buildings

MDPI

Article

Intelligent Decision-Making Method for Multi-Source Heterogeneous 3D Digitization of Ancient Architecture—An Integrated Technical System of Perception–Fusion–Construction–Delivery for High-Precision Digital Assets

Hongzhen Hou [1], Pan Deng [2], Mingduan Zhou [3,*] and Hongzhou Hou [4]

[1] Siwei Shijing Technology (Beijing) Co., Ltd., Beijing 102209, China; 18911364246@163.com
[2] School of Civil and Transportation Engineering, Beijing University of Civil Engineering and Architecture, Beijing 102616, China
[3] School of Geomatics and Urban Spatial Informatics, Beijing University of Civil Engineering and Architecture, Beijing 102616, China
[4] Weituo Times Architectural Design Co., Ltd., Beijing 100025, China
* Correspondence: mdzhou@bucea.edu.cn

Abstract

The 3D digitization of ancient architecture has long been constrained by large spatial-scale spans, heterogeneous precision levels, and inconsistent deliverable standards, with acquisition selection, fusion strategies, and modeling methods depending heavily on expert experience. This paper formalizes these experience-dependent procedures into computable decision models for intelligent multi-source heterogeneous 3D digitization. A three-level decision model is constructed with the target identifiable minimum feature size Rmin and precision levels as decision variables. At the perception level, a scale-aware sensor-selection model organizes six modalities into four complementary acquisition channels and back-calculates the operational parameters from Rmin. At the fusion level, a precision-level-aware three-layer architecture of "datum–skeleton–detail" and a five-component error-budget model establish an explicit and verifiable transfer from instrument precision to deliverable precision. At the construction level, a semantics-driven pipeline automatically produces PBR-LOD digital assets from the C1–C7 semantic categories. Ten ancient-architecture and immovable-relic scenes were used for validation. The horizontal and vertical mean errors of the fused models were 3.1 mm and 3.6 mm, satisfying the <10 mm requirement; the predicted total error (4.5 mm) agreed with the measured value within 1%; and 543 component assets with 8192 × 8192 PBR maps and three LOD levels were delivered with a naming-compliance rate of 100%. The method transforms heritage digitization from experience-dependent operations into reproducible and auditable digital-asset production.

Keywords: ancient-architecture digitization; intelligent decision model; multi-source data fusion; terrestrial laser scanning; oblique photogrammetry; error budget; PBR materials; LOD model; digital assets

Check for updates

Academic Editor: Fabrizio Banfi

Received: 22 October 2026
Revised: 10 November 2026
Accepted: 17 November 2026
Published: 24 November 2026

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

1. Introduction

As typical representatives of immovable cultural relics, the 3D digitization of ancient architecture constitutes the fundamental work for preventive heritage conservation, virtual restoration, cultural dissemination, and digital-content re-creation. Over the past two

decades, measured-geometry technologies represented by terrestrial laser scanning (TLS) and photogrammetry have become the mainstream means of heritage digitization, and have been verified in landmark projects including the Digital Michelangelo Project [1] and the Great Buddha Digitization Project in Nara [2]. With the advancement of Real-Scene 3D China, digital-twin cities, and the generated artificial-intelligence content (AIGC) industry, high-precision 3D digital assets have become a common foundational resource for cultural heritage conservation, virtual display, film-and-game production, and cultural-tourism digitization. Digital-content producers represented by AI creation and application platforms have further placed threefold demands on ancient-architecture-type assets: scale (a single platform often requires thousands of independently callable high-precision models), standardization (a unified classification, naming, unit, coordinate, and material specification is required before assets can be directly retrieved and rendered by engines), and high fidelity (not only visual realism, but also geometric compliance with the measurement requirements of archival documentation and refined asset management). In essence, this has advanced the traditional "ancient-architecture surveying and mapping" into a new stage of "industrial production of ancient-architecture digital assets", and has turned "which device to use for acquisition", "how to fuse multi-source data", and "how to generate reusable assets" into computable problems that must be automatically derived from object features and quantitative indicators.

In response to the above requirements, extensive research has been carried out worldwide on multi-source heterogeneous 3D data acquisition, multi-source point-cloud registration and fusion, and heritage 3D assetization, forming a technical foundation that can be used for reference. A review is given below.

TLS [3] can acquire high-density point clouds in a non-contact manner and has become one of the mainstream technologies for the documentation of cultural heritage [4]; point clouds have further supported the automatic reconstruction of as-built building information models [5]. Levoy et al. [1] verified the ability of laser scanning to restore millimeter-level details of sculptures; Remondino [6] pointed out the complementary relationship between photogrammetry and laser scanning, establishing the basic division of labor that "geometry comes from scanning, and texture comes from imagery". Photogrammetry workflows based on Structure-from-Motion (SfM) [7] and Multi-View Stereo (MVS) [8] have low cost and realistic texture, and UAV oblique photogrammetry has been used to record large sites and historic districts [9,10], with accuracy mainly constrained by ground sample distance (GSD), control-point layout, and image overlap [11–14]. Handheld structured-light/laser scanners are distinguished by portability and sub-millimeter accuracy, and are suitable for the refined acquisition of small- and medium-sized relics such as inscriptions, sculptures, and painted sculptures [15]. Ikeuchi et al. [2] embodied the hierarchical acquisition philosophy from the whole to details through multi-scale digitization of the Great Buddha; polarization imaging [16] and Reflectance Transformation Imaging [17] can improve the readability of low-contrast engraved marks. The capability boundaries of the three types of technologies are also relatively clear: TLS has a ranging accuracy of about ±1 mm, with point spacing growing linearly with distance, making it difficult to remain stably superior to 1.6 mm at 10 m and insufficient to resolve inscription strokes 0.3–1.5 mm wide; moreover, TLS responds unstably to highly reflective, transparent, and dark surfaces. The geometric-accuracy ceiling of UAV photogrammetry is determined by GSD, typically 5–15 mm at flight heights of 30–60 m. Handheld scanners have limited scan coverage, accumulate registration drift over long-distance continuous scanning, and lack natural color. No single technology can span scale ranges exceeding three orders of magnitude, which constitutes the necessity of multi-source collaborative acquisition and integration [18–20]; however, modality selection still relies on expert experience at present, lacking a computable selection method that takes the target scale as a decision variable.

The unification of multi-source point clouds is the core challenge of cross-scale heritage digitization. The classical ICP algorithm [7,21] and its variants such as point-to-plane ICP and generalized ICP [22] converge stably when the initial value is good, but are sensitive to the initial value and easily fall into local optima; in engineering, the two-level strategy of "target-based coarse registration + ICP fine registration" is commonly adopted [23]. For the ternary fusion of "TLS–photogrammetry–handheld scanning", the main challenges lie in three types of differences: precision-level difference, where the local accuracy of handheld scanning (0.1 mm) is much higher than that of TLS (1 mm), while the spatial-datum stability of TLS is superior to that of handheld scanning, so a simple global ICP will "pull the high-accuracy data off"; scale-and-density difference, where photogrammetric point clouds tend to be sparse under eaves and indoors, causing a weighting imbalance in direct fusion; and topological difference, where the same object has different geometric expressions under different modalities (for example, a stele surface is planar under TLS but is a curved surface with engraved undulations under handheld scanning), so ICP will incorrectly eliminate true details. Existing studies have proposed hierarchical registration, weighted least squares, and the "datum-first" strategy [24], but few studies combine them with an explicit error-budget model and quantitatively verify the fusion accuracy with independent check points; fusion accuracy is thus difficult to predict in advance and lacks a reproducible decision basis.

Level of detail (LOD) is an indispensable mechanism for 3D assets in real-time rendering and network transmission scenarios [25]. Biljecki et al. [26,27] systematically improved the LOD specification for 3D city models, pointing out that the definition of LOD should take into account both geometric complexity and semantic richness, but LOD research in the cultural heritage field mostly focuses on the semantic hierarchy of HBIM [28,29], with insufficient discussion of engine-level LOD production workflows constrained by face count and texture specifications. In terms of materials, physically based rendering has become an industry consensus [30], and the glTF 2.0 specification [31] adopts the metallic/roughness workflow, requiring four types of maps (base color, metallic, roughness, and normal) to describe surface optical properties; limestone and wood have large roughness dynamic ranges, painted layers exhibit anisotropic highlights, and gold foil and colored glaze have strong specular reflection, all of which are difficulties in texture acquisition and albedo inversion [32,33]. In terms of deliverable reuse, Scopigno et al. [34] pointed out that heritage 3D models should go beyond the "documentation" level to serve dissemination and reuse; Champion and Rahaman [35] demonstrated their conditions for serving as sustainable scholarly resources; and Münster [36] sorted out the application landscape of digital 3D technologies in humanities research. However, existing research has not yet connected "multi-scale high-precision acquisition" with "engine-level assetization production" into a complete technical chain with full-process accuracy verification; material and LOD production is still mainly manual, lacking a computable and reproducible automated workflow, which is exactly the entry point of this paper.

Synthesizing the above research status, the existing workflows still have systematic deficiencies in four aspects: scale continuity, precision levels, deliverable forms, and workflow intelligence. First, the scale contradiction: ancient-architecture scenes often range from several thousand to tens of thousands of square meters, while the restoration of bracket-set mortise-and-tenon joints, inscription characters, and tile-end patterns requires millimeter or even sub-millimeter resolution. Taking an inscription group as an example, the group layout is on the order of 10 m, a single stele body is on the order of 1 m, and the inscription strokes are on the order of 10^{-4} m, so scale requirements of more than three orders of magnitude coexist within the same site. Ground laser scanning has point spacing constrained by angular resolution and distance, close-range photogrammetry has difficulties in control

and matching, and handheld scanning has limited working coverage; no single technology can bridge this span, and existing studies have not provided a systematic selection method that takes the target scale as the decision variable and explicitly defines the applicable boundaries and handover conditions of each modality. Second, the precision-level contradiction: high-precision local data (handheld scanning, 0.1 mm) and low-precision global data (photogrammetry, centimeter level) are at different orders of magnitude; if a global registration is executed indiscriminately, the "averaging" effect will make the accuracy converge toward the lowest one, causing detail loss or datum drift. Existing workflows generally lack precision-level awareness and also lack a quantitative error-budget model linking the "instrument nominal precision" with the "deliverable precision". Third, the deliverable-form contradiction: traditional deliverables mostly remain at the level of point clouds or one-time-format meshes, lacking physically based rendering (PBR) materials, level-of-detail (LOD) structures, and a standardized naming-and-classification system, so they cannot directly enter real-time rendering engines and AIGC creation workflows, forming an "asset silo" of "reconstruct once, use once". Key engineering details such as the four-channel material production for physically based rendering and the ORM channel merging of metallic and roughness maps during glTF export are rarely systematically reported in the cultural heritage literature. Fourth, insufficient intelligence: key decisions such as device selection, resolution levels, registration strategies, and modeling methods depend heavily on personal experience, and quality control is mainly post hoc manual inspection, lacking computable and reproducible decision models and automated inspection means.

