, R. An Approach to the Space Variant Restoration and Enhancement of Images. In Proceedings of the Symposium on Current Mathematical Problems in Image Science, Naval Postgraduate School, Monterey CA, USA, 10–12 November, 1976.
50. Gruen, A.; Li, H. Road Extraction from Aerial and Satellite Images by Dynamic Programming. *ISPRS J. Photogramm. Remote Sens.* **1995**, *50*, 11–20. https://doi.org/10.1016/0924-2716(95)98233-P.

51. Gaiani, M.; Remondino, F.; Apollonio, F.I.; Ballabeni, A. An Advanced Pre-Processing Pipeline to Improve Automated Photogrammetric Reconstructions of Architectural Scenes. *Remote Sens.* **2016**, *8*, 178. https://doi.org/10.3390/rs8030178.
52. Liu, K.; Liao, Y.; Yang, K.; Xi, K.; Chen, Q.; Tao, P.; Ke, T. Efficient Radiometric Triangulation for Aerial Image Consistency across Inter and Intra Variances. *Int. J. Appl. Earth Obs. Geoinf.* **2024**, *130*, 103911. https://doi.org/10.1016/j.jag.2024.103911.
53. Yang, Z.; Xu, Y.; Song, H.; Yu, K. Data-Driven Structural Damage Monitoring and Assessment Based on Unmanned Aerial Vehicle Images: A Survey. *Int. J. Digit. Earth* **2025**, *18*, 2528617. https://doi.org/10.1080/17538947.2025.2528617.
54. Furukawa, Y.; Curless, B.; Seitz, S.M.; Szeliski, R. Towards Internet-Scale Multi-View Stereo. In *Proceedings of the 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*; IEEE: New York, NY, USA, 2010; pp. 1434–1441.
55. Furukawa, Y.; Hernández, C. Multi-View Stereo: A Tutorial. *Found. Trends® Comput. Graph. Vis.* **2015**, *9*, 1–148.
56. Boissonnat, J.-D. Geometric Structures for Three-Dimensional Shape Representation. *ACM Trans. Graph. (TOG)* **1984**, *3*, 266–286. https://doi.org/10.1145/357346.357349.
57. Tsai, V.J. Delaunay Triangulations in TIN Creation: An Overview and a Linear-Time Algorithm. *Int. J. Geogr. Inf. Sci.* **1993**, *7*, 501–524. https://doi.org/10.1080/02693799308901979.
58. Mohamed, N.; Al-Jaroodi, J.; Jawhar, I.; Idries, A.; Mohammed, F. Unmanned Aerial Vehicles Applications in Future Smart Cities. *Technol. Forecast. Soc. Change* **2020**, *153*, 119293. https://doi.org/10.1016/j.techfore.2018.05.004.
59. Wu, S.; Feng, L.; Zhang, X.; Yin, C.; Quan, L.; Tian, B. Optimizing Overlap Percentage for Enhanced Accuracy and Efficiency in Oblique Photogrammetry Building 3D Modeling. *Constr. Build. Mater.* **2025**, *489*, 142382. https://doi.org/10.1016/j.conbuildmat.2025.142382.
60. Chen, Y.; Liu, X.; Zhu, B.; Zhu, D.; Zuo, X.; Li, Q. UAV Image-Based 3D Reconstruction Technology in Landslide Disasters: A Review. *Remote Sens.* **2025**, *17*, 3117. https://doi.org/10.3390/rs17173117.
61. Gu, K.; Jakhetiya, V.; Qiao, J.-F.; Li, X.; Lin, W.; Thalmann, D. Model-Based Referenceless Quality Metric of 3D Synthesized Images Using Local Image Description. *IEEE Trans. Image Process.* **2017**, *27*, 394–405. https://doi.org/10.1109/TIP.2017.2733164.
62. Yan, H.; Dai, W.; Xu, W.; Shi, Q.; Sun, K.; Lu, Z.; Wang, R. A Method for Correcting InSAR Interferogram Errors Using GNSS Data and the K-Means Algorithm. *Earth Planets Space* **2024**, *76*, 51. https://doi.org/10.1186/s40623-024-01999-5.
63. Gillani, H.H.; Qureshi, M.A.; Beghdadi, A.; Cheikh, F.; Ullah, M. Distortion Classification in Computer Vision Applications: Current Progress, Challenges, and Perspectives. *ACM Comput. Surv.* **2025**, *58*, 147. https://doi.org/10.1145/3773023.
64. Zhou, Y.; Daakir, M.; Rupnik, E.; Pierrot-Deseilligny, M. A Two-Step Approach for the Correction of Rolling Shutter Distortion in UAV Photogrammetry. *ISPRS J. Photogramm. Remote Sens.* **2020**, *160*, 51–66. https://doi.org/10.1016/j.isprsjprs.2019.11.020.
65. Tian, W.; Sanchez-Azofeifa, A.; Kan, Z.; Zhao, Q.; Zhang, G.; Wu, Y.; Jiang, K. NR-IQA for UAV Hyperspectral Image Based on Distortion Constructing, Feature Screening, and Machine Learning. *Int. J. Appl. Earth Obs. Geoinf.* **2024**, *133*, 104130. https://doi.org/10.1016/j.jag.2024.104130.