

Special Section on 3DOR2026

SHREC 2026: Reconstruction of high-frequency geometry in synthetic heritage[☆]

Cristián Llull^{a,b},^{*}, Ivan Sipiran^{a,b,c}, Francisca Gil^c, Nelson Baloian^a, Anthony Belessis^g, Vlassis Fotis^g, George Ioannakis^f, Efthymios Koukoulis^g, Konstantinos Moustakas^g, Sotiris Papatheodorou^g, Ioannis Pratikakis^{e,f}, Ioannis Romanelis^g, Kriti Singh^{d,1}, Eleftheria Solomonidi^g, Christoforos Vlachos^g, Vaia Zachari^{e,f}

^a Department of Computer Science, University of Chile, Beauchef 851, Santiago, Chile

^b Shape Vision Lab, Santiago, Chile,

^c CENIA, Santiago, Chile

^d Amazon, Seattle, WA, United States

^e Democritus University of Thrace, Greece

^f ATHENA Research Center, Greece

^g Visualization and Virtual Reality Group, Department of Electrical & Computer Engineering, University of Patras, Greece

ARTICLE INFO

Keywords:

3D shape retrieval
3D reconstruction
Cultural heritage

ABSTRACT

The geometric evaluation of 3D reconstruction techniques, spanning from classical structure-from-motion (SfM) to contemporary neural implicit representations and Gaussian splatting, remains a challenge. In cultural heritage applications, the preservation of high-frequency surface details, such as tool marks, inscriptions and reliefs, is often of greater significance than global shape approximation. Traditional metrics, notably the Chamfer Distance (CD), are frequently inadequate as their reliance on global error averaging masks the loss of critical micro-topography. To address this, in this Shape Retrieval Challenge (SHREC) track we introduce a novel benchmarking framework comprising: (1) a controlled, procedurally generated synthetic dataset of Khachkar-inspired stone reliefs, fully reproducible; and (2) the Feature-Aware Chamfer Distance (FACD), a feature-aware distance function designed to prioritize structurally and significantly geometric features. Three teams each submitted one run, and the organizers contributed five baseline reconstructions, for a total of eight configurations evaluated under a common protocol. Our evaluation reveals that standard CD frequently misranks models by favoring global smoothness over detail preservation. By providing more archaeologically plausible assessments, our FACD reveals a persistent fidelity gap, showing that current state-of-the-art methods still lack the nuance required for feature optimization.

1. Introduction

The digitization of cultural heritage represents a critical race against physical degradation and environmental loss. High-fidelity “digital twins” serve as indispensable records for non-invasive archaeological analysis, restoration planning, and virtual museology. However, the

democratization of passive, image-based 3D capture has revealed a significant “fidelity gap”: while recent advances in Neural Radiance Fields [1] (NeRF), 3D Gaussian Splatting [2] (3DGS), and transformer-based architectures [3,4] have revolutionized photometric visual realism, geometric topological precision has frequently lagged behind active-sensor technologies (e.g., laser scanners).

[☆] This article is part of a Special issue entitled: ‘CAG_SS_3DOR2026’ published in Computers & Graphics.

^{*} Corresponding author at: Department of Computer Science, University of Chile, Beauchef 851, Santiago, Chile.

[@cllullt](https://twitter.com/cllullt), [Linkedin: cristian-llull-torres](https://www.linkedin.com/in/cristian-llull-torres/) (C. Llull).

E-mail addresses: clull@dcc.uchile.cl (C. Llull), isipiran@dcc.uchile.cl (I. Sipiran), francisca.gil@cenia.cl (F. Gil), nbaloian@dcc.uchile.cl (N. Baloian), ambelessis@ece.upatras.gr (A. Belessis), vfotis@ece.upatras.gr (V. Fotis), ekoukoulis@ece.upatras.gr (E. Koukoulis), moustakas@ece.upatras.gr (K. Moustakas), s.papatheodorou@ac.upatras.gr (S. Papatheodorou), ipratika@ee.duth.gr (I. Pratikakis), iroman@ece.upatras.gr (I. Romanelis), krsingn@amazon.com (K. Singh), up1092837@ac.upatras.gr (E. Solomonidi), chris.vlachos@ece.upatras.gr (C. Vlachos), vzachari@ee.duth.gr (V. Zachari).

URLs: <https://www.cllullt.cl> (C. Llull), <https://www.ivan-sipiran.com> (I. Sipiran).

¹ Work done independently, not affiliated with or endorsed by Amazon.

This fidelity gap is exacerbated by a foundational flaw in how 3D reconstruction algorithms are evaluated. Standard geometric evaluation relies heavily on global spatial averages, such as the Chamfer Distance (CD), which inherently penalize outlier noise while heavily favoring smooth, interpolated surfaces. Consequently, algorithms are optimized for global topological completeness at the expense of preserving semantically critical, high-frequency topography. Models that produce visually pleasing but geometrically “melted” meshes systematically outrank those that successfully capture intricate micro-topographies. This bias is catastrophic for heritage artifacts — such as inscribed steles or Khachkars — where the engravings and archaeological value are encoded precisely within these fine-scale reliefs.

This bias motivates our central research question: how can we quantitatively decouple the preservation of structural micro-geometry from global shape approximation to accurately benchmark 3D reconstruction models in heritage contexts.

To address this, we formalize the challenge as the SHREC 2026 track on the retrieval of high-frequency geometry in synthetic heritage, establishing a coherent evaluation ecosystem. Organized since 2006 in conjunction with the Eurographics Symposium on 3D Object Retrieval, the 3D Shape Retrieval Challenge (SHREC) evaluates the effectiveness of 3D-shape retrieval and analysis algorithms on common benchmarks. This track followed the standard SHREC protocol: the organizers issued a call for participation, released the rendered image collection together with the ground-truth camera poses, and set a submission deadline for the reconstructed models. Each participating team contributed a description of its methods. Three teams each submitted one run, and the organizers contributed five baseline reconstructions using standard pipelines and recent transformer-based variants; all were evaluated centrally under a common protocol. The submissions are not ranked to determine a winner—they provide a diverse pool of reconstructions against which the ordering induced by CD and FACD can be compared. This ecosystem comprises three contributions:

1. **Feature-Aware Distance function:** We propose the Feature-Aware Chamfer Distance (FACD), a novel evaluation criterion that computes vertex-specific criticality weights directly from local differential geometry (curvedness). By shifting from global averaging to intrinsic curvature-based weighting, the FACD mathematically penalizes the detail-erasure bias of traditional evaluation criteria.
2. **Controlled Micro-Topography Dataset:** We introduce a fully reproducible, procedurally generated synthetic dataset of Khachkar-inspired stone reliefs. By providing perfect ground truth independent of instrumental scanner noise, this dataset isolates algorithmic geometric extraction from real-world photometric confounders.
3. **Comprehensive Algorithmic Benchmark:** We evaluate eight diverse reconstruction configurations across our dataset. Our analysis disentangles the variables of evaluation design and algorithmic capability, showing that standard CD systematically misranks methods. We demonstrate that while traditional explicit pipelines (e.g., COLMAP) exhibit severe fragility (up to an 11.1% failure rate on homogeneous stone), they often preserve sharp features better than highly-ranked implicit neural surfaces, a nuance captured exclusively by our FACD.

2. Related work

2.1. Evaluation benchmarks and datasets

The foundation of quantitative 3D reconstruction relies on benchmarks that provide high-fidelity ground truth. However, as reconstruction methods have evolved, significant structural gaps have emerged regarding their applicability to cultural heritage.

Early geometric-fidelity gold standards, such as DTU [5] and Tanks & Temples [6], rely on high-precision laser-scanned ground truths. Other traditional benchmarks leverage terrestrial LiDAR [7] or Structured Light Scanners [8] across macro environments (e.g., ETH3D [9]). While rigorous, these datasets primarily consist of industrial, architectural, or everyday objects and thus fail to capture the specific complexities of untextured materials and the high-frequency geometries inherent to cultural artifacts. Conversely, synthetic datasets like BlendedMVS [10] offer perfect ground-truth geometry at scale, yet they lack realistic surface weathering, erosion patterns, and complex physical materiality.

Heritage-specific benchmarks have recently begun to emerge but remain limited in topological scope. Collaborative efforts such as NERFBK [11] and BaliMask3D [12] provide museum-quality objects with challenging sub-surface or reflective materials, yet they lack the varied, stress-testing geometries necessary to evaluate micro-relief structural variations. Broader initiatives like Culture3D [13] focus predominantly on large-scale landmark documentation via aerial drone imagery, leaving micro-level tool markings, inscriptions, and high-frequency reliefs as an open challenge. The current absence of a standardized benchmark that simultaneously combines precise geometric ground truth, heritage-relevant micro-topographies and controlled material homogeneity actively hinders the targeted development of high-fidelity preservation algorithms.

2.2. Geometric evaluation

The standard evaluation of reconstruction quality currently bifurcates into photometric (image-based) and geometric (spatial) assessment. While photometric quality is assessed via PSNR [14], SSIM [15] and LPIPS [16], geometric quality compares extracted 3D meshes or point clouds against a ground truth, relying almost exclusively on the Chamfer Distance (CD) [17,18] and the F-Score [6].