2. Research Objectives and Contributions

The overall objective of this study is to transform ancient-architecture 3D digitization from an experience-dependent operation into a computable, reproducible, and auditable digital-asset production process. Three specific objectives are addressed. The first is to establish a scale-aware sensor-selection model that takes the target identifiable minimum feature size Rmin as the decision variable, so that acquisition modalities, operational parameters, and handover conditions are automatically derived from object features rather than from expert judgment. The second is to construct a precision-level-aware fusion architecture and a verifiable error-budget model, so that the abstract deliverable-accuracy indicator is decomposed into operational parameters that can be controlled item by item and predicted before processing. The third is to build a semantics-driven production pipeline that converts multi-scale high-precision acquisition results into standardized, engine-callable PBR-LOD digital assets.

The contribution of this work to the knowledge base of heritage digitization is threefold. First, it provides a transferable decision framework that converts undocumented expert experience into explicit quantitative rules, which makes acquisition schemes reusable and auditable across sites. Second, it establishes a quantitative error-transfer model that links instrument nominal precision with deliverable precision, filling a methodological gap between instrument specification and acceptance indicators. Third, it demonstrates, with multi-site experimental evidence, that precision-level isolation rather than global averaging is the necessary condition for preserving high-precision details in cross-modality fusion, and that this mechanism can be verified by independent check points. The proposed method is validated through the digitization of ten ancient-architecture and immovable-relic scenes.

3. Data Collection

3.1. Experimental Subjects and Indicator System

The experiments take ten ancient-architecture and immovable-cultural-relic scenes as subjects, anonymously denoted as Site A–Site J. The types cover single halls, pavilions, stone

memorial archways, courtyard groups, inscription groups, cliff/grotto statues, and building-environment composite scenes, with the spatial scale extending from the component level to a group level of about 120 m × 90 m, basically covering the main types of ancient-architecture heritage in China in the three dimensions of form, material, and spatial organization. In order to unify the object boundary and the deliverable calibers, the experiment establishes a three-level object system of "scene layer–building layer–component layer", and divides the component layer into seven major categories, C1–C7, according to the structural logic and material attributes. The object attributes, detail-object categories, acquisition modalities, and acquisition difficulties of each site are shown in Table 1, and the technical-indicator systems corresponding to each level are shown in Table 2; together they form the basis for accuracy control and deliverable acceptance.

Table 1. Integrated table of experimental objects and classification system.

No.	Type	Main Structure/ Material	Scale	Key Detail Objects (with Category Code)	Acquisition Modality	Acquisition Difficulty
Site A	Single hall (main hall)	Timber frame, bracket sets, green-tile roof	Five bays, about 22 m × 14 m	Ridge beasts, tile ends, door-and-window carvings (C2)	TLS + oblique photogrammetry + close-range photogrammetry	Severe under-eave occlusion, partial high reflectivity of the roof
Site B	Single hall (main hall)	Timber structure, stone column bases, painted patterns	About 30 m × 18 m	Beam-and-lintel painted patterns (C5), bracket sets (C1)	TLS + oblique photogrammetry + close-range photogrammetry	Anisotropic highlights of painted patterns, large indoor light ratio
Site C	Pavilion	Multi-eave timber pavilion	About 24 m high, three stories	Railings (C3), sparrow brackets (C1), eave ornaments (C4)	TLS + oblique photogrammetry	Self-occlusion of multi-story eaves, insufficient facade overlap
Site D	Stone memorial archway	Stone archway with relief carvings	About 11 m high, three bays and four pillars	Stone relief carvings, plaque inscriptions (C5)	TLS + handheld scanning + oblique photogrammetry	Large depth span of high relief, sub-millimeter requirement for inscriptions
Site E	Courtyard group	Enclosing walls, covered corridors, pavements	About 60 m × 45 m	Ground pavements, steps, railings (C3)	TLS + oblique photogrammetry	Large group scale, multi-layer occlusion
Site F	Courtyard group	Multi-entry courtyards, mixed brick-stone-timber	About 48 m × 40 m	Brick carvings (C5), gatehouses (C1), pavements (C3)	TLS + oblique photogrammetry + close-range photogrammetry	Poor spatial connectivity, narrow passages
Site G	Inscription group	Bluestone/white-marble inscriptions	Single stele 2.2–3.6 m high	Inscription texts, stele-head patterns (C5)	Handheld scanning + close-range photogrammetry	Engraved depth 0.3–1.2 mm, low contrast under strong light
Site H	Cliff/grotto statues	Sandstone statues	Statues 1.5–4.0 m high	Buddhist-robe folds, face, halo (C5)	Handheld scanning + laser-assisted scanning + close-range photogrammetry	Non-contactable, weathered surface, dust
Site I	Single building (ancestral hall)	Timber structure, brick walls, stone steps	About 18 m × 12 m	Column bases, beam frames (C1), door-stop stones (C3)	TLS + oblique photogrammetry + close-range photogrammetry	Many later alterations, mixed forms
Site J	Immovable scene (building + environment)	Buildings, terrain, vegetation	About 120 m × 90 m	Retaining walls (C1), ancient trees and paths (C7)	Oblique photogrammetry + TLS + close-range photogrammetry	Vegetation occlusion and dynamic interference, rugged terrain

Table 2. Technical-indicator system.

Indicator Category	Specific Indicator	Indicator Requirement
Ground-part acquisition	Three-dimensional laser scanning equipment accuracy	±1 mm
Non-ground-part acquisition	UAV oblique photogrammetry	≥40 megapixels (supplemented by ≥20 megapixel small UAV when necessary)
Acquisition overlap ratio	Laser scanning/oblique photogrammetry	≥40%
Acquisition coverage ratio	Laser scanning/oblique photogrammetry	≥70%
Photo acquisition	Camera pixels/format/illumination/coverage	≥50 megapixels; JPG; uniform fill light; ≥80%
Sculpture-and-inscription acquisition	Handheld-scanner accuracy/overlap/coverage	≤±0.1 mm; ≥50%; ≥80%
Deliverable point cloud	Point-spacing relative accuracy of building scanning	≤10 mm
Deliverable point cloud	Point-spacing relative accuracy of Buddha-statue scanning	≤5 mm
Deliverable accuracy	Horizontal and vertical errors	Both <10 mm
Model format	Format and LOD	FBX, glTF; each contains LOD0 (100% faces), LOD1 (50–60%), LOD2 (20–30%)
Map	Types	Base color, normal, metallic, roughness (metallic and roughness merged on glTF export)
Map	Format/resolution	PNG; no less than 8192 × 8192 (single/multiple)
Asset organization	Standard/unit	Organized, classified, and annotated uniformly by the Eeasten standard; unified unit is meter
Map naming	Specification	[TextureFunction_SequenceNumberOfThisType_UVLayer]

In the three-level object system, the scene layer faces the overall integrity of the site and the environment, the building layer faces the form and spatial relationships of individual buildings, and the component layer faces the measurable geometry and PBR-material assets. C1–C7 are divided according to the dual consistency of the structural logic and the material attributes: C1–C4 are building-body components, distinguished by the functions of load bearing, decoration, ground, and roof; C5 is the attached objects, characterized by sub-millimeter detail requirements; and C6 and C7 are layout-and-environment organizing elements. Among them, the C7 category is a dynamic or semi-dynamic object, whose modeling goal is to provide spatial context for the scene rather than a measurement basis, so the "graded and degraded specification" strategy is adopted: the geometric accuracy is controlled at 5–10 cm, while the material still meets the PBR specification, so as to ensure visual coordination when rendered with the main assets in the same engine.

There are intrinsic tensions among the above indicators that need to be distinguished. First, "acquisition coverage ≥70%" targets building scenes (laser/oblique photogrammetry), "photo acquisition coverage ≥80%" targets photos, and "sculpture-and-inscription acquisition coverage ≥80%" targets handheld scanning; the three have different baselines and

cannot be inferred from one another. Second, "device accuracy ±1 mm" is the instrument nominal accuracy, while "point-spacing relative accuracy ≤10 mm" and "horizontal/vertical errors <10 mm" are deliverable accuracies; they belong to different levels, and the former is a necessary but not sufficient condition of the latter. This paper clarifies their transfer relationship through the error-budget model in Section 4.3.3, and verifies it with independent check points in Section 5.2.

3.2. Implementation of Multi-Source Collaborative Acquisition

The execution effect of the decision model proposed in Section 4 in actual operations, namely whether the Rmin-driven modality selection and resource allocation are reasonable, needs to be tested through multi-site engineering practice. This section reports the implementation process and resource-allocation results of the multi-source collaborative acquisition, in which the regulatory effect of the selection model on resource allocation, the quantitative benefit of precision-level isolation, and the predictive capability of the error-budget model constitute the three key points of verification in this paper.

The acquisition is implemented collaboratively through two channels, ground and non-ground. The non-ground channel uses UAV oblique photogrammetry to acquire the roof, high eaves, and the whole group; the ground channel uses TLS to acquire the building body and confined spaces; non-contactable or high-detail objects such as inscriptions and statues are supplemented by handheld scanning and laser-assisted scanning; and close-range photogrammetry is used for texture-and-color restoration. The total amount of raw data is about 4.6 TB (including about 1.1 TB of fused point clouds), and the implementation over-view is shown in Table 3.

Table 3. Integrated table of acquisition workload and per-site implementation.

Site	TLS Stations	UAV Flights	Handheld Components	Fused Point Cloud (100 Million Points)	Acquisition Overlap	Acquisition Coverage
Workload over-view	10 completed scenes; 842 TLS stations; 96 UAV flights; about 286,000 images total (about 214,000 UAV + 72,000 ground); fused point cloud about 9.63 billion points; 118 handheld-scanned components (37 inscriptions, 64 Buddha statues/sculptures, 17 other attached objects); raw data about 4.6 TB (fused point cloud about 1.1 TB)					
Per-site implementation						
Site A	82	9	3	9.8	46%	82%
Site B	98	11	4	12.6	48%	79%
Site C	112	12	0	14.6	44%	75%
Site D	70	8	12	7.2	52%	86%
Site E	158	13	0	16.8	42%	73%
Site F	120	11	4	13.4	45%	78%
Site G	42	5	44	5.6	58%	88%
Site H	52	6	51	6.9	56%	85%
Site I	54	9	0	5.2	47%	81%
Site J	54	12	0	4.2	49%	76%
Total/worst value	842	96	118	96.3	≥42%	≥73%

The overlap ratio is calculated as the ratio of the projected area in the common-view region where both adjacent stations have valid point clouds to the projected area of a single

station; the coverage ratio is calculated as the ratio of the projected area of the valid point cloud to the theoretical area of the target surface. Since the denominator of the coverage ratio is the theoretical surface area, surfaces that are persistently occluded by vegetation or structurally inaccessible are counted as uncovered portions, which is why the coverage ratios of the sites are generally lower than 90%, with the lowest being 73% at Site E. The measured overlap and coverage ratios of each channel are all superior to the design thresholds: the overlap ratio of the scanning and oblique-photogrammetry channels is not lower than 42% and the coverage ratio is not lower than 73%, respectively, which is higher than the indicator requirements of ≥40% and ≥70%; the coverage ratios of the photo and handheld-scanning channels also satisfy the corresponding thresholds.