Chamfer Distance remains the dominant geometric evaluation criterion due to its differentiability and computational efficiency. It computes the average nearest-neighbor distance between a reconstructed source point cloud S and a target ground-truth point cloud T :

$$CD = \frac{1}{2|S|} \sum_{x \in S} \min_{y \in T} \|x - y\|_2 + \frac{1}{2|T|} \sum_{y \in T} \min_{x \in S} \|x - y\|_2 \quad (1)$$

F-Score addresses global outlier sensitivity by serving as the harmonic mean of precision (P) and recall (R) calculated for a specific spatial distance threshold d :

$$P(d) = \frac{100}{|T|} \sum_{t \in T} [\min_{s \in S} \|t - s\| < d] \quad (2)$$

$$R(d) = \frac{100}{|S|} \sum_{s \in S} [\min_{t \in T} \|s - t\| < d]$$

$$F(d) = \frac{2P(d)R(d)}{P(d) + R(d)} \quad (3)$$

While the F-Score excellently quantifies regional and global spatial completeness, both CD and F-Score remain fundamentally feature-agnostic. They treat a point lying on an archaeologically critical inscription with the identical statistical weight as an uninformative point situated on a flat, featureless support base.

Recent variations have attempted to resolve specific spatial distribution biases. The Density-Aware Chamfer Distance (DCD) [19] corrects for non-uniform point sampling densities, while the Normal-guided Chamfer Distance (NCD) [20] incorporates local surface orientation vectors to better capture topological flow. Newer paradigms, such as the Flexible-weighted Chamfer Distance (FCD) [21] and the Asymmetric Contrastive Chamfer Loss (ACCL) [22], aim to decouple local edge precision from global structural integrity. Nevertheless, none of these existing metrics derive semantic importance directly from the intrinsic differential geometry of the object. This critical gap underscores the need for our proposed FACD, which explicitly embeds localized, feature-aware weighting into the geometric evaluation architecture.

2.3. Reconstruction paradigms

Reconstruction methodologies evaluated in our benchmark span the three dominant eras of 3D vision: explicit geometric pipelines, implicit learnable representations, and feed-forward transformers.

Explicit Structure-from-Motion (SfM) pipelines, exemplified by COLMAP [23], established the early benchmarks for 3D reconstruction accuracy [24]. However, due to their strict reliance on explicit handcrafted feature tracking [25], these frameworks suffer severe performance degradation on textureless or homogeneous heritage stone.

Conversely, optimization-based techniques utilize coordinate-based neural field representations, such as NeRF [1] and its variants [26–28], alongside Neural Surface Representations (NeuS) [29] and its multi-view surface derivatives [30–32]. These methods frequently embed multiresolution hash-grids [33] to map coordinates to volume density or signed distance functions (SDF). While highly robust to untextured regions, they carry an inherent interpolation bias.

Explicit optimization pipelines utilizing 3D Gaussian Splatting (3DGS) [2] and its surface-constrained extensions (e.g., SuGaR [34], Sparse2DGS [35], MeshSplatting [36]) attempt to jointly refine geometry and photorealistic appearance rapidly, though mesh extraction remains noisy.

Finally, feed-forward learning approaches, spearheaded by DUST3R [3] and VGGT [4], along with their scalable multi-view successors (e.g., MAST3R [37], Fast3R [38], FastVGGT [39]), directly regress 3D coordinate maps from input imagery in a single pass. While unprecedented in speed, these models often rely heavily on structural priors and training data distributions (e.g., [40,41]), risking the “hallucination” of geometry that smooths over unique features—a phenomenon our benchmark is specifically designed to quantify.

3. Methodology

To determine whether a novel evaluation criterion can accurately retrieve geometric information, a robust dataset with zero instrumental bias is required. In cultural heritage, the scientific value of an artifact often lies in its high-frequency surface details — such as tool marks, erosion patterns and reliefs — rather than its low-frequency global shape. To systematically evaluate this “fidelity gap,” we developed a fully reproducible benchmarking framework using synthetic ground truth, which guarantees absolute precision and scalability while allowing us to leverage known data distributions to explicitly evaluate local topological features.

The evaluation task requires participants to reconstruct complex 3D objects from a set of **90 high-quality 2D renderings per object**, simulating a realistic photogrammetry capture setup with physically based materials and dual-sun high-dynamic-range (HDRI) lighting. Researchers were permitted to employ any 3D reconstruction paradigm (e.g., Structure-from-Motion, Neural Radiance Fields, 3D Gaussian Splatting) with the primary objective of maximizing the geometric fidelity of the reconstructed micro-topography.

3.1. Dataset construction

The dataset consists of procedurally sculpted meshes derived from canonical 3D base forms, featuring a static, approximated diffuse stone material (high roughness, low metallicity). For each object, participants are provided with:

- **Images:** 90 PNG images (1920 × 1080 resolution) rendered from a helical trajectory.
- **Camera Poses:** Full camera intrinsics and extrinsics formatted for standard COLMAP pipelines.

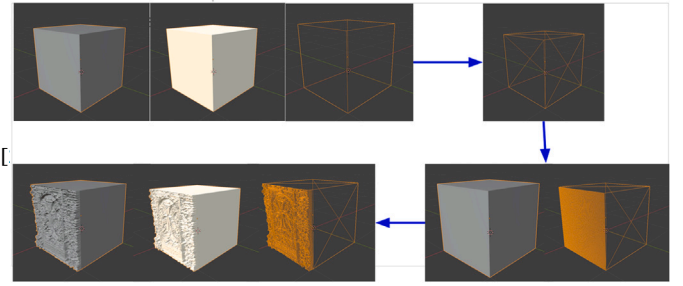


Fig. 1. Complete steps for generating the base cube. First: create a mesh. Second: for each vertex, apply a height based on the original image’s luminosity information.

3.1.1. 3D model generation

We present a controlled synthetic dataset inspired by the intricate stone reliefs found in Armenian Khachkars (decorative memorial steles), which are inscribed on UNESCO’s List of Intangible Cultural Heritage of Humanity.

To generate these reliefs without the biases of real-world scanner noise, we bypass the ill-posed problem of pure image-to-geometry generation [42]. Instead, we extract disparity/depth maps from front-facing photographs of real Khachkars by converting the image luminance into displacement values—a valid approximation given that shallow carvings illuminated by direct sunlight exhibit strong shading-to-depth correlations.

The procedure begins by scaling a grayscale Khachkar image to match the intended mesh vertex resolution (e.g., 300 × 300 pixels for a 300 × 300 grid). The ground truth is then procedurally sculpted in Blender [43], similar to the methods proposed in [44]. For each vertex on the target face, we apply a displacement along its normal vector proportional to the pixel’s luminosity multiplied by a base displacement amplitude factor (strength $s = 0.02$). See Fig. 1 for the sculpting process visualization.

Based on the degree of structural complexity, the dataset is stratified into two subsets: *Primitives* and *Derivatives*.

Primitives. To introduce geometric variety, the base mesh is morphed into four distinct aspect ratios: the canonical *Normal* cube, a *Large* version (2× higher than wide), a *Wide* version (2× wider than high), and a *Realistic* version mimicking the proportions of an actual stele. For the *Realistic* version, the displacement amplitude is scaled by a factor of 5 to preserve the prominence of the carved features despite the altered dimensions.

Derivatives. To rigorously test the reconstruction paradigms across complex global geometries, we apply non-linear morphing to the base shapes, yielding 6 distinct geometric derivatives: Cone, Cylinder, Egg, Package, Pyramid, and Sphere. This expansion yields a total of 48 structurally complex 3D objects per base image, enabling extensive stress testing of the robustness of surface extraction.

To evaluate algorithmic robustness against untextured, homogeneous regions, the flat surfaces of the remaining four visible faces may optionally be sculpted with textures from real Khachkar back surfaces, randomly selected from a pool to enhance realism, as shown in Fig. 2. This distinction leads to two classifications of the objects:

- **all:** The main face is sculpted with the primary relief, while the other visible faces are displaced using a secondary “rock noise” texture sampled from the back of a real Khachkar.
- **one face:** Only the main face is sculpted; the remaining faces are left perfectly flat and textureless.

In total, a single source image generates eight distinct Primitive configurations (4 aspect ratios × 2 texture variants) and six Derivatives for

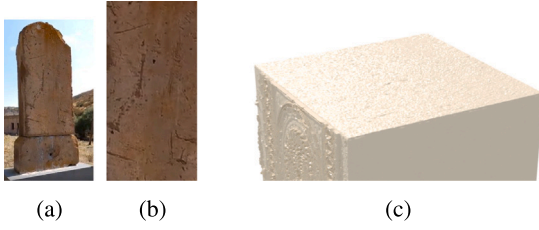


Fig. 2. Apply the back texture (a) of the Khachkar (b) to the ground truth flat faces. In the image, it appears on top of the object in (c).

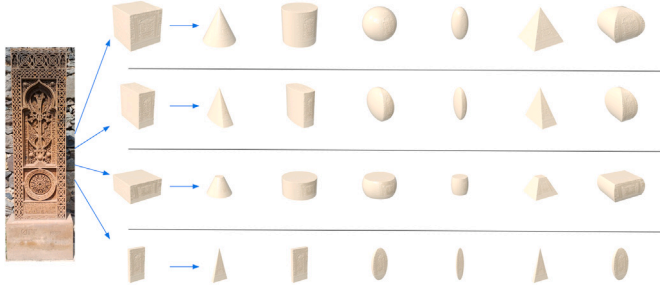


Fig. 3. Half of the objects generated from a single image. First column: the four Primitives from the image at the left. Next to each primitive: its Derivatives, in order: Cone, Cylinder, Sphere, Egg, Pyramid and Package.