The allocation of acquisition resources shows a pattern consistent with the object-scale structure: the sites dominated by building groups have the most TLS stations (Site E reaches 158 stations), the sites dominated by inscriptions and statues have the most handheld-scanned components (Site H reaches 51 pieces and Site G 44 pieces), while Site J, dominated by environment and terrain, mainly uses UAVs (only 12 flights and 54 stations). This indicates that the Rmin-driven selection model can automatically invest the acquisition resources into the modality matching the object-scale structure, and the resource allocation does not need to be re-decided site by site.

3.3. Data Processing

The data-processing procedure is summarized as follows. After the on-site completion checks (QC1), the TLS station point clouds were processed through the four-step workflow of coarse registration, global adjustment, ICP fine registration, and voxel down-sampling described in Section 4.3; the UAV images were processed by SfM–MVS aerial triangulation with ground-control-point constraints, and the aerial-triangulation results were checked for spatial consistency with the TLS datum before participating in the fusion; the handheld scans were registered to their local datum surfaces defined by the TLS control surfaces; the fused point clouds were denoised by statistical outlier and radius filtering and down-sampled by voxel (5 mm for buildings and 1 mm for sculptures); the meshes were constructed by screened Poisson reconstruction [37] with octree depths of 11–14; the PBR maps were baked from the calibrated images and the geometric differences, and LOD1/LOD2 were generated by quadric-error-metrics simplification [38]. Every deliverable passed the QC2–QC4 inspections before acceptance, and the acceptance thresholds are given in Section 5.1.

4. Intelligent Multi-Source Heterogeneous 3D Digitization Method

4.1. Overall Framework of the Intelligent Decision Model

In view of the problems that ancient architecture has complex structures, diverse materials, and intricate details, and that the manual decision links are numerous and poorly reproducible, this paper formalizes the expert experience into a three-level decision model and connects it into a complete automatic decision chain according to the order of "perception–fusion–construction–delivery":

The perception level (intelligent perception layer) is the starting point of decision-making. It takes the target identifiable minimum feature size Rmin as input and automatically outputs the acquisition modality, operational parameters, and modality-handover conditions, solving the problem of "which device and which parameters to use for acquisition". The fusion level (intelligent fusion layer) takes the precision levels as the decision variable and automatically outputs the registration path and error-budget scheme, solving the problem of "how to fuse multi-source data without losing accuracy". The construction level (intelligent construction layer) uses the semantic category as the decision entry, auto-

matically matches the modeling strategy and controls the simplification ratio and texture pixel density, solving the problem of "how to generate reusable, engine-callable assets". The delivery level (intelligent quality-control and delivery layer) uses quantitative thresholds as the inspection basis, automatically judging item by item the field overlap and coverage ratios, point-cloud registration, model geometric accuracy, and map specifications; failure to meet the standards triggers rollback and reprocessing, solving the problem of "how to accept and audit". The four-level decisions are closely interlocked, forming a full-process closed loop from field acquisition to asset delivery, and its overall framework, input–output relationships, and the determination thresholds of each layer are shown in Figure 1.

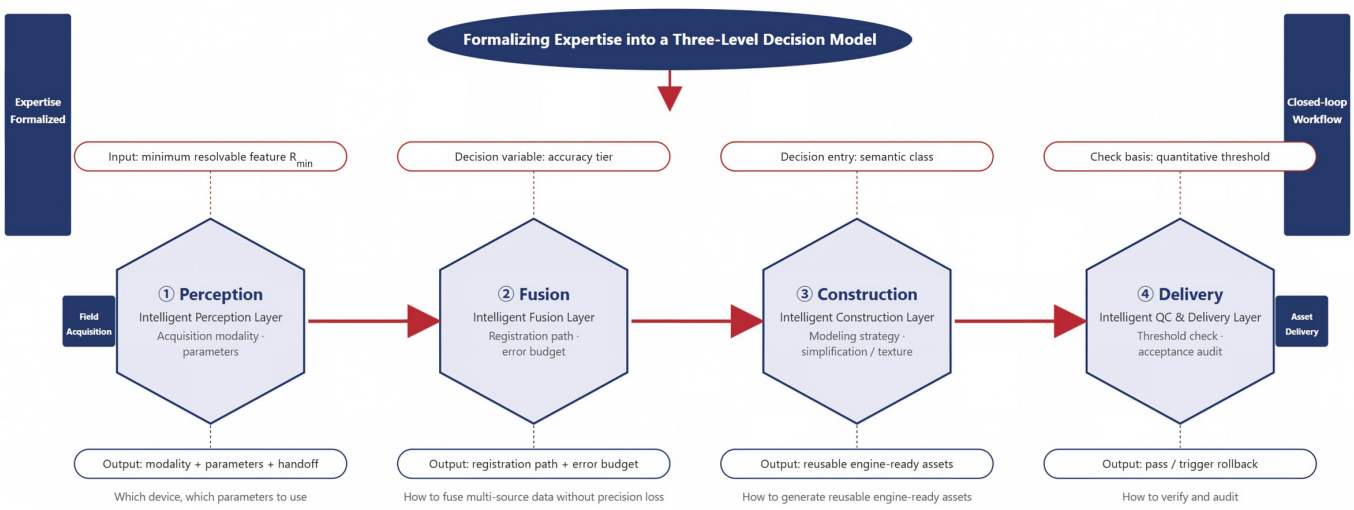


Figure 1. Overall framework of the intelligent decision model for the multi-source heterogeneous 3D digitization of ancient architecture.

The core design principle of this framework is "decision front-loading and unidirectional precision transmission": all the information required for decision-making is determined at the perception level, and no empirical judgment is re-introduced in subsequent links; the precision is transmitted unidirectionally from the finest modality to the deliverables, rather than being averaged during fusion and simplification. The formal descriptions, solution methods, and verification results of the three decision models are given below, respectively.

It should be clarified which stages of the proposed workflow are genuinely automated. The decision links—modality selection and parameter back-calculation (Section 4.2), registration-path and error-budget determination (Section 4.3), strategy matching and LOD-ratio allocation (Section 4.4), and the item-by-item pass/fail judgments of QC1–QC4 (Section 5.1)—are executed automatically from quantitative inputs without expert re-decision. The computational operations of coarse registration, global adjustment, ICP fine registration, mesh construction, QEM simplification, and ORM map merging are executed automatically by software once the decision results are given. Manual intervention remains in three links: the LOD0 master modeling and UV optimization, the regional annotation of metallic components such as gold foil and copper tiles, and the final esthetic correction after simplification. Compared with conventional workflows, in which device selection, fusion strategy, and acceptance are re-decided empirically at each site, the proposed method replaces these experience-dependent decisions with deterministic and auditable rules; the measurable proxies include the automatic structuring of acquisition resources across ten heterogeneous sites without per-site re-decision (Section 3.2), a budget-to-measured deviation within 1% (Section 5.2), and a 100% naming-compliance rate of 543 delivered

assets (Section 5.3). A controlled benchmark of processing time and human-hours against conventional workflows was not included in the present experiment and is planned as future work.

4.2. Scale-Aware Intelligent Sensor-Selection Model

The essence of acquisition-scheme design is to answer three questions: "which device to use, which parameters to adopt, and when to change the device". Traditional operations rely mainly on empirical judgment, and the solutions for the same type of object are often inconsistent across different projects, making them difficult to reuse and audit. This paper formalizes this decision process into a quantitative selection model that takes the target identifiable minimum feature size Rmin as the unique decision variable, so that the acquisition scheme is automatically derived from the object features and reused across sites. The model is solved in three steps.

Step 1: Define the accuracy requirement. Let the minimum feature size that needs to be reliably identified be Rmin (mm), for example, the width of inscription strokes, the thinnest line width of tile-end patterns, and the minimum concave–convex dimensions of bracket-set components. According to the engineering extension of the sampling theorem, to reliably reconstruct a feature, the number of transverse sampling points should be no less than ns (ns = 3–5 is taken in this paper, and the upper limit is adopted for complex curved surfaces), so the required acquisition point spacing is

$$\Delta s_req = Rmin/ns \quad (1)$$

Step 2: Calculate the actual point spacing of each modality. For ground laser scanning, the point spacing $\Delta s_TLS = d \cdot \Delta\theta$ (where d is the distance from the instrument to the target and $\Delta\theta$ is the angular resolution); for UAV photogrammetry, the point spacing equals the ground resolution $\Delta s_UAV = GSD = p \cdot H/f$ (where p is the pixel size, H is the relative flight height, and f is the focal length); for handheld scanners, the nominal 3D resolution under the effective working distance is adopted. Δs_TLS grows linearly with distance, which is the fundamental reason why TLS is difficult to undertake inscription acquisition: when d = 10 m and $\Delta\theta = 0.009°$, Δs already reaches about 1.6 mm; if Rmin = 0.5 mm, then $\Delta s \leq 0.17$ mm is required, corresponding to $\Delta\theta \leq 0.001°$, which has exceeded the capability of current phase-based scanners.

Step 3: Execute the selection criteria. A modality is judged applicable if and only if, within its effective working distance, $\Delta s_modality \leq \Delta s_req$ is satisfied; if multiple modalities satisfy the condition simultaneously, the one with higher efficiency and lower cost is selected; if no modality satisfies the condition, the "hierarchical combination" strategy is executed, in which the large-scale modality provides the geometric skeleton and datum and the small-scale modality provides the detail layer. On this basis, a six-modality intelligent selection matrix is established (Table 4). This matrix is a core contribution of the acquisition methodology in this paper: using Rmin as the query key, the modality, parameters, and handover conditions can be determined, transforming device selection from "empirical judgment" into "table-lookup by Rmin" deterministic decision-making.

The Rmin gates in Table 4 are conservative engineering thresholds rather than direct outputs of Equation (1), because the "actual point spacing" column lists the capability envelope over the full effective working distance, whereas Equation (1) is evaluated at the distance where the target feature is located. For example, TLS at d = 10 m with $\Delta\theta = 0.009°$ yields $\Delta s \approx 1.6$ mm, which satisfies Equation (1) for Rmin $\geq$ 8 mm when ns = 5; the listed gate of 15 mm additionally reserves a margin of about 1.5–2 times for registration error, surface noise, and incidence-angle effects. The same conservative rule is applied to the other modalities: the handheld gates (0.2 mm for M4, 0.6 mm for M5, and 0.3 mm for M6) follow

from ns = 3–5 applied to the upper end of the nominal resolution ranges, and the UAV gate of 20 mm accounts for GSD fluctuation caused by attitude instability and matching degradation on repetitive roof textures. Accordingly, the applicability limits of each modality are as follows: TLS is unsuitable for highly reflective, transparent, or dark surfaces and for features finer than its distance-dependent spacing; UAV photogrammetry is constrained by flight permissions, GSD, and under-eave occlusion; close-range photogrammetry requires rich texture and dense control points; and handheld scanning is limited in coverage and subject to accumulated registration drift, which is why Channel 3 introduces control-point constraints.