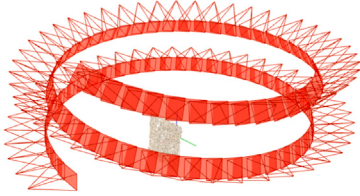


Fig. 4. 90 cameras around the object visualized in COLMAP.

each Primitive, thus obtaining 48 3D objects from a single image with this procedure. See Fig. 3. These distinctions (image ID, all/one_face, Primitive and Derivative) are further used to identify objects in the dataset.

3.1.2. Rendering generation

Camera poses were algorithmically generated in Blender. Inspired by the capture methodology of NeRF-BK [11], 90 cameras are distributed along a dual-level helical path encircling the object, providing overlapping coverage from both high and low elevation angles (see Fig. 4). The scenes were rendered using the CYCLES path-tracing engine to provide physically accurate global illumination, shadows and realistic stone reflectance.

Each pose is then rendered using Blender. In Fig. 5 is an example. The scene is modeled with two suns and an HDRI background to achieve more realism. However, a white background is left around the object. We use the CYCLES rendering engine to provide realistic illumination.

3.2. Feature-aware geometric evaluation

Standard geometric metrics, such as the conventional Chamfer Distance (CD), treat every reconstructed point equally. This global averaging inherently biases evaluations toward methods that produce smooth, interpolated surfaces while masking the catastrophic loss of high-frequency micro-geometry (e.g., carvings, sharp corners). To resolve this, we introduce the **Feature-Aware Chamfer Distance**

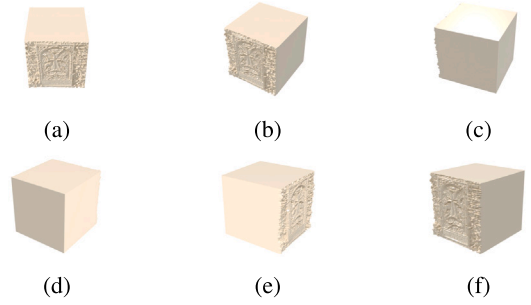


Fig. 5. Sample of six frames taken of the object.

(FACD), which assigns a continuous, feature-aware weight to each ground-truth vertex based on its structural and engraving-preservation significance.

3.2.1. Intrinsic curvature-based weighting

The synthetic dataset permits the utilization of manually-assigned heuristic weights to each vertex. However, such methodology is not scalable and introduces subjective bias. Therefore, we present a novel, fully automated methodology to compute intrinsic feature weights directly from the ground-truth differential geometry.

We quantify local surface saliency using the Root Mean Square (RMS) curvature, or *curvedness* (C). First, a geometric purification step is applied to the mesh to remove non-manifold artifacts, degenerate faces (area $< 10^{-8}$), and orphaned vertices, ensuring mathematical stability. For every valid vertex i , we compute the principal curvatures k_1 and k_2 using the ‘libigl’ [45] library. The raw curvedness is defined as:

$$C_i = \sqrt{\frac{1}{2} (k_1^2 + k_2^2)} \quad (4)$$

To distinguish deliberate carvings from background stone noise, we isolate the “active” curvatures ($C_i > 10^{-5}$) and compute their 90th percentile, C_{upper} . We then apply a smart clamping and normalization strategy to map the raw curvatures to a normalized scale $\hat{C}_i \in [0, 1]$:

$$\hat{C}_i = \frac{\min(C_i, C_{\text{upper}}) - C_{\text{lower}}}{C_{\text{upper}} - C_{\text{lower}}} \quad (5)$$

where $C_{\text{lower}} = 0$. Finally, to actively penalize the smoothing of sharp details, we apply a non-linear amplification ($\gamma = 2.0$) and map the values to a continuous weight scale:

$$W_i = w_{\text{base}} + w_{\text{scale}} \cdot (\hat{C}_i)^\gamma \quad (6)$$

Using $w_{\text{base}} = 1.0$ and $w_{\text{scale}} = 299.0$, the final vertex weights organically span a continuous range from 1.0 (flat background) to 300.0 (sharp carvings and critical edges). This maps geometric topology to semantic importance without human intervention. Without γ , many vertices at medium curvedness (rough stone texture, shallow relief edges) receive weights of half the scale, which may introduce noise into the comparison. With $\gamma = 2$, those same vertices compress closer to the baseline, which better reflects the categorical intent. See Fig. 9 for a colormap of the vertices weights.

The scale of 300 provides an expressive representation of the vertices’ weights. However, note that any scale can be used, since the ranking is a comparison of values on the same scale.

3.2.2. Metric formulation

Let S be the set of reconstructed source points and T be the set of ground-truth target points. Let W_i represent the calculated intrinsic weight for a target point $t \in T$.

The traditional **Chamfer Distance (CD)** is defined as the mean of the bidirectional point-to-point nearest-neighbor distances:

$$CD_{FWD} = \frac{1}{|S|} \sum_{x \in S} \min_{y \in T} \|x - y\|_2^2 \quad (7)$$

$$CD_{BWD} = \frac{1}{|T|} \sum_{y \in T} \min_{x \in S} \|x - y\|_2^2$$

$$CD = \frac{1}{2} (CD_{FWD} + CD_{BWD}) \quad (8)$$

To heavily penalize asymmetric failure modes (e.g., when an algorithm hallucinates geometry far from the surface, or fails to complete a shape), we evaluate the **Harmonic Chamfer Distance (HCD)**:

$$HCD = \frac{2 \cdot CD_{FWD} \cdot CD_{BWD}}{CD_{FWD} + CD_{BWD} + \epsilon} \quad (9)$$

and the **Harmonic Chamfer Distance variant (HCD₂)**:

$$HCD_2 = 2 \frac{(1 + CD_{FWD})(1 + CD_{BWD})}{(1 + CD_{FWD}) + (1 + CD_{BWD})} - 1 \quad (10)$$

that penalizes more when either FWD or BWD becomes 0.

Finally, our proposed **Feature-Aware Chamfer Distance (FACD)** directly integrates the intrinsic curvature weights into the backward distance calculation, forcing the evaluation to prioritize the high-frequency carvings:

$$FACD_{FWD} = \sum_{x \in S} w(x) \min_{y \in T} \|x - y\|_2^2$$

$$FACD_{BWD} = \sum_{y \in T} w(\hat{x}(y)) \min_{x \in S} \|x - y\|_2^2 \quad (11)$$

$$FACD = \frac{1}{2} (FACD_{FWD} + FACD_{BWD}) \quad (12)$$

Similarly, the **Harmonic Feature-Aware Chamfer Distance (HFACD and HFACD₂)** replaces CD_{BWD} with $FACD_{BWD}$ and CD_{FWD} with $FACD_{FWD}$ in the harmonic formulation. By contrasting these feature-aware formulations against standard distance metrics, we successfully decouple the evaluation of global shape accuracy from the preservation of high-frequency features.

3.3. Note on reconstruction alignment

To evaluate the structural accuracy of the 3D reconstructions, they must be registered to the ground-truth coordinate system. We perform this registration using a Similarity Transform (Sim3), estimated via the camera trajectories. Using camera centers as registration landmarks ensures that the subsequent geometric evaluation — specifically, the high-frequency detail analysis — is not biased by surface-to-surface alignment optimizations such as ICP.

To decouple the spatial registration from the surface evaluation, we leverage the **camera optical centers** as corresponding, invariant spatial landmarks. We assume possessing exact, one-to-one correspondences between the estimated camera poses and their true coordinates.

To prevent localized Structure-from-Motion (SfM) tracking errors from introducing global scale or rotational bias, we implement a robust Umeyama (1991) alignment with a robust Procrustes superimposition approach grounded in Iterative Least Trimmed Squares [46]. Let $P = \{p_1, p_2, \dots, p_N\}$ and $Q = \{q_1, q_2, \dots, q_N\}$ represent the sets of paired camera optical centers from the reconstruction and the ground truth, respectively. We aim to compute an optimal Similarity Transform, Sim(3)—comprising a uniform scale $s \in \mathbb{R}^+$, rotation $R \in SO(3)$, and translation $t \in \mathbb{R}^3$ —while dynamically rejecting a predefined proportion α of severe outliers (in our experiments, $\alpha = 0.2$).

The iterative trimming procedure is defined as follows:

1. **Initial Transformation:** We compute the initial Sim(3) parameters $\{s, R, t\}$ using the closed-form Umeyama [47] algorithm on the full set of corresponding pairs.
2. **Residual Evaluation:** We compute the Euclidean residual for each pair: $e_i = \|q_i - (sRp_i + t)\|_2$.

3. **Greedy Outlier Rejection:** The pair (p_k, q_k) associated with the maximum residual e_k is classified as an outlier and removed from the active sets P and Q .
4. **Iterative Refinement:** Steps 1 through 3 are repeated until the targeted subset size $N' = \lfloor (1 - \alpha)N \rfloor$ is reached.

By removing the point with the highest residual and strictly re-estimating the transformation matrix at each step, we ensure that extreme outliers do not artificially shift the alignment.