Table 4. Scale-adaptive six-modality intelligent selection matrix.

Modality	Representative Device	Effective Working Distance/Flight Height	Actual Point Spacing	Identifiable Rmin	Served Category	Handover Condition with Adjacent Modality
M1 Stationary ground laser scanning	FARO Focus S150 (±1 mm, range 150 m)	2–50 m	1.6–10 mm	≥15 mm	C1, C3, C4, C6	Hand over to M3/M5 when Rmin < 15 mm or occlusion rate > 25%
M2 UAV oblique photogrammetry	DJI Air3/Air3s (48–50 MP), Phantom 4P (1-inch 20 MP, mechanical shutter)	Flight height 25–90 m	GSD 5–15 mm	≥20 mm	C4, C6, C7	Hand over to M1 when Rmin < 20 mm or geometric measurement is required
M3 Ground close-range photogrammetry	≥50 MP full-frame camera + prime lens	0.5–8 m	0.1–1 mm	≥0.5 mm	C2, C4, C5 (texture)	Hand over to M4/M6 when the surface has no texture or strong specularity
M4 Ultra-high-precision handheld scanning	Spider (point accuracy about 0.1 mm, resolution about 0.05 mm)	0.2–0.5 m	0.05–0.2 mm	≥0.2 mm	C5 (inscription, shallow-relief patterns)	Hand over to M5 when the object size > 1.5 m or the overall form is required
M5 Handheld overall-form scanning	Artec EVA (point accuracy about 0.1 mm, resolution about 0.2 mm)	0.4–1.0 m	0.2–0.5 mm	≥0.6 mm	C5 (whole stele, statue)	Control accumulated registration drift with M1 targets when present
M6 Industrial-grade laser-assisted scanning	Scantech/Type-pro devices + global tracking	0.3–4 m	0.05–1 mm	≥0.3 mm	C2, C5 (large stone carvings, refined Buddha statues)	Complementary to M4, used for high-precision acquisition over large working volumes

The six modalities differ in accuracy, efficiency, cost, and coverage, and thus undertake complementary roles. M1 (stationary TLS) offers the most stable spatial datum and millimeter-to-centimeter spacing over tens of meters at moderate cost, and therefore serves as the skeleton and datum provider. M2 (UAV oblique photogrammetry) provides the highest efficiency for roofs and group-scale coverage, but its geometric accuracy is GSD-limited and degrades under eaves. M3 (close-range photogrammetry) reaches sub-millimeter spacing at low equipment cost, but demands dense control, rich texture, and heavy manual work, so it serves mainly texture acquisition and medium-detail geometry. M4–M6 provide the finest resolution—M4 for sub-millimeter details of small objects, M5 for meter-scale overall forms, and M6 for large working volumes beyond the reach of handheld tracking—at the cost of equipment price, operator skill, and limited coverage, which is why they are organized as detail channels operating under the datum of M1.

Organization of the Six Modalities into Four Functional Channels

The six modalities are further organized into four functionally complementary acquisition channels at the operational level, so that the selection result can be directly mapped into an executable operational organization scheme and significantly reduce repeated on-site work. Channel 1 is the global-skeleton acquisition (M1), which uses stationary ground laser scanning to acquire the unified geometric skeleton of the building body, indoor space, and the group, and undertakes the establishment of the whole-site spatial datum; the resolution level is back-calculated by $\Delta s = d\Delta\theta$ combined with the target Rmin, so the parameters become the execution result of the selection model rather than empirical default values. Channel 2 is the overall-form acquisition (M2), which uses UAV oblique photogrammetry to acquire the overall geometry and texture of roofs, high eaves, and the group according to the composite route of "five-direction oblique + orbiting + close-range". Channel 3 is the high-precision detail acquisition (M4/M5/M6), which targets attached objects with Rmin < 20 mm, acquires them block by block according to Rmin, and introduces control-point constraints to suppress the accumulated registration drift; a non-contact working mode is adopted for non-contactable objects. Channel 4 is the texture-and-color acquisition (M2/M3), which uses cross-polarized photography to eliminate specular highlights and obtain nearly pure diffuse-reflection images, supplemented by color-chart calibration to ensure cross-station color consistency. The object division, operational parameters, and modality-handover conditions of the four channels constitute a complete executable specification of the acquisition scheme.The specific details are shown in the Figure 2.

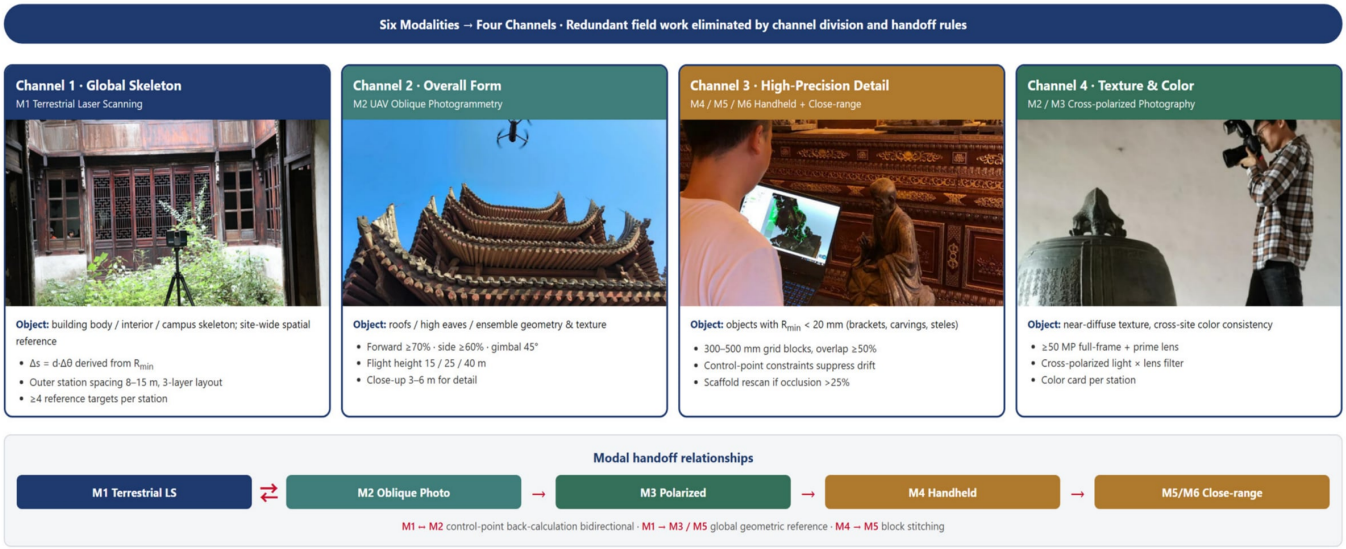


Figure 2. Multi-source heterogeneous collaborative acquisition system and equipment working modes.

In addition, unified lighting and on-site quality-inspection rules are executed in the field. Operations are preferably carried out in the period of uniform diffuse light; for sensitive objects such as inscriptions and painted patterns, portable fill-light groups are used to ensure uniform fill light, avoid highlight overflow and shadow dead zones, and thus guarantee the unbiased estimation of the PBR base-color map. After each scan station is completed, the point-cloud completeness and target-recognition quality are checked immediately, and the station is re-scanned on site when the target residual exceeds the limit (>3 mm).

4.3. Precision-Level-Aware Multi-Source Registration and Fusion and Error Budget

4.3.1. Two-Level Control Network and Global Datum

The premise for multi-source data fusion is a unified spatial datum. This paper uniformly establishes a two-level control network at each site. The first-level control network is the site datum, which establishes no fewer than three Class-E control points through GNSS static observation or traverse survey, adopts the CGCS2000 coordinate system with Gauss–Krüger projection and the 1985 National Height Datum, and controls the point-position mean error within ±5 mm. The second-level control network is the scanning control, which lays out target spheres and planar targets around and inside the building, and measures the target-center coordinates by total station (angle measurement 2″, distance measurement ±(2 + 2 ppm)) according to the attached traverse or free-stationing method, with an overall accuracy better than ±3 mm. The target layout follows three principles: the envelope principle, that the targets should enclose the scanned object rather than lie on one side, so as to reduce extrapolation errors; the layering principle, that targets should be laid out separately at different elevations such as the ground, eaves, and roof ridge; and the redundancy principle, that no fewer than four visible targets should be provided at each station, and a station whose registration residual exceeds the limit can be rejected and re-computed. This control network constitutes the common registration datum of multi-source data and is also the physical implementation of the "datum layer" in the three-layer fusion architecture.

4.3.2. Four-Step Registration and Three-Layer Fusion Architecture

In office processing, the TLS point cloud is used as the global datum, and a four-step workflow of "coarse registration–global adjustment–ICP fine registration–voxel down-sampling" is adopted, in which each step automatically judges whether it passes according to a quantitative threshold, and returns for reprocessing if it does not pass (Figure 3). In S1 coarse registration, with the target coordinates measured by the second-level control network as constraints, the rigid-body transformation of each station to the unified coordinate system is solved by least squares:

$$\min \Sigma \| Ri \cdot pij + ti - qij \|^2 \quad (2)$$

where pij is the observed coordinate of the j-th target at the i-th station, qij is its control-network coordinate, and Ri and ti are the rotation matrix and translation vector, respectively. In S2 global adjustment, all stations, targets, and control points are incorporated into the same overall adjustment model, and the station pose and target coordinates are solved simultaneously, eliminating the error accumulation caused by station-by-station transmission. This is the key to "horizontal and vertical errors both <10 mm": if station-by-station serial registration is adopted, the error will accumulate linearly or super-linearly with the number of stations and easily exceed 10 mm after dozens of stations, whereas the global adjustment makes the error grow according to the √N law. In S3, point-to-plane ICP fine registration is performed on the overlapping regions on the basis of the target-based coarse registration, minimizing $E(R,t) = \Sigma[(R \cdot pk + t - qk) \cdot nk]^2$ (where qk is the target point and nk is its normal vector). In S4, down-sampling is performed according to the voxel size v (v = 5 mm for buildings and v = 1 mm for sculptures), and statistical outlier removal (SOR) and radius filtering are used for denoising.

In view of the fusion distortion caused by precision-level differences, this paper further adopts the precision-level-aware three-layer fusion architecture of "datum–skeleton–detail". The datum layer consists of the first- and second-level control networks with an accuracy of ±2–5 mm, providing the global spatial datum and not participating in deformation. The skeleton layer consists of the TLS and photogrammetric point clouds/meshes with an accuracy of ±1–10 mm, providing the overall geometry and measurement basis of the building; the UAV aerial-triangulation results must pass the spatial-consistency check with

the TLS datum before they can participate in the fusion, otherwise they are used only for texture and visual gap-filling. The detail layer consists of the handheld high-precision scanning data with an accuracy of ≤±0.1 mm, which is superimposed onto the skeleton layer after local registration while preserving its original resolution.

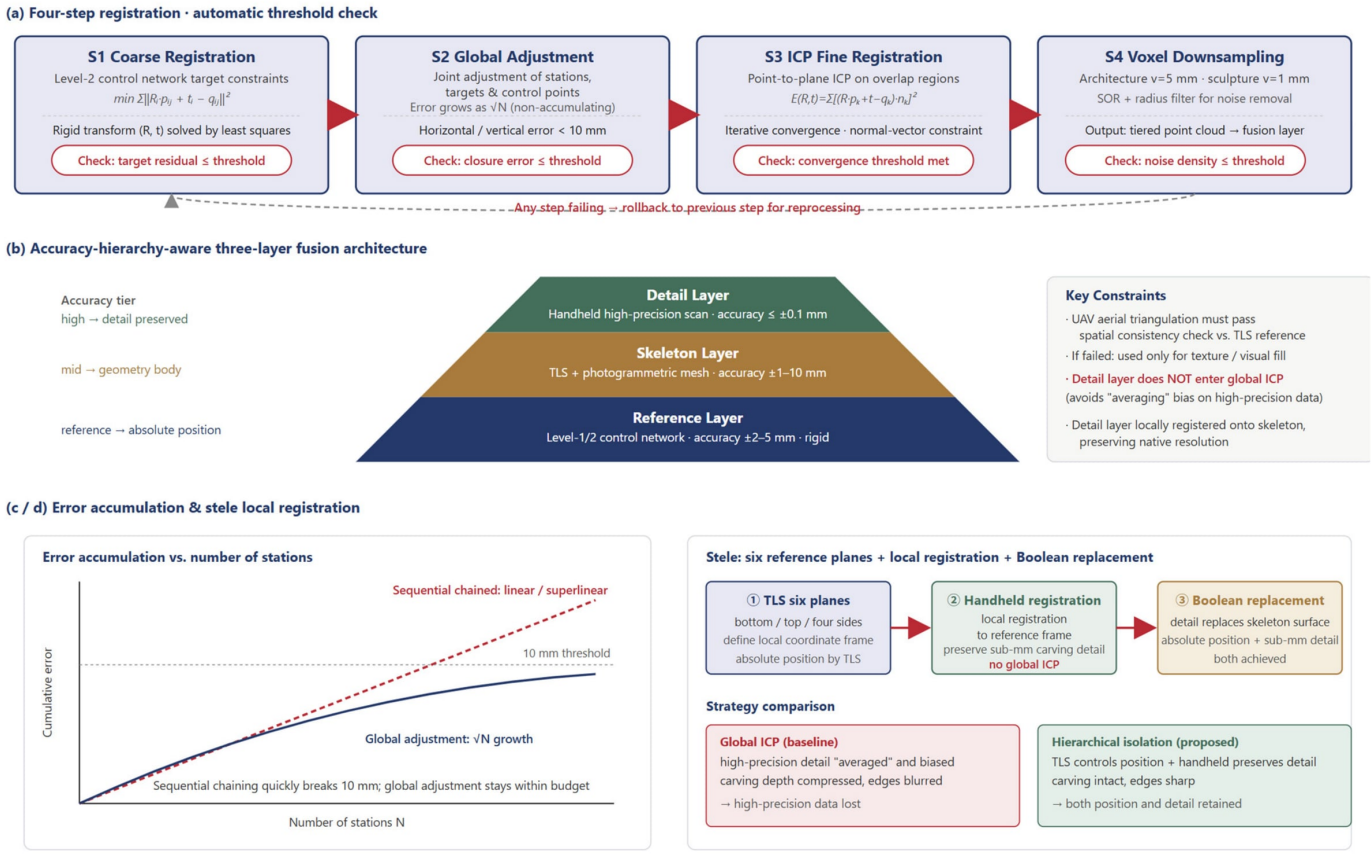


Figure 3. Multi-source data registration-fusion workflow and precision-level isolation mechanism.

The key constraint of this architecture is that the detail layer does not participate in the global ICP. In implementation, for objects such as inscriptions, the six datum surfaces of the stele (bottom, top, and four sides) are first determined by the TLS point cloud, then the handheld scanning data are registered to the local coordinate system composed of these datum surfaces, and finally the detail layer replaces the surface of the corresponding region in the skeleton layer through Boolean operations. In this way, the absolute position of the inscription is controlled by TLS, while the sub-millimeter details of the engraved marks are preserved, avoiding the "averaging" pull-off of the high-accuracy data by the global ICP (verification is given in Section 5.2).

4.3.3. Five-Component Error-Budget Model

In order to judge whether the deliverable accuracy can meet the indicator requirements and to answer the ambiguous link of "how the instrument nominal precision is transferred into the deliverable precision" in existing workflows, this paper establishes an error-budget model. The total error is synthesized from five independent or weakly correlated components:

$$\sigma_total = \sqrt{(\sigma^2_inst + \sigma^2_geo + \sigma^2_reg + \sigma^2_model + \sigma^2_check)} \quad (3)$$

where σ_inst is the point-position error caused by the instrument ranging/angle-measurement error, σ_geo is the control-network and coordinate-transfer error, σ_reg is the point-cloud

registration error, σ_model is the modeling (surface fitting, meshing) error, and σ_check is the error of the inspection measurement itself. For building-type deliverables, if the conventional operational values σ_inst = 1.0 mm, σ_geo = 2.0 mm, σ_reg = 4.0 mm, σ_model = 3.0 mm, and σ_check = 1.5 mm are adopted, then σ_total = 5.7 mm, and taking twice the mean error as the tolerance limit (confidence level about 95%) gives 11.4 mm, which does not meet the indicator requirement of <10 mm. Accordingly, this paper compresses two dominant components in a targeted manner during implementation: through global adjustment and highly redundant target layout, the registration component σ_reg is compressed to within 3.0 mm; and through surface-fitting-residual inspection, the modeling error σ_model is controlled to within 2.0 mm.

The assumption that the five components are independent or weakly correlated is justified by the physical separation of their sources: σ_inst originates in the instrument ranging and angle-encoding errors, σ_geo in the control network and coordinate transfer, σ_reg in the station registration transformation, σ_model in surface fitting and meshing, and σ_check in the inspection measurement itself. The five sources involve different instruments, different observation epochs, and different error-propagation mechanisms, so their cross-correlations are practically weak; under this condition, the root-sum-square synthesis is a conservative upper-bound approximation, and residual correlations would reduce rather than inflate the synthesized value.

$$\sigma_\text{total} = \sqrt{(1.0^2 + 2.0^2 + 3.0^2 + 2.0^2 + 1.5^2)} = 4.5 \text{ mm} \quad (4)$$

that is, twice the mean error is 9.0 mm, which meets the requirement of <10 mm with a margin. This process of "budget—not satisfied—targeted compression—satisfied again" is exactly the engineering value of the error-budget model: it decomposes the abstract accuracy indicator into operational parameters that can be controlled item by item. For sculpture-and-inscription-type deliverables, because the handheld high-precision modality is introduced, σ_inst is reduced to 0.1 mm and σ_reg to 0.3 mm, and the total mean error can be controlled at the order of 0.5 mm, which is far superior to the requirement of "point-spacing relative accuracy ≤5 mm for Buddha-statue scanning".

4.4. Semantics-Driven Component-Based Modeling and Automatic Production of PBR-LOD Assets

4.4.1. Scene Modeling and Three Semantics-Driven Component Modeling Strategies

Scene-level modeling takes the dense-matching point cloud of oblique photogrammetry as input, and generates the terrain, group-layout, and natural-environment models through filtering (progressive-encrypted-TIN ground/non-ground separation), thinning, and TIN/Poisson mesh construction; the octree depth is set according to the target-detail levels (11–12 for the main building body and 13–14 for detail parts) [37]. Hole repair distinguishes true holes (door-and-window openings are not repaired) from data-missing holes (repaired by surface continuation), and non-manifold-edge cleaning and normal unification are completed. The texture is obtained from multi-view images through view selection and seam-balancing fusion.

Component-level modeling uses the fused point cloud as the geometric reference, semantically splits the building according to the seven major categories in Table 1, and automatically matches three modeling strategies by the semantic category, realizing the automatic linkage of "category determination–strategy selection–quality constraint". Strategy A (scan-direct meshing, applicable to C2, C5) directly meshes morphologically complex components that are difficult to parameterize (ridge beasts, tile ends, inscriptions, statues, stone and brick carvings) from the point cloud, preserving the original geometric details with the highest geometric authenticity, and these are the main body of high-precision digital assets. Strategy B (semi-parametric reconstruction, applicable to C1, C3) first fits

the axis, section contour, and key dimensions of regularly shaped components with modular relationships (columns, column bases, beam frames, bracket sets, steps, railings, and pavements), and then generates the mesh driven by parameters; such models have controllable face count, clear semantics, and are convenient for batch reuse. Bracket sets of the same form can generate a master instance and adapt to different sizes through parameter adjustment, echoing the cai-fen modular thinking of the Yingzao Fashi (Treatise on Architectural Methods). Strategy C (instantiation and group organization, applicable to C6, C7) organizes the courtyard group layout and the natural environment in the form of "component instantiation + layout relationship". The component splitting follows the principle of "independently reusable", so that components such as bracket sets, ridge beasts, and tile ends can be individually called and recombined in the asset library.

4.4.2. Intelligent Production of PBR Materials

Four types of maps, namely base color (BaseColor), normal (Normal), metallic (Metallic), and roughness (Roughness), are made for each component, in PNG format with a resolution of no less than 8192 × 8192 (single or multiple images). The base-color map takes the cross-polarized photography image as the main source, and is projected into the UV space after color calibration (color-chart constrained), white-balance unification, and illumination-nonuniformity correction; the texture synchronized by the handheld device first performs color transfer to match the photographic tones. The normal map is baked from the geometric difference between the high-poly model and the low-poly (or LOD1/LOD2) model, with the baking-ray distance set to 2–5% of the bounding-box diagonal; for stone surfaces, high-frequency normal information generated from the high-precision handheld data is additionally superimposed to enhance the engraved marks, so that the tactile sense of chisel marks and inscribed lines is maintained under the low-poly LOD. The roughness map is inverted from the specular component of the polarized images; the specular-reflection intensity is estimated from the ratio of the two groups of images with and without the polarizing plate, and is then mapped to the $[0, 1]$ value domain according to an empirical model; for special surfaces such as painted patterns and gold foil, manual regional annotation and local correction are supplemented. The metallic map is generated as a binary or near-binary map by manual annotation for the limited number of gold-foil, copper-tile, and iron components, and the metallic value of all other materials is set to zero.

The UV-unwrapping rule is that different material regions of the same object are placed in different UV layers (layer numbers starting from zero), and seams are preferentially placed at geometric edges and visually occluded regions to avoid crossing the main viewing surfaces. The texture pixel density (texel density, TD) is defined as

$$\mathrm{TD} = \mathrm{T_res}/\mathrm{L} \tag{5}$$

where T_res is the map resolution (pixel) and L is the actual physical size (m) corresponding to the UV layer. The minimum TD is back-calculated according to the object Rmin:

$$\mathrm{TD} \geq \mathrm{ns}/\mathrm{Rmin} \tag{6}$$

For example, when the inscription Rmin = 0.5 mm, TD ≥ 4000–8000 px/m; if the actual size of the stele surface is 2 m × 1.5 m, a single 8192 × 8192 map (8192/2 = 4096 px/m) only reaches the lower limit, so multiple 8K maps are used for partitioned baking rather than simply raising the resolution of a single map. This rule converts the format-level requirement of "8K map" into computable shooting and baking parameters, which is another embodiment of the connection between the acquisition link and the asset-production link. According to the glTF 2.0 specification, the metallic and roughness maps are merged into a single

map and occupy the B and G channels, respectively (the R channel is reserved for ambient occlusion AO):

$$\mathrm{ORM(R, G, B)} = \mathrm{(AO, Roughness, Metallic)} \quad (7)$$

The ORM merging is automatically executed during glTF export. Both the merged ORM map and the unmerged independent metallic and roughness maps are submitted to meet the dual-channel delivery requirement. The maps are named according to the specification of [TextureFunction_SequenceNumberOfThisType_UVLayer] (for example, BaseColor_0001_UV0.png, Normal_0001_UV1.png).

4.4.3. Automatic Generation of Hierarchical LOD

According to the three-level LOD requirement, LOD0 has 100% of the faces, LOD1 has 50–60% of the faces, and LOD2 has 20–30% of the faces. LOD0 is the "master" of the asset. The modeler manually completes the mesh cleaning (removing floating facets, repairing non-manifold edges), UV optimization, normal- and tangent-space correction, and map baking; the wiring is dominated by quads and the topology is regular, and its geometric accuracy is exactly the object assessed by the acceptance indicators. LOD1/LOD2 adopts the quadric error metrics (QEM) [38] edge-collapse simplification, and the collapse cost is

$$\mathrm{Cost(v)} = \Sigma\ \mathrm{viT}\cdot(\mathrm{Qp} + \mathrm{Qq})\cdot\mathrm{vi} \quad (8)$$

where Qp and Qq are the quadric-error matrices of the two vertices at the two ends of the collapsed edge, and vi is the vertex after collapse. Three constraints are imposed during the simplification: feature-edge protection, which imposes a high-cost weight on edges with a dihedral angle greater than the threshold (30°) to prevent building edge lines and bracket-set contours from disappearing during simplification; boundary locking, which fixes the model boundary and non-manifold-edge vertices so that they do not participate in the collapse; and UV-seam protection, which densifies the weight of vertices at UV seams to prevent the simplification from destroying the texture mapping. After the automatic simplification, the modeler manually corrects it, focusing on ensuring the continuity of the roof-tile ridges, the recognizability of the bracket-set hierarchy, the identifiability of the ridge-beast contour, and the readability of the inscription characters; LOD2 allows blurred characters but does not allow the character structure to disappear.

4.4.4. Standardized Organization of Assets

All assets uniformly use meters as the unit and are unified into the coordinate system of the whole-site control network; they are organized, classified, and annotated according to the Eeasten standard to form a hierarchical asset catalog. Each model carries a four-segment identifier of category code, site code, building code, and component serial number, and is accompanied by metadata such as the site, acquisition date, acquisition device and modality, accuracy level, material type, and LOD level, supporting multi-dimensional retrieval by the platform according to "component type + era + material + region". The models are delivered in both FBX and glTF formats, each containing three-level LOD and the complete set of four-category PBR maps, and the naming-specification compliance rate is an acceptance and delivery item. The hierarchical relationships and automation nodes of the above "point cloud–mesh–component–PBR asset" chain are summarized in Figure 4.

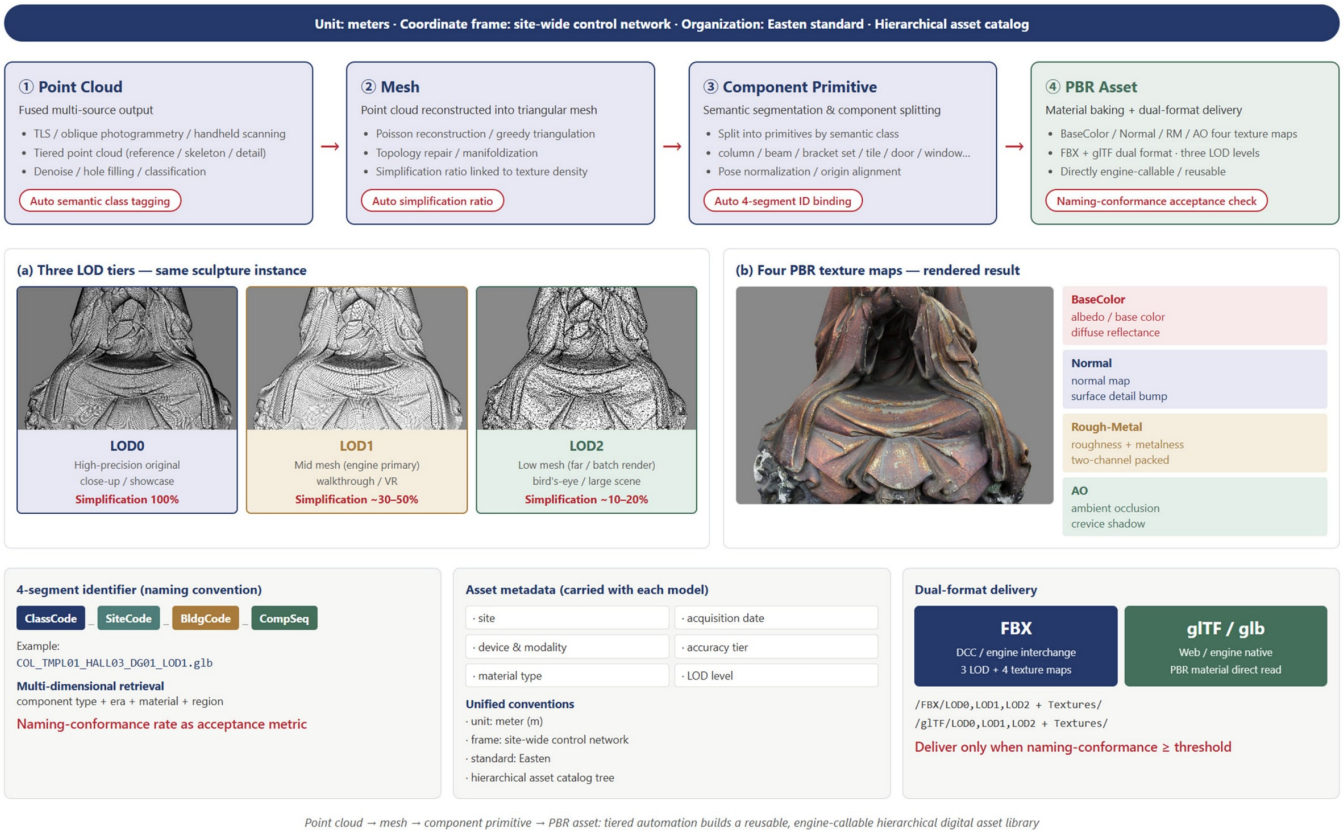


Figure 4. Intelligent construction pipeline of 3D digital assets (point cloud → mesh → component → PBR asset).

5. Results and Accuracy Verification

5.1. Full-Process Quality-Control Mechanism

This paper establishes a quality-control and accuracy-verification mechanism that runs through the whole process of acquisition, fusion, modeling, and delivery, serving as the execution basis for the automatic determination and rollback of each link.

Quality control is the closed-loop link of the intelligent workflow. This paper establishes a four-level automatic inspection mechanism, QC1–QC4, so that accuracy control runs through the whole process of acquisition, fusion, modeling, and delivery. QC1 is the field acquisition inspection: after each scan station is completed, the point-cloud completeness and target-recognition quality are checked immediately, and the station is re-scanned on site when the target residual exceeds the limit (>3 mm); after the UAV flight, the image overlap and the GCP residual are checked. QC2 is the post-registration inspection: the registration accuracy is evaluated with independent check points that do not participate in the registration (8–15 per site), and the point-position mean error is calculated

$$\mathrm{RMSE} = \sqrt{(\Sigma \|\Delta pi\|^2 / n)} \quad (9)$$

and the horizontal (XY) and vertical (Z) errors are counted separately, corresponding to the indicator that "horizontal and vertical errors are both <10 mm". QC3 is the model-accuracy inspection, which uses two methods in combination: the scale-bar/check-target method and the point-cloud-to-model deviation method. The former lays out scale bars and check targets of known length at the site, measures them on the final model, and compares them with the total-station measured values; the latter computes the nearest distance from the model surface to the original point cloud and counts the mean, standard deviation, and 95% quantile of the deviation distribution. QC4 is the texture-quality inspection, which

checks the map resolution (≥8192 × 8192), naming compliance, UV-layer completeness, and seam visibility for each model, as well as the color deviation in the base-color map, targeting a ΔE00 within five with the color chart as the reference.

The thresholds of the four-level inspection are all taken from the technical-indicator system in Table 2, and the result of each level directly determines whether the next step is executed or rolled back: if the field inspection fails, supplemental acquisition is performed on site; if the registration inspection fails, the control constraints are adjusted and the adjustment is re-run; and if the model-and-texture inspection fails, the workflow returns to the modeling and baking links. In this way, accuracy control changes from traditional post hoc sampling inspection to full-process quantifiable auditing, and also gives each decision model in this paper a verifiable closed loop.

5.2. Verification of Fusion Accuracy and Error Budget

The fusion accuracy adopts the joint adjustment workflow of "target constraint + ICP", and the deliverable accuracy is evaluated with independent check points that do not participate in the registration (eight to fifteen per scene), with the point-position mean error calculated according to Equation (9). The mean error of the target-registration residuals is 1.0–2.6 mm, which drops to 0.6–1.8 mm after ICP fine registration; the overall point-position mean error of the independent check points is 1.5–5.4 mm, of which the horizontal is 0.9–3.6 mm and the vertical 1.2–4.1 mm, which are all significantly superior to the <10 mm requirement. The internal registration RMS of the handheld-scanned components does not exceed 0.1 mm (Spider average 0.05 mm, Artec EVA average 0.09 mm), indicating that the accuracy of the detail layer is not weakened in the cross-modality registration.

The statistical calibers need to be clearly distinguished: the average of the per-site mean errors is 2.58 mm horizontally and 2.93 mm vertically, reflecting the average control level of each site; the mean error summarized over all check points is 3.1 mm horizontally and 3.6 mm vertically (maximum 8.3/9.1 mm), reflecting the overall accuracy of the whole survey area. In terms of the error budget, with σ_inst taking 1.0 mm, σ_geo 2.0 mm, σ_reg 3.0 mm, σ_model 2.0 mm, and σ_check 1.5 mm, the budget value of σ_total synthesized according to $\sqrt{(\Sigma\sigma^2)}$ is 4.5 mm; the measured components and synthesized value are 0.9, 1.9, 3.1, 2.0, 1.4, and 4.5 mm, respectively, and the deviation between the measured mean of the total mean error and the budget value is within 1%, indicating that the error-budget model has reliable predictive capability and that the registration and modeling components after targeted compression are indeed the dominant links of the deliverable accuracy. The inspection results and the measured-vs-budget comparison of the error budget for each site are shown in Table 5.