4. Evaluated reconstruction methods

The evaluation encompasses eight distinct reconstruction configurations: three submitted by participating teams and five contributed by the organizer as baselines. Together they span the current spectrum of 3D vision approaches, from explicit photogrammetry to implicit neural rendering, Gaussian splatting and feed-forward transformer architectures:

- **SAMplify_SuGaR** (Team 1): Integrates SAM2 [48] segmentation with surface-aligned Gaussian Splatting (SuGaR) [34].
- **VGGT_frame** (Team 2): A feed-forward transformer model (VGGT [4]) optimized for efficiency using a deterministic 16-frame subset.
- **NeRF_InstantNGP** (Team 3): Utilizes NeRF [1] with multiresolution hash encoding [33] and Poisson surface reconstruction.
- **COLMAP** [23] (baseline): The classical explicit photogrammetric baseline.
- **NeuS** [29] (baseline): The SDF-based implicit neural surface baseline.
- **VGGT Variants** (baselines): Three distinct full-image configurations of VGGT [4] with varying point-count constraints and input resolutions (vggt_100k, vggt, and vggt_high_def).

4.1. Team 1 - SAMplify_SuGaR (*sam_sugar*)

https://github.com/TEXTailES/SAMplify_SuGaR

Team 1 reconstructs each object through a four-stage pipeline that integrates promptable foreground segmentation, Structure-from-Motion (SfM), and surface-aligned Gaussian splatting [2] into a unified background-free workflow (see Fig. 6).

In the first stage, foreground regions are isolated using Segment Anything 2 (SAM2) [48]. An operator provides initial positive and negative prompt points on a reference image, which SAM2 propagates across the remaining views to generate consistent binary masks. In the second stage, these masked images are processed with COLMAP [23] to estimate camera poses and produce a sparse initialization cloud. By restricting feature matching to the object region, this method suppresses unreliable background correspondences, yielding highly stable camera pose estimation.

In the third stage, the sparse reconstruction initializes a 3D Gaussian Splatting representation, optimized under the SuGaR [34] regularizer. This encourages the Gaussians to remain concentrated on the object surface and align their principal axes with the underlying geometry, significantly reducing floaters. Finally, a triangle mesh is extracted via Poisson reconstruction over the SuGaR density/SDF estimates, decimated to a target face count, and textured by projecting the refined Gaussian radiance onto the mesh surface.

4.2. Team 2 - VGGT with frame selection (*vggt_frame*)

<https://www.kaggle.com/code/kritisi19/vggt-shrec>

Team 2 employs the Visual Geometry Grounded Transformer (VGGT) [4] for feed-forward 3D reconstruction. Although the full set of 90 views per object is available, using all frames proved computationally infeasible in our setting. Specifically, experiments with FastVGGT [39] on the full image set yielded overlapping and duplicated geometric

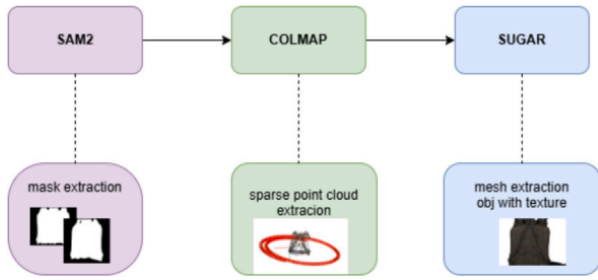


Fig. 6. SAM_SuGaR pipeline, integrating SAM2 masks with COLMAP initialization and SuGaR regularization.

Table 1

Time and resources used by the models to reconstruct the objects. This is obtained using our parameters, optimized for greater speed than in the original paper.

Method	VRAM	Time	Hardware
COLMAP	4 GB	100 min	Intel i7-12700H RTX A2000 (8 GB)
NeuS	8 GB	10 h	Intel i7-9700F RTX 2080 (8 GB)
VGGT	37 GB	30 s	AMD Ryzen 7 5700X RTX A6000 (48 GB)

structures in the reconstructed outputs. Additionally, running VGGT on all frames resulted in out-of-memory (OOM) errors on the available hardware. To circumvent this, the method deterministically selects a fixed subset of 16 frames, evenly distributed across the acquisition sequence.

{1, 5, 7, 11, 14, 17, 22, 24, 35, 57, 59, 61, 71, 77, 82, 87}

This deterministic selection reduces memory consumption while maintaining multi-view coverage across the objects.

Input images are processed at their original 1920×1080 resolution in a single forward pass. VGGT predicts per-pixel depth maps, confidence values and camera pose encodings, which are translated into standard 3×4 transformation matrices consistent with the COLMAP convention. When available, predicted world-space 3D points are used directly; otherwise, depth maps are unprojected into 3D space.

To refine the reconstruction, a global confidence threshold is computed at the 90th percentile, retaining only the highest-certainty points. Additionally, a brightness-based filtering heuristic is applied (removing points with an RGB sum ≥ 740) to eliminate background noise. The implementation was optimized to run on a single NVIDIA T4 GPU.

4.3. Team 3: NeRF with InstantNGP (*nerf_instantngp*)

Team 3 proposes an optimization-based methodology that utilizes Neural Radiance Fields (NeRFs). The pipeline consists of training a NeRF model, extracting a dense point cloud via raycasting and generating a triangle mesh using Poisson surface reconstruction [49] (see Fig. 7).

The architecture leverages Instant-NGP [33], using a multiresolution hash-grid encoding to efficiently map spatial coordinates while preserving high-frequency detail (Fig. 8). The implementation uses $L = 16$ levels with $F = 2$ feature dimensions per level, with grid resolutions ranging from 16 to 2048. The density MLP employs a single hidden layer of 64 neurons, followed by a view-dependent RGB head. Hierarchical sampling is guided by two proposal networks.

For point cloud extraction, an arbitrary number of camera rays are sampled, and depth is determined using a median-based estimator (the depth at which accumulated weight reaches 50%). Points with low accumulated weight are discarded. Before Poisson reconstruction, the point cloud undergoes statistical outlier removal (discarding points

> 2 standard deviations from the 20-nearest-neighbor mean). The final mesh is generated using an octree depth of 10. The average processing time per object is approximately 335 s using an NVIDIA RTX 4070 for the entire pipeline, including NeRF training (283 s.), point cloud extraction and mesh generation.

For development, the Nerfstudio [50] was employed, specifically the Nerfacto model, which is a configurable NeRF pipeline.

4.4. Organizer baselines: Established pipelines and VGGT variants

The organizers provided a comprehensive comparative baseline by evaluating three distinct reconstruction approaches — explicit photogrammetry, implicit neural surfaces, and feed-forward transformers — across five standardized configurations.

COLMAP [23] was selected to represent the classical explicit geometrical baseline. To ensure perfect alignment with the ground truth, COLMAP was provided with the exact original camera poses generated during the dataset rendering process.

NeuS [29] was evaluated as the foundational implicit neural rendering baseline utilizing Signed Distance Functions (SDF). Ground-truth cameras were provided by converting the COLMAP format into the standard npz format used by VolSDF [51]. To maintain computational feasibility across the large dataset, iterations were limited to 140,000 steps (rather than the 500,000 proposed in the original work).

VGGT Variants [4] were evaluated to test the scalability of feed-forward architectures utilizing all 90 images. Unlike the other baselines, VGGT does not accept external camera poses, requiring a post-processing algorithm for rigid mesh alignment. The organizers evaluated three variations:

- **vgt_100k**: Limits the output to 100,000 points with an input image resolution of 1024.
- **vgt**: Processes inputs at 1024 resolution with an effectively unlimited point count (capped at 5 million).
- **vgt_high_def**: Processes inputs at the original 1920 resolution, constrained to 1 million points to ensure dense surface coverage.

A summary of the hardware requirements and inference times for these baselines is provided in Table 1.

5. Results and discussions

The novel automatic Feature-Aware weighting pipeline was successfully applied across the entire dataset. Fig. 9 visualizes this application across three different geometries. The generated color maps confirm the correct identification of semantically relevant features, assigning baseline weights to flat background surfaces (darker regions) while strongly amplifying the weights of the object's structure and high-frequency sculpted surfaces (brighter regions).

Although participants were provided with the entire dataset of 616 objects, the computational overhead of full reconstruction varied widely across methods. Consequently, a representative subset of 280 objects was evaluated across all configurations. NeuS was the only exception, reconstructing a subset of 242 objects due to severe processing time limits. The evaluated dataset comprises:

- **Primitives Subset**: Complete reconstruction of all primitive topological variants across all configurations.
- **Derivatives Subset**: Complete reconstruction of all derivative geometries corresponding to base images ID 1, 3, and 10, including models with no starting image (10_flat, featuring completely flat surfaces) and those with only texturized faces (3_all_texture).

In total, evaluating this subset required over 140 cumulative days of continuous computational time. This highlights the severe resource constraints of modern optimization-heavy pipelines. NeuS alone accounted for approximately 100 days of processing.

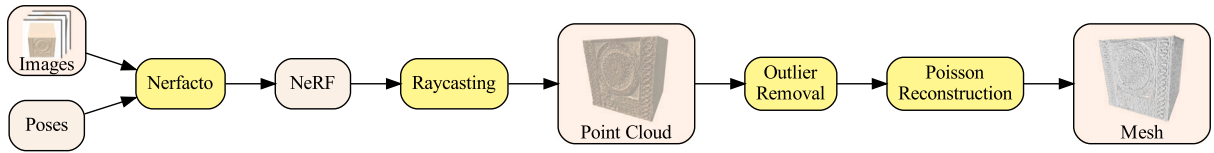


Fig. 7. Block diagram of the NeRF-based extraction pipeline proposed by Team 3.

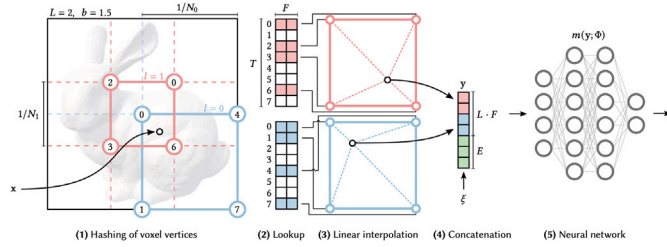


Fig. 8. Multiresolution hash grid encoding used in Instant-NGP. Source: Figure adapted from [33].

Table 2

Reconstruction methods performance by CD and FACD. Lower is better. In bold: the minimum; the * indicates the maximum. Truncated to 4 decimals for clarity.

Method	CD		FACD	
	Mean	Median	Mean	Median
colmap	0.0055	0.0014	0.0742	0.0043
nerf_instantngp	0.0003	0.0001	0.0050	0.0007
neus	0.0102	0.0010	0.8378	0.0102
sam_sugar	0.4865*	0.0013	73.1297*	0.0901
vggt	0.0057	0.0013	0.5679	0.1121
vggt_100k	0.0055	0.0014	0.5768	0.1227
vggt_frame	0.1244	0.0062*	3.1470	0.3387*
vggt_high_def	0.0449	0.0015	1.0419	0.1258

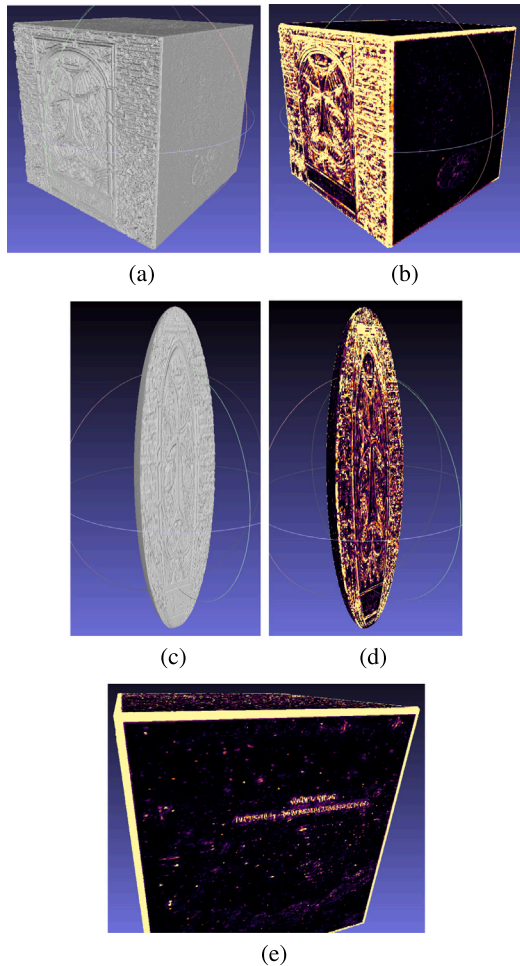


Fig. 9. Visualizing the Feature-Aware weighting pipeline of relief 1, normal and realistic egg derivative. (a) and (c) The original Khachkar geometry. (b) and (d) The final continuous semantic weight map mapped to the mesh, where flat background rock receives a baseline weight of 1.0 (black), and significantly high-frequency carvings receive amplified weights up to 300.0 (yellow). (e) Feature-Aware text on the side of the relief 3_all_texture.

Before assessing fine geometric precision, we quantified the baseline reliability of the reconstruction paradigms. A stark contrast in robustness emerged (Table 2 and Fig. 10). The hash-encoded neural radiance field (nerf_instantngp) demonstrated exceptional robustness and precision, consistently ranking first across almost all spatial criteria (CD, HCD₂, FACD, HFACD, HFACD₂). This suggests it maximizes both overall structure and feature preservation.

neus closely followed, maintaining highly stable performance (ranks 2–3). In contrast, traditional explicit feature-matching pipelines exhibited severe fragility on homogeneous stone artifacts. colmap presented an 11.1% algorithmic failure rate (Table 3), confirming that classical local-feature matching without strong geometric priors remains a bottleneck for automated heritage documentation.

The VGGT family variations generally occupied the middle-to-lower tiers of the rankings. It is crucial to note that, unlike the other methods, VGGT was not provided with initial camera poses, which consistently resulted in inaccurate global positioning of the object. Finally, while sam_sugar achieved a 0% algorithmic failure rate, returning a mesh for every input, it consistently ranked last in mean error, suggesting the presence of severe outliers.

While median aggregation mitigates the impact of statistical anomalies, it can still be skewed when a specific method fails structurally across a large subset of objects. In this study, we distinguish between a method’s “failure rate” (the inability to produce a valid reconstruction, such as colmap’s 11.1%) and its “outlier rate” (successful completions that yield catastrophic geometric noise).

To identify the true accuracy hierarchy of the algorithms when they successfully converge, we conducted a multidimensional outlier ablation. Because distances are highly right-skewed, we used Tukey’s [52, 53] rule based on the interquartile range (IQR) with a 3 factor to objectively define a threshold for catastrophic failure, filtering out any reconstruction with FACD > 0.916.

Table 3 details the impact of these catastrophic outliers on method evaluation. While sam_sugar never failed to run, the IQR threshold reveals that 26.1% of its completions were catastrophic noise explosions. Conversely, when colmap did manage to converge, it produced highly reliable geometries (only 1.1% outliers). Crucially, when these unsuccessful reconstructions are removed from the dataset, the global CD rankings shift significantly. Standard CD artificially penalizes robust methods (such as neus and vggt) when outliers are present and shifts their rankings when noise is removed. In stark contrast, the rankings derived from the proposed FACD distance remain unchanged.

Reconstruction Method	Rankings: Whole Dataset (Lower is better)		Rankings: Primitives Only		Rankings: Derivatives Only	
	CD	FACD	CD	FACD	CD	FACD
colmap	7	2	8	3	3	2
nerf_instantngp	1	1	1	1	1	1
neus	2	3	5	2	2	3
sam_sugar	6	6	6	8	4	4
vggt	3	4	2	4	5	5
vggt_100k	4	5	3	5	6	6
vggt_frame	8	8	7	7	8	8
vggt_high_def	5	7	4	6	7	7

Fig. 10. Ranking for each method reconstruction, evaluating reconstructed objects by all teams. Each row corresponds to a reconstruction method. The numbers correspond to the rank given by the criterion of that column. Colors correspond to the rank number for visualization only. Squares from left to right: (1) all datasets reconstructed by all the objects; (2) primitives subset; (3) derivatives subset of object with IDs 1, 3 and 10.

Table 3

Robustness analysis and rank shifts after removing catastrophic outliers (FACD > 0.916). *Fail (%)* indicates algorithms that returned no valid data. *Outlier (%)* indicates successful runs that resulted in catastrophic geometric noise. Rank shifts indicate movement after noise removal (positive means improved rank).

Method	Fail (%)	Outlier (%)	Rank shift						
			CD	HCD	HCD ₂	FACD	HFACD	HFACD ₂	
colmap	11.1	1.1	1	0	1	0	0	0	
nerf_instantngp	0.0	0.0	0	0	0	0	0	0	
neus	3.7	2.5	-1	0	-1	0	0	0	
sam_sugar	0.0	26.1	1	0	1	0	0	0	
vggt	0.4	7.9	-1	0	-1	0	0	0	
vggt_100k	0.4	8.2	-1	0	-1	0	0	0	
vggt_frame	0.0	16.1	0	0	0	0	0	0	
vggt_high_def	0.0	8.9	1	0	1	0	0	0	

This suggests that FACD inherently provides a highly stable evaluation hierarchy that naturally resists catastrophic variance.

Table 4 provides examples of objects. As can be seen, not all evaluation criteria yield the same rankings, revealing a gap in correctly ranking objects with criteria that do not measure the features. An interesting case is 5_one_face_wide, as the ranking dramatically changes between CD (vggt better) and FACD (nerf_instantngp better). In fact, CD classifies colmap as the worst, while FACD leaves colmap as the second best. Images have been shown to align qualitatively better with the FACD.

A critical finding from isolating the Primitive subset is the extreme divergence in how traditional photogrammetry (COLMAP) is penalized by different evaluation criteria. On basic Primitive shapes, COLMAP ranks the worst (rank 8) under standard Chamfer Distance (CD), yet jumps to rank 2 under Weighted Chamfer Distance (FACD) (see Fig. 10). This indicates that while COLMAP likely struggles with scattered outlier noise far from the true surface (heavily penalized by CD), the underlying density and structural core of its reconstructions remain highly accurate (rewarded by FACD). Interestingly, on the more complex Derivative subset, COLMAP's performance stabilizes across evaluations, jumping to rank 3 in CD and rank 1 in HCD.