The measured results also allow an accuracy comparison among the modalities in Table 4. The sites dominated by handheld detail scanning (Sites G and H) achieved check-point 3D RMSE of 1.5–1.8 mm, whereas the sites dominated by TLS and UAV fusion achieved 3.6–5.4 mm (Table 5); within all sites, the target-registration residuals decreased from 1.0 to 2.6 mm to 0.6–1.8 mm after ICP fine registration; and the internal registration RMS of handheld scanning (≤0.1 mm) is one order of magnitude finer than the global fusion accuracy. This stratification is consistent with the design intent of the decision model: the global absolute accuracy is governed by the TLS datum and the control network, while the fine modality contributes local detail fidelity (95% quantile 0.21–0.36 mm for sculpture-and-inscription models, Table 6) rather than global absolute accuracy.

Table 5. (A) Integrated table of per-site registration accuracy and measured error-budget comparison, (B) Measured-vs-budget comparison of the error budget (building-type deliverables, unit: mm).

(**A**)

Site	Target-Registration Residual RMSE/mm	Post-ICP Residual RMSE/mm	Check-Point Planimetric RMSE/mm	Check-Point Elevation RMSE/mm	Check-Point 3D RMSE/mm
Site A	2.1	1.4	2.6	3.1	4.0
Site B	2.4	1.6	3.0	3.4	4.5
Site C	2.6	1.8	3.4	3.9	5.2
Site D	1.8	1.2	2.4	2.8	3.6
Site E	2.5	1.7	3.6	4.1	5.4
Site F	2.3	1.5	2.9	3.2	4.3
Site G	1.0	0.6	0.9	1.2	1.5
Site H	1.1	0.7	1.1	1.4	1.8
Site I	2.2	1.4	2.7	3.0	4.0
Site J	2.4	1.6	3.2	3.6	4.8

(**B**)

Error Component	Budget Value	Measured Mean	Remark
σ_inst Instrument error	1.0	0.9	FARO Focus S150 nominal ±1 mm
σ_geo Control and coordinate transfer	2.0	1.9	First-level network ±5 mm, second-level target better than ±3 mm
σ_reg Registration error	3.0	3.1	Per-station residual after global adjustment
σ_model Modeling error	2.0	2.0	Surface fitting and meshing residual
σ_check Inspection-measurement error	1.5	1.4	Measured
σ_total Total mean error	4.5	4.5	$\sqrt{(\Sigma\sigma^2)}$

Table 6. Geometric-fidelity verification results of LOD0 models.

Category	Sample Count	Mean Deviation/mm	Deviation Std/mm	95% Quantile/mm	Max Deviation/mm	Corresponding Indicator	Ratio to Indicator
C1 Structural components	101	2.4	1.6	5.8	9.2	≤10 mm	58%
C2 Decorative components	76	2.6	1.9	6.1	9.8	≤10 mm	61%
C3 Ground and steps	28	1.8	1.2	4.2	7.4	≤10 mm	42%
C4 Roof and eaves	56	2.2	1.5	5.4	8.9	≤10 mm	54%
C5 Attached objects (inscription)	15	0.09	0.07	0.23	0.46	≤5 mm	4.6%
C5 Attached objects (statue)	16	0.14	0.11	0.36	0.71	≤5 mm	7.2%
C5 Attached objects (stone/brick carving)	8	0.11	0.08	0.28	0.55	≤5 mm	5.6%
C5 Attached objects (painted pattern)	4	0.08	0.06	0.21	0.38	≤5 mm	4.2%
C6 Courtyard group	14	2.1	1.4	5.0	8.3	≤10 mm	50%
C7 Geographical environment	8	2.3	1.7	5.6	9.1	≤10 mm	56%
Sampling total	326	—	—	—	—	—	About 60% of total components

5.3. Model Geometric Fidelity and Component Asset Library

Table 6 gives the LOD0-model geometric-fidelity verification results. The mean deviation in the building-type models is 1.8–2.6 mm and the 95% quantile is 4.2–6.1 mm, accounting for 42–61% of the indicator; the mean deviation in the sculpture-and-inscription-type models is 0.08–0.14 mm and the 95% quantile is 0.21–0.36 mm, accounting for 4.2–7.2% of the indicator; the sampled specimens total 326, accounting for about 60% of the total components, with sufficient representativeness. In terms of detail fidelity, the inscription model acquired by Spider can clearly resolve engraved marks as shallow as 0.28 mm, with sharp stroke edges and clean bottoms, and no identifiable geometric offset is found in the overlay check with the original rubbing image.

Table 7 gives the statistics of the component-level assets and Table 8 gives the statistics of the three-level LOD. The composition of the asset library shows an asymmetric "quantity–complexity" distribution: the C1 category has the largest number (168) but the lowest average face count (86,400), because most of them are regular geometric bodies; the C5 category has a low proportion (13.1%) but the highest average face count (386,500), because it must preserve micrometer-level surface undulations. This distribution directly determines that the LOD generation strategy must be customized by category: a more aggressive simplification ratio can be adopted for the C1 category, while the feature edges of the C5 category must be strictly protected, otherwise the characters and patterns will become unrecognizable after simplification. The actual ratios of LOD1 and LOD2 are dynamically allocated according to the visual task: the regular C1 components take the lower limit of the interval (52–55%) to obtain a smaller file size, while the high-detail C5 components take the upper limit (57–59%) to ensure the recognizability of the characters and patterns; this "visual-task-oriented ratio allocation" is more consistent with the actual usage needs of the assets than a unified ratio. All 543 components complete the delivery of three-level LOD and four-category PBR maps, are uniformly organized according to the Eeasten standard, and adopt the naming specification of [TextureFunction_SequenceNumberOfThisType_UVLayer], with a naming-compliance rate of 100%; the deliverables are submitted in both FBX and glTF formats, with the unified unit of meter. So far, the three-level decision model proposed in this paper has passed the verification on the three indicators of accuracy, asset specification, and delivery standardization.

Table 7. Statistics of component-level assets.

Category	Count	Proportion	Modeling Strategy	Average Triangle Count (LOD0)	Typical Objects
C1 Structural components	168	30.9%	Strategy B (semi-parametric)	86,400	Walls, column bases, columns, beam frames, bracket sets
C2 Decorative components	126	23.2%	Strategy A (scan meshing)	214,600	Ridge beasts, tile ends, door-and-window carvings
C3 Ground and steps	47	8.7%	Strategy B	32,800	Pavements, steps, railings
C4 Roof and eaves	93	17.1%	Strategy A + B hybrid	126,200	Roof ridges, tile surfaces, eave ornaments
C5 Attached objects	71	13.1%	Strategy A	386,500	Inscriptions, statues, stone/brick carvings, painted patterns
C6 Courtyard and group layout	24	4.4%	Strategy C (instantiation)	512,000	Covered corridors, enclosing walls, courtyards
C7 Geographical and natural environment	14	2.6%	Strategy C	268,000	Terrain, trees, surrounding environment
Total	543	100%	—	—	—

Table 8. Statistics of the three-level LOD.

Level	Face-Ratio Requirement	Measured Ratio	Average Face Count	Average File Size (FBX/glTF)/MB
LOD0	100%	100%	181,000	118/74
LOD1	50–60%	52.4–58.9%	99,600	66/41
LOD2	20–30%	22.1–29.6%	47,100	31/19

The total number of LOD0 triangles is about 98 million, with an average of about 181,000 faces per component. Synthesizing Tables 6–8, the geometric fidelity and asset structure show a significant stratification across categories: the building-type and attached-object-type correspond to differentiated indicators of ≤10 mm and ≤5 mm, respectively, with sufficient measured margin; and both the accuracy margin and the LOD ratio can be automatically adapted according to the category and the visual task. The final deliverables are fed into the 3D digital-asset library of the AI-creation platform, and can support applications such as game-engine development, film virtual production, digital cultural-tourism experiences, and the permanent digital archiving of cultural heritage.

6. Conclusions

Aiming at the demand of digital-content creation platforms for high-fidelity Chinese ancient-architecture digital assets, and at the problems of "large scale span, heterogeneous precision levels, inconsistent deliverable standards, and insufficient workflow intelligence" of ancient-architecture heritage, this paper proposes an intelligent multi-source heterogeneous 3D digitization decision-making method with Rmin and precision levels as decision variables, and completes a full-process experimental verification with ten ancient-architecture and immovable scenes as subjects. The main conclusions are as follows:

(1) A six-modality intelligent selection model and four functional acquisition channels with the target identifiable minimum feature size Rmin as the decision variable are established, transforming device selection from expert experience into a deterministic decision that can be automatically derived from object features, and explicitly giving the handover conditions between modalities. The experiments show that acquisition-resource allocation is automatically adjusted with the object-scale structure: the sites dominated by building groups have the most TLS stations (Site E, 158), the sites dominated by inscriptions and statues have the most handheld-scanned components (Site H, 51), and the resource investment is structurally complementary.

(2) A precision-level-aware three-layer fusion architecture of "datum–skeleton–detail" and a four-step registration-adjustment workflow are proposed, clarifying the key constraint that the detail layer does not participate in the global registration and that its local coordinate system is defined by the TLS control surfaces, thereby solving the fusion-distortion problem of cross-precision-level data. The established five-component error-budget model is verified by experiments, and the deviation between the budget value (4.5 mm) and the measured value (4.5 mm) is within 1%, so that the accuracy indicators can be decomposed into operational parameters that can be controlled item by item.

(3) The stable attainment of high-precision deliverables is realized: the horizontal and vertical mean errors of the multi-source fused models are 3.1 mm and 3.6 mm, respectively (maximum 8.3/9.1 mm), all meeting the <10 mm requirement with sufficient margin; the 95% quantile geometric deviation is 4.2–6.1 mm for building-type models and 0.21–0.36 mm for sculpture-and-inscription-type models; and the resolution capability of inscription engraved marks reaches 0.28 mm. It should be noted that the instrument

nominal accuracy (TLS ±1 mm), the internal registration RMS of handheld scanning (≤0.1 mm), and the deliverable check-point accuracy are distinct calibers: the first two characterize local or instrument-level consistency, while the deliverable accuracy after global integration is governed by the TLS datum and the control network and is quantitatively transferred through the error-budget model.

(4) A semantics-driven component-based automatic production pipeline for PBR-LOD assets is constructed: the C1–C7 semantic categories are used as the decision entry to automatically match three modeling strategies, producing 543 component-level assets in seven categories, all equipped with four types of 8192 × 8192 PBR maps and LOD0/LOD1/LOD2 three-level models, delivered in both FBX and glTF formats with a unified naming specification and a naming-compliance rate of 100%, realizing the form leap from "digital recording" to "digital assets".

(5) A QC1–QC4 four-level full-process quality-control and accuracy-verification system is formed, which sets quantitative thresholds for key links such as the field overlap and coverage ratios, point-cloud registration, model geometric accuracy, and map specifications, and automatically inspects each item, so that accuracy control changes from "post hoc sampling inspection" to "full-process quantifiable auditing".

(6) The error-transfer mechanism of "multiple devices do not imply high accuracy" is revealed: without precision-level awareness, the averaging effect of ICP will make the high-accuracy data converge toward the low-accuracy data; only through the three mechanisms of hierarchical isolation, datum individualization, and unidirectional precision transmission can the final accuracy be determined by the finest modality rather than by the coarsest modality.

This study also has several limitations. First, the workflow automation is not complete: LOD0 master modeling, UV optimization, and the regional annotation of metallic materials still rely on manual operations, and a controlled benchmark of processing time and human intervention against conventional workflows has not yet been conducted. Second, the five error components are assumed to be independent or weakly correlated; under site conditions that produce strong correlations, the square-root synthesis may deviate from the measured value. Third, the roughness inversion adopts an empirical mapping model, and the material inversion of anisotropic painted surfaces remains difficult. Fourth, the ten experimental sites are concentrated in one regional and typological spectrum, and the generalization to other climate conditions, material systems, and larger group scales requires further verification.

In the future, four directions can be further deepened: first, introducing variable-angle reflectance measurement to improve the credibility of the material inversion of painted-pattern types; second, establishing a "multi-phase scanning–change detection" mechanism to incorporate the deformation monitoring of the main body into the routine workflow; third, for AIGC applications, exploring a path to automatically generate stylized variant assets based on PBR materials and semantic metadata under the premise of preserving geometric authenticity; and fourth, introducing deep-learning-based point-cloud–mesh semantic segmentation, automatic decimation-quality evaluation, and intelligent material inversion to further reduce manual cost and push the production of ancient-architecture digital assets toward scale and assembly line.

Author Contributions: Methodology, H.H. (Hongzhen Hou) and P.D.; software, H.H. (Hongzhen Hou) and P.D.; validation, P.D. and H.H. (Hongzhou Hou); formal analysis, H.H. (Hongzhen Hou); investigation, H.H. (Hongzhou Hou); resources, H.H. (Hongzhou Hou); data curation, P.D.; writing—original draft, H.H. (Hongzhen Hou); writing—review and editing, P.D., M.Z. and H.H. (Hongzhou Hou); visualization, P.D.; supervision, M.Z.; project administration, M.Z.; funding acquisition, M.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author.

Conflicts of Interest: Author Hongzhen Hou was employed by the company Siwei Shijing Technology (Beijing) Co., Ltd. Author Hongzhou Hou was employed by the company Weituo Times Architectural Design Co., Ltd. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Levoy, M.; Pulli, K.; Curless, B.; Rusinkiewicz, S.; Koller, D.; Pereira, L.; Ginzton, M.; Anderson, S.; Davis, J.; Ginsberg, J.; et al. The Digital Michelangelo Project: 3D scanning of large statues. In *SIGGRAPH '00: Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques*; ACM: New York, NY, USA, 2000; pp. 131–144.
2. Ikeuchi, K.; Oishi, T.; Takamatsu, J.; Sagawa, R.; Nakazawa, A.; Kurazume, R.; Nishino, K.; Kamakura, M.; Okamoto, Y. The Great Buddha Project: Digitally archiving, restoring, and analyzing cultural heritage objects. *Int. J. Comput. Vis.* **2007**, *75*, 189–208. [CrossRef]
3. Stylianidis, E.; Remondino, F. *3D Recording, Documentation and Management of Cultural Heritage*; Whittles Publishing: Dunbeath, UK, 2016.
4. Tang, P.; Huber, D.; Akinci, B.; Lipman, R.; Lytle, A. Automatic reconstruction of as-built building information models from laser-scanned point clouds: A review of related techniques. *Autom. Constr.* **2010**, *19*, 829–843. [CrossRef]
5. Remondino, F. Heritage recording and 3D modeling with photogrammetry and 3D scanning. *Remote Sens.* **2011**, *3*, 1104–1138. [CrossRef]
6. Yastikli, N. Documentation of cultural heritage using digital photogrammetry and laser scanning. *J. Cult. Herit.* **2007**, *8*, 423–427. [CrossRef]
7. Besl, P.J.; Mckay, N.D. A method for registration of 3-D shapes. *IEEE Trans. Pattern Anal. Mach. Intell.* **1992**, *14*, 239–256. [CrossRef]
8. Grussenmeyer, P.; Alby, E.; Landes, T.; Koehl, M.; Guillemin, S.; Hullo, J.F.; Assali, P.; Smigiel, E. Accurate and Detailed Survey of Heritage Objects at Different Scales: From Requirements to Implementation. In *XXII ISPRS Congress*; ISPRS: Amsterdam, The Netherlands, 2012.
9. Grilli, E.; Remondino, F. Machine learning generalisation across different 3D architectural heritage. *ISPRS Int. J. Geo-Inf.* **2020**, *9*, 379. [CrossRef]
10. Schönberger, J.L.; Frahm, J.M. Structure-from-motion Revisited. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*; IEEE: Piscataway, NJ, USA, 2016; pp. 4104–4113.
11. Furukawa, Y.; PONCE. Accurate, dense, and robust multiview stereopsis. *IEEE Trans. Pattern Anal. Mach. Intell.* **2010**, *32*, 1362–1376. [CrossRef] [PubMed]
12. Westoby, M.; Brasington, J.; Glasser, N.F.; Hambrey, M.J.; Reynolds, J.M. Structure-from-Motion' photogrammetry: A low-cost, effective tool for geoscience applications. *Geomorphology* **2012**, *179*, 300–314. [CrossRef]
13. Nex, F.; Remondino, F. UAV for 3D mapping applications: A review. *Appl. Geomat.* **2014**, *6*, 1–15. [CrossRef]
14. Murtiyoso, A.; Grussenmeyer, P. Documentation of heritage buildings using close-range UAV images: Dense matching issues, comparison and case studies. *Photogramm. Rec.* **2017**, *32*, 206–229. [CrossRef]
15. Sapirstein, P. Accurate measurement with photogrammetry at large sites. *J. Archaeol. Sci.* **2016**, *66*, 137–145. [CrossRef]
16. Remondino, F.; Rizzi, A. Reality-based 3D documentation of natural and cultural heritage sites—Techniques, problems, and examples. *Appl. Geomat.* **2010**, *2*, 85–100. [CrossRef]
17. Farella, E.M.; Torresani, A.; Remondino, F. Quality features for the integration of terrestrial laser scanner and UAV images. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2016**, *41*, 285–291.
18. Luhmann, T.; Robson, S.; Kyle, S.; Boehm, J. *Close-Range Photogrammetry and 3D Imaging*, 2nd ed.; De Gruyter: Berlin, Germany, 2014.
19. Nocerino, E.; Menna, F.; Remondino, F. Investigation of indoor and outdoor performance of two portable mobile mapping systems. *Proc. SPIE 2017, 10332* **2017**, *10332*, 125–139. [CrossRef]
20. Earl, G.; Basford, P.; Bischoff, A.; Bowman, A.R.; Crowther, C.; Dahl, J.; Hodgson, M.; Isaksen, L.; Kotoula, E.; Martinez, K.; et al. Reflectance transformation imaging systems for ancient documentary artefacts. *Proc. EVA Lond.* **2011**, *2011*, 147–154.
21. Zhang, Z. Iterative point matching for registration of free-form curves and surfaces. *Int. J. Comput. Vis.* **1994**, *13*, 119–152. [CrossRef]
22. Rusinkiewicz, S.; Levoy, M. Efficient Variants of the ICP Algorithm. In *Third International Conference on 3-D Digital Imaging and Modeling (3DIM)*; IEEE: Piscataway, NJ, USA, 2001; pp. 145–152.

23. Rodríguez-Gonzálvez, P.; Fernández-Palacios, B.J.; Muñoz-Nieto, Á.L.; Arias-Sanchez, P.; Gonzalez-Aguilera, D. Mobile LiDAR system: New possibilities for the documentation and dissemination of large cultural heritage sites. *Remote Sens.* **2017**, *9*, 189. [CrossRef]
24. Croce, V.; Caroti, G.; De Luca, L.; Jacquot, K.; Piemonte, A.; Véron, P. From the semantic point cloud to heritage-building information modeling: A semiautomatic approach exploiting machine learning. *Remote Sens.* **2021**, *13*, 461. [CrossRef]
25. Luebke, D.; Reddy, M.; Cohen, J.D.; Varshney, A.; Watson, B.; Huebner, R. *Level of Detail for 3D Graphics*; Morgan Kaufmann: San Francisco, CA, USA, 2003.
26. Biljecki, F.; Ledoux, H.; STOTER. An improved LOD specification for 3D building models. *Comput. Environ. Urban Syst.* **2016**, *59*, 25–37. [CrossRef]
27. Biljecki, F.; Ledoux, H.; Stoter, J.; Zhao, J. Formalisation of the level of detail in 3D city modelling. *Comput. Environ. Urban Syst.* **2014**, *48*, 1–15. [CrossRef]
28. Murphy, M.; McGOVERN, E.; Pavia, S. Historic building information modelling—Adding intelligence to laser and image based surveys of European classical architecture. *ISPRS J. Photogramm. Remote Sens.* **2013**, *76*, 89–102. [CrossRef]
29. Barazzetti, L.; Banfi, F.; Brumana, R.; Previtali, M. Creation of parametric BIM objects from point clouds using NURBS. *Photogramm. Rec.* **2015**, *30*, 339–362. [CrossRef]
30. Burley, B. Physically Based Shading at Disney. In *ACM SIGGRAPH 2012 Course Notes: Practical Physically Based Shading in Film and Game Production*; ACM: New York, NY, USA, 2012.
31. Khronos Group. *GlTF 2.0 Specification*; Khronos Group: Beaverton, OR, USA, 2022. Available online: https://www.khronos.org/gltf/ (accessed on 15 October 2026).
32. Dellepiane, M.; Callieri, M.; Ponchio, F.; Scopigno, R. Mapping highly detailed colour information on extremely dense 3D models: The case of David. *Comput. Graph. Forum* **2008**, *27*, 2178–2187. [CrossRef]
33. Mcdermott, W. *The PBR Guide*; Allegorithmic: Montreal, QC, Canada, 2018.
34. Scopigno, R.; Callieri, M.; Cignoni, P.; Corsini, M.; Dellepiane, M.; Ponchio, F.; Ranzuglia, G. 3D models for cultural heritage: Beyond plain documentation. *Int. J. Herit. Digit. Era* **2012**, *1*, 33–42.
35. Champion, E.; Rahaman, H. 3D digital heritage models as sustainable scholarly resources. *Sustainability* **2019**, *11*, 2425. [CrossRef]
36. Münster, S. Digital 3D technologies for humanities research and education: An overview. *Appl. Sci.* **2022**, *12*, 2426. [CrossRef]
37. Kazhdan, M.; Hoppe, H. Screened Poisson surface reconstruction. *ACM Trans. Graph.* **2013**, *32*, 29. [CrossRef]
38. Garland, M.; Heckbert, P.S. Surface Simplification Using Quadric Error Metrics. In *24th Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH '97)*; ACM: New York, NY, USA, 1997; pp. 209–216.

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