Our rank correlation analysis using Kendall's τ and Spearman's ρ identified the specific topological configurations that cause evaluation breakdowns. The criteria showed perfect agreement ($\tau = 1.0$) on globally symmetric, fully visible objects such as all_wide_sphere, all_wide_egg, and one_face_cone. Conversely, the most severe evaluation discordance occurred on the one_face_large experiment ($\tau \approx 0.066$, $\rho \approx 0.085$). Objects characterized by large, single planar faces appear to exacerbate the differing sensitivities of CD (outlier-sensitive) versus FACD. When methods attempt to reconstruct these large, flat regions, incomplete filling or sparse point sampling can cause massive rank

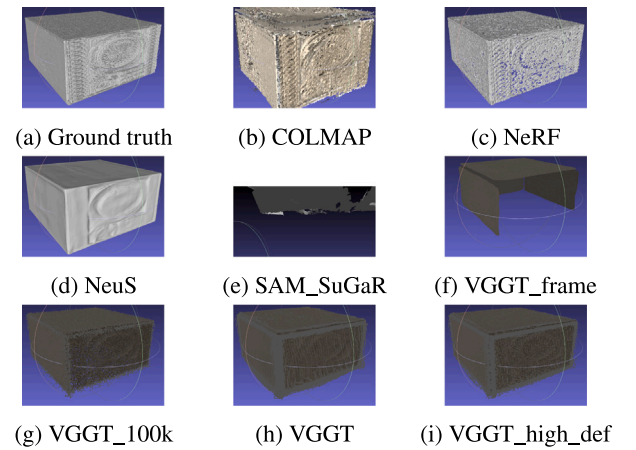


Fig. 11. Captures of the 8 reconstructions of one of the objects where CD and FACD agreed the most, i.e., had the same ranking (object 10_all_wide). To the naked eye, it is simpler to rank than in Table 4 example 2.

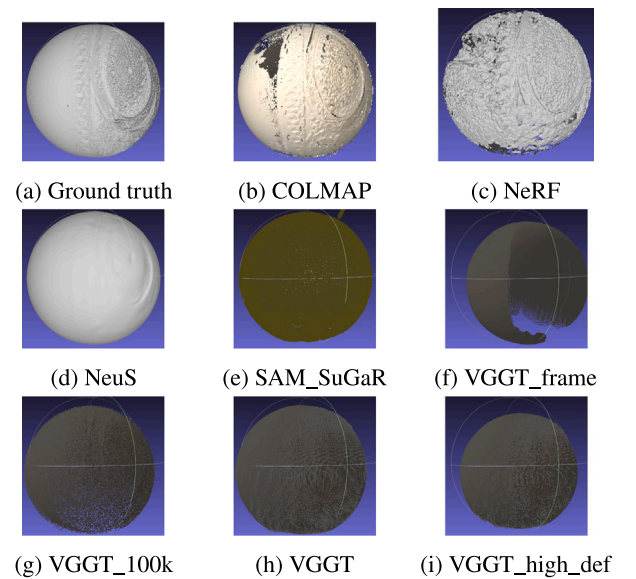
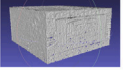
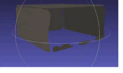
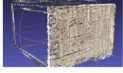
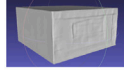
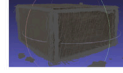
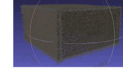
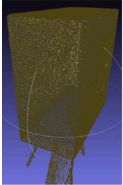
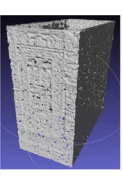
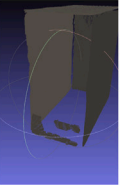
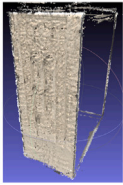
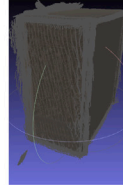
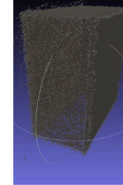
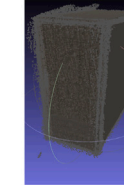
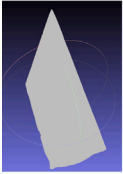
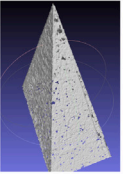
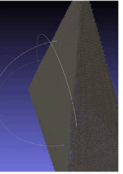
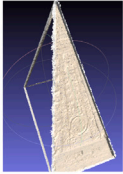
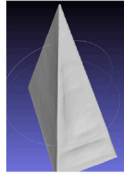
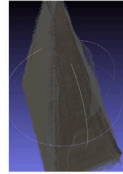
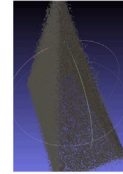
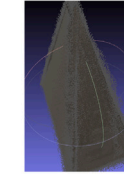
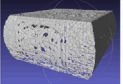
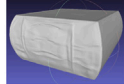
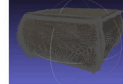
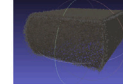
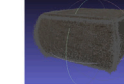


Fig. 12. Captures of the 8 reconstructions of one of the objects where CD and FACD agreed the most, i.e., had the same ranking (object 10_all_sphere).

Table 4

Qualitative and quantitative comparison of reconstruction methods across three representative objects. For each criterion, the best performing method is **bolded**, the second best is underlined, and the worst performing method is marked with an asterisk (*). Lower values indicate better geometric fidelity.

Metric	sam_sugar	nerf_instantngp	vggt_frame	colmap	neus	vggt	vggt_100k	vggt_high_def
Example 1: 5_one_face_wide (High-Frequency Relief)								
Mesh								
CD	0.001107	0.001536	0.004310	0.008255*	0.004984	0.000483	<u>0.000584</u>	0.001570
FACD	0.151228	0.004841	0.189191*	<u>0.013523</u>	0.015887	0.033173	0.040463	0.031262
HFACD	0.010094	<u>0.002234</u>	0.013780	0.012663	0.001208	0.028819	0.039635*	0.024355
HFACD ₂	0.132688	0.004828	0.161285*	<u>0.013512</u>	0.015657	0.033033	0.040431	0.031053
Example 2: 6_one_face_large (High-Frequency Relief)								
Mesh								
CD	0.008046	0.000868	0.004428	0.013582*	0.000950	0.000621	<u>0.000723</u>	0.000751
FACD	0.137071	0.001656	0.199842*	0.041618	<u>0.006523</u>	0.052972	0.059380	0.060753
HFACD	0.001896	<u>0.001590</u>	0.015621	0.036873	0.001262	0.047115	0.057467*	0.055099
HFACD ₂	0.120776	0.001656	0.169159*	0.041428	<u>0.006489</u>	0.052678	0.059273	0.060429
Example 3: 3_one_face_large_pyramid (Complex Topology & Untextured Stone)								
Mesh								
CD	0.003957	0.000045	0.011078*	0.002745	0.001544	0.001131	0.001188	<u>0.000501</u>
FACD	0.163908	0.000206	0.370843*	<u>0.003935</u>	0.011992	0.050859	0.052775	0.025580
HFACD	0.032765	0.000173	0.333738*	<u>0.003221</u>	0.003676	0.007058	0.015222	0.012872
HFACD ₂	0.145440	0.000206	0.360805*	<u>0.003932</u>	0.011894	0.048739	0.050892	0.025263
Example 4: 1_all_wide_package (Straight and Curved Edges)								
Mesh								
CD	0.002684	0.000078	0.004088*	<u>0.000161</u>	0.002916	0.001500	0.001723	0.001582
FACD	0.777684*	0.000988	0.451766	<u>0.002468</u>	0.009531	0.210643	0.235414	0.220522
HFACD	0.003638	0.000973	0.411524*	0.002380	<u>0.001749</u>	0.178404	0.221590	0.188629
HFACD ₂	0.439062	0.000988	0.439244*	<u>0.002468</u>	0.009458	0.205034	0.232780	0.214760

inversions between the point-to-point and density-weighted evaluation paradigms.

As demonstrated in Table 5, analyzing the dataset by topological class reveals the profound degree to which standard geometric evaluation masks semantic failure. We utilize the FACD/CD Ratio to explicitly quantify this discrepancy.

A clear morphological pattern emerges among the most challenging object types. The top five hardest topologies are uniformly “flat” variants (e.g., flat_realistic_egg, flat_sphere). Because these objects lack macroscopic structural complexity and homogeneous stone

textures provide few features for explicit tracking, reconstruction algorithms frequently fail to converge on the true surface, often producing smoothed, blob-like geometries (see Figs. 11 and 12).

Crucially, the standard CD fails to capture the severity of these catastrophic failures. For instance, on the flat_sphere, algorithms achieved a seemingly acceptable median CD of 0.01844. However, the automated curvature weighting in our FACD unmasks the complete loss of micro-topographic integrity, yielding a value of 5.08860—a staggering 275.9x divergence. This ratio illustrates the “smoothing bias”: models minimize spatial point-to-point distances by averaging

Table 5

Analysis of Object Difficulty. We determine the difficulty by the median *FACD* value across all methods. Note the extreme divergence between the standard Chamfer Distance (*CD*) and our Feature-Aware criterion (*FACD*) in the hardest cases. Values limited to 4 decimals for better visualization.

Object type	Median CD	Median FACD	FACD/CD Ratio
<i>Top 5 Hardest Object Types (Highest FACD)</i>			
flat_realistic_egg	0.0718	18.1190	252.5x
flat_pyramid	0.1110	17.5057	157.7x
flat	0.0402	9.2104	229.2x
flat_sphere	0.0184	5.0886	275.9x
flat_egg	0.0216	3.8825	179.9x
<i>Top 5 Easiest Object Types (Lowest FACD)</i>			
realistic_all_texture_cylinder	0.0007	0.0111	16.3x
large_all_texture_pyramid	0.0007	0.0133	18.0x
realistic_all_texture	0.0006	0.0148	24.3x
one_face_realistic_cylinder	0.0004	0.0159	41.8x
one_face_realistic	0.0004	0.0165	42.3x

over the surface, effectively smoothing out the high-frequency features that FACD prioritizes.

Conversely, the “easiest” objects are characterized by prominent macrostructures (e.g., pyramid, cylinder) and high-contrast surfaces (all_texture). These features provide robust anchors for both explicit SfM matching and implicit neural priors, resulting in highly accurate geometric retrieval. In these ideal conditions, the FACD/CD ratio stabilizes at the lower end of our amplified evaluation scale (ranging from 16x to 24x), indicating that standard CD aligns with semantic reality only when the underlying topology is exceptionally well-structured and textured. Ultimately, this divergence confirms that evaluating untextured heritage surfaces with traditional global averages yields a highly misleading picture of engraving fidelity.

6. Conclusion

This paper presented the findings of the SHREC 2026 track on the retrieval and evaluation of high-frequency geometry in cultural heritage 3D reconstruction. We addressed the discrepancy between visual realism and geometric accuracy by introducing a procedurally generated benchmarking framework for Khachkar-inspired reliefs, coupled with a novel curvature-based evaluation criterion, the Feature-Aware Chamfer Distance (FACD).

Our extensive evaluation across eight diverse reconstruction paradigms yielded several pivotal insights into the current state of 3D digitization. First, we empirically demonstrated that traditional global spatial averages, such as the standard Chamfer Distance (CD), inherently exhibit a smoothing bias, leading to the misranking of reconstruction algorithms.

Second, our robustness and ablation analyses highlighted a pronounced dichotomy between implicit and explicit representations. Neural implicit surfaces (e.g., NeRF_InstantNGP, NeuS) exhibited unparalleled convergence stability on untextured, diffuse materials. However, NeuS frequently failed to preserve unique, high-frequency tool marks, while the winner method (NeRF_InstantNGP) successfully recovered these nuances. Conversely, explicit feature-matching pipelines (COLMAP) demonstrated significant fragility, evidenced by substantial failure rates, but provided superior structural fidelity and edge preservation in successful reconstructions. Crucially, this structural nuance was exclusively captured by our FACD curvature-based evaluation criterion.

Despite these insights, we acknowledge certain limitations in our current framework. Primarily, while our procedurally generated dataset provides perfect ground truth and rigorous topological control, it inherently lacks the full spectrum of real-world photometric complexities, including varied environmental lighting, occlusions and biological

weathering. Furthermore, because FACD relies on curvature to scale semantic weights, its performance has yet to be demonstrated on real-world scans, where sensor noise may manifest differently than in our synthetic dataset.

Ultimately, this benchmarking initiative establishes that advancing the digitization of cultural heritage requires a paradigm shift away from feature-agnostic spatial averages. Future work will focus on broadening the applicability of the automated FACD. Furthermore, we encourage the research community to explore feature-aware optimization strategies and design hybrid architectures that fuse the convergence robustness of neural priors with the topological precision of explicit geometric extraction.

CRedit authorship contribution statement

Cristián Llull: Writing – original draft, Visualization, Validation, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Ivan Sipiran:** Validation, Supervision, Project administration, Formal analysis, Conceptualization. **Francisca Gil:** Writing – review & editing, Visualization, Validation, Supervision, Methodology, Investigation, Formal analysis, Conceptualization. **Nelson Baloian:** Supervision, Methodology, Formal analysis. **Anthony Belessis:** Validation, Methodology, Data curation. **Vlassis Fotis:** Methodology, Investigation. **George Ioannakis:** Methodology, Investigation, Formal analysis. **Efthymios Koukoulis:** Writing – original draft, Methodology, Investigation. **Konstantinos Moustakas:** Methodology, Investigation. **Sotiris Papatheodorou:** Writing – original draft, Visualization, Methodology, Investigation. **Ioannis Pratikakis:** Supervision, Methodology, Investigation. **Ioannis Romanelis:** Methodology, Investigation. **Kriti Singh:** Writing – original draft, Validation, Software, Methodology, Investigation. **Eleftheria Solomonidi:** Methodology, Investigation. **Christoforos Vlachos:** Methodology, Investigation. **Vaia Zachari:** Writing – original draft, Validation, Methodology, Investigation.

Declaration of generative AI and AI-assisted technologies in the manuscript preparation process

During the preparation of this work, the author(s) used Gemini in order to improve writing quality, search for state-of-the-art projects and summarize broad topics. After using this tool/service, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the published article.

Declaration of competing interest

The authors declare no competing interests.

Acknowledgments

This document is the result of the research project funded by the ANID-Subdirección de Capital Humano/Doctorado Nacional/2026. The work of Ivan Sipiran was supported by ANID Chile - Fondecyt Regular N°1251263 and the National Center for Artificial Intelligence CENIA FB210017, Basal ANID. The work of Francisca Gil was supported by the National Center for Artificial Intelligence CENIA FB210017, Basal ANID. This work has been partially funded by the European Union’s Horizon Europe research and innovation programme under Grant Agreement No. 101158328 (TEXTaiLES).

Data availability

Data will be made available on request.

References

- [1] Mildenhall B, Srinivasan PP, Tancik M, Barron JT, Ramamoorthi R, Ng R. NeRF: representing scenes as neural radiance fields for view synthesis. 65, (1):New York, NY, USA: Association for Computing Machinery; 2021, p. 99–106. <http://dx.doi.org/10.1145/3503250>.
- [2] Kerbl B, Kopanas G, Leimkuehler T, Drettakis G. 3D Gaussian splatting for real-time radiance field rendering. *ACM Trans Graph* 2023;42(4):1–14. <http://dx.doi.org/10.1145/3592433>, URL <https://dl.acm.org/doi/10.1145/3592433>.
- [3] Wang S, Leroy V, Cabon Y, Chidlovskii B, Revaud J. DUST3r: Geometric 3D vision made easy. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. CVPR, 2024, p. 20697–709. <http://dx.doi.org/10.1109/CVPR52733.2024.01956>.
- [4] Wang J, Chen M, Karaev N, Vedaldi A, Ruppert C, Novotny D. VGGT: Visual geometry grounded transformer. In: 2025 IEEE/CVF conference on computer vision and pattern recognition. CVPR, 2025, p. 5294–306. <http://dx.doi.org/10.1109/CVPR52734.2025.00499>.
- [5] Jensen R, Dahl A, Vogiatzis G, Tola E, Aanaes H. Large scale multi-view stereopsis evaluation. In: Proceedings of the IEEE conference on computer vision and pattern recognition. CVPR, 2014, p. 406–13. <http://dx.doi.org/10.1109/CVPR.2014.59>.
- [6] Knapitsch A, Park J, Zhou Q-Y, Koltun V. Tanks and temples: Benchmarking large-scale scene reconstruction. *ACM Trans Graph* 2017;36(4):78:1–78:13. <http://dx.doi.org/10.1145/3072959>.
- [7] Strecha C, Hansen W, Van Gool L, Fua P, Thoennessen U. On benchmarking camera calibration and multi-view stereo for high resolution imagery. *CVPR* 2008. <http://dx.doi.org/10.1109/CVPR.2008.4587706>.
- [8] McPherron SP, Gernat T, Hublin J-J. Structured light scanning for high-resolution documentation of in situ archaeological finds. *J Archaeol Sci* 2009;36(1):19–24. <http://dx.doi.org/10.1016/j.jas.2008.06.028>, URL <https://www.sciencedirect.com/science/article/pii/S030544030800160X>.
- [9] Schops T, Schonberger JL, Galliani S, Sattler T, Schindler K, Pollefeys M, Geiger A. A multi-view stereo benchmark with high-resolution images and multi-camera videos. In: 2017 IEEE conference on computer vision and pattern recognition (CVPR). IEEE; 2017, p. 2538–47. <http://dx.doi.org/10.1109/cvpr.2017.272>, URL <http://ieeexplore.ieee.org/document/8099755/>.
- [10] Yao Y, Luo Z, Li S, Zhang J, Ren Y, Zhou L, Fang T, Quan L. BlendedMVS: A large-scale dataset for generalized multi-view stereo networks. *Comput Vis Pattern Recognit (CVPR)* 2020. <http://dx.doi.org/10.1109/CVPR42600.2020.00186>, arXiv:1911.10127.
- [11] Yan Z, Mazzacca G, Rigon S, Farella EM, Trybala P, Remondino F. Nerfbk: a holistic dataset for benchmarking nerf-based 3d reconstruction. *Int Arch Photogramm Remote Sens Spat Inf Sci* 2023;XLVIII-1/W3-2023:219–26. <http://dx.doi.org/10.5194/isprs-archives-XLVIII-1-W3-2023-219-2023>, URL <https://isprs-archives.copernicus.org/articles/XLVIII-1-W3-2023/219/2023/>.
- [12] Putra I, Yuniarti A, Fabroyir H. BaliMask3D dataset for 3D completion and reconstruction of traditional Balinese masks. *Sci Data* 2025;12. <http://dx.doi.org/10.1038/s41597-025-05505-8>.
- [13] Zheng X, Zhang S, Lin W, Zhang A, Mayol-Cuevas WW, Liu Y, Shen J. CULTURE3D: A large-scale and diverse dataset of cultural landmarks and terrains for Gaussian-based scene rendering. In: Proceedings of the IEEE/CVF international conference on computer vision. ICCV, 2025, p. 29064–74. <http://dx.doi.org/10.1109/ICCV56789.2025.00429>, URL <https://ieeexplore.ieee.org/document/11443386/>.
- [14] Setiadi DRIM. PSNR vs SSIM: imperceptibility quality assessment for image steganography. *Multimedia Tools Appl* 2021;80(6):8423–44. <http://dx.doi.org/10.1007/s11042-020-10035-z>.
- [15] Wang Z, Bovik A, Sheikh H, Simoncelli E. Image quality assessment: From error visibility to structural similarity. *IEEE Trans Image Process* 2004;13:600–12. <http://dx.doi.org/10.1109/TIP.2003.819861>.
- [16] Zhang R, Isola P, Efros AA, Shechtman E, Wang O. The unreasonable effectiveness of deep features as a perceptual metric. In: 2018 IEEE/CVF conference on computer vision and pattern recognition. 2018, p. 586–95. <http://dx.doi.org/10.1109/CVPR.2018.00068>.
- [17] Barrow HG, Tenenbaum JM, Bolles RC, Wolf HC. Parametric correspondence and chamfer matching: two new techniques for image matching. In: Proceedings of the 5th international joint conference on artificial intelligence - volume 2. IJCAI '77, San Francisco, CA, USA: Morgan Kaufmann Publishers Inc.; 1977, p. 659–63, URL <https://api.semanticscholar.org/CorpusID:1621080>.
- [18] Fan H, Su H, Guibas LJ. A point set generation network for 3D object reconstruction from a single image. In: Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR). 2017, p. 2463–71. <http://dx.doi.org/10.1109/CVPR.2017.264>.
- [19] Wu T, Pan L, Zhang J, Wang T, Liu Z, Lin D. Density-aware chamfer distance as a comprehensive metric for point cloud completion. In: Proceedings of the 35th international conference on neural information processing systems. NIPS '21, Red Hook, NY, USA: Curran Associates Inc.; 2021, URL <https://proceedings.neurips.cc/paper/2021/file/f3bd5ad57c8389a8a1a541a76be463bf-Paper.pdf>.
- [20] Li J, Tan J, Ou Z, Wang H. NCD: Normal-guided chamfer distance loss for watertight mesh reconstruction from unoriented point clouds. In: Computer graphics forum. 44, (6):Wiley Online Library; 2025, e70088. <http://dx.doi.org/10.1111/cgf.70088>.
- [21] Li J, Tian S, Yu L, Ning X. Flexible-weighted chamfer distance: Enhanced objective function for point cloud completion. *IEEE Trans Pattern Anal Mach Intell* 2026;48(07):8106–19. <http://dx.doi.org/10.1109/TPAMI.2026.3669003>, URL <https://doi.ieeecomputersociety.org/10.1109/TPAMI.2026.3669003>.
- [22] Martin D, Bowman B. Dual-constraint contrastive completion: Bridging local precision and global integrity in 3D point cloud reconstruction. 2026, <http://dx.doi.org/10.20944/preprints202604.0233.v1>, Preprints.
- [23] Schönberger JL, Frahm J-M. Structure-from-motion revisited. In: 2016 IEEE conference on computer vision and pattern recognition. CVPR, 2016, p. 4104–13. <http://dx.doi.org/10.1109/CVPR.2016.445>.
- [24] Snavely N, Seitz SM, Szeliski R. Photo tourism: exploring photo collections in 3D. *ACM Trans Graph* 2006;25(3):835–46. <http://dx.doi.org/10.1145/1141911.1141964>.
- [25] Scharstein D, Szeliski R. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *Int J Comput Vis* 2002;47(1):7–42. <http://dx.doi.org/10.1023/A:1014573219977>.
- [26] Barron JT, Mildenhall B, Tancik M, Hedman P, Martin-Brualla R, Srinivasan PP. Mip-NeRF: A multiscale representation for anti-aliasing neural radiance fields. In: 2021 IEEE/CVF international conference on computer vision. ICCV, 2021, p. 5835–44. <http://dx.doi.org/10.1109/ICCV48922.2021.00580>.
- [27] Rosinol A, Leonard J, Carlone L. NeRF-SLAM: Real-time dense monocular SLAM with neural radiance fields. 2023, p. 3437–44. <http://dx.doi.org/10.1109/IROS5552.2023.10341922>.
- [28] Reiser C, Szeliski R, Verbin D, Srinivasan P, Mildenhall B, Geiger A, Barron J, Hedman P. MERF: Memory-efficient radiance fields for real-time view synthesis in unbounded scenes. *ACM Trans Graph* 2023;42(4). <http://dx.doi.org/10.1145/3592426>.
- [29] Wang P, Liu L, Liu Y, Theobalt C, Komura T, Wang W. NeuS: Learning neural implicit surfaces by volume rendering for multi-view reconstruction. In: Ranzato M, Beygelzimer A, Dauphin Y, Liang PS, Vaughan JW, editors. Proceedings of the 35th international conference on neural information processing systems. NIPS '21, 34, Red Hook, NY, USA: Curran Associates, Inc.; 2021, p. 27171–83, URL https://proceedings.neurips.cc/paper_files/paper/2021/file/e41e164f7485ec4a28741a2d0ea41c74-Paper.pdf.
- [30] Li Z, Müller T, Evans A, Taylor RH, Unberath M, Liu M-Y, Lin C-H. Neuralangelo: High-fidelity neural surface reconstruction. In: 2023 IEEE/CVF conference on computer vision and pattern recognition. CVPR, 2023, p. 8456–65. <http://dx.doi.org/10.1109/CVPR52729.2023.00817>.
- [31] Wang Y, Han Q, Habermann M, Daniilidis K, Theobalt C, Liu L. NeuS2: Fast learning of neural implicit surfaces for multi-view reconstruction. In: 2023 IEEE/CVF international conference on computer vision. ICCV, Los Alamitos, CA, USA: IEEE Computer Society; 2023, p. 3272–83. <http://dx.doi.org/10.1109/ICCV51070.2023.00305>, URL <https://doi.ieeecomputersociety.org/10.1109/ICCV51070.2023.00305>.
- [32] Wang Y, Skorokhodov I, Wonka P. HF-NeuS: improved surface reconstruction using high-frequency details. In: Proceedings of the 36th international conference on neural information processing systems. NIPS '22, Red Hook, NY, USA: Curran Associates Inc.; 2022, URL https://proceedings.neurips.cc/paper_files/paper/2022/file/0ce8e3434c7b486bbddff9745b2a1722-Paper-Conference.pdf.
- [33] Müller T, Evans A, Schied C, Keller A. Instant neural graphics primitives with a multiresolution hash encoding. *ACM Trans Graph* 2022;41(4). <http://dx.doi.org/10.1145/3528223.3530127>.
- [34] Guédon A, Lepetit V. SuGaR: Surface-aligned Gaussian splatting for efficient 3D mesh reconstruction and high-quality mesh rendering. In: Proceedings - 2024 IEEE/CVF conference on computer vision and pattern recognition, CVPR 2024. Proceedings of the IEEE computer society conference on computer vision and pattern recognition, IEEE Computer Society; 2024, p. 5354–63. <http://dx.doi.org/10.1109/CVPR52733.2024.00512>.
- [35] Wu J, Li R, Zhu Y, Guo R, Sun J, Zhang Y. Sparse2DGS: Geometry-prioritized Gaussian splatting for surface reconstruction from sparse views. 2025, p. 11307–16. <http://dx.doi.org/10.1109/CVPR52734.2025.01056>, URL https://openaccess.thecvf.com/content/CVPR2025/papers/Wu_Sparse2DGS_Geometry-Prioritized_Gaussian_Splatting_for_Surface_Reconstruction_from_Sparse_Views_CVPR_2025_paper.pdf.
- [36] Held J, Son S, Vandeghen R, Rebain D, Gadelha M, Zhou Y, Cioppa A, Lin MC, Van Droogenbroeck M, Tagliasacchi A. MeshSplatting: Differentiable rendering with opaque meshes. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. CVPR, 2026, p. 7320–9, URL https://openaccess.thecvf.com/content/CVPR2026/papers/Held_MeshSplatting_Differentiable_Rendering_with_Opaque_Meshes_CVPR_2026_paper.pdf.
- [37] Leroy V, Cabon Y, Revaud J. Grounding image matching in 3D with MAST3r. In: Computer vision – ECCV 2024: 18th European conference, milan, Italy, September 29–October 4, 2024, proceedings, part LXXII. Berlin, Heidelberg: Springer-Verlag; 2024, p. 71–91. http://dx.doi.org/10.1007/978-3-031-73220-1_5.

- [38] Yang J, Sax A, Liang KJ, Henaff M, Tang H, Cao A, Chai J, Meier F, Feiszli M. Fast3R: Towards 3D reconstruction of 1000+ images in one forward pass. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. CVPR, 2025, <http://dx.doi.org/10.1109/cvpr52734.2025.02042>.
- [39] Shen Y, Zhang Z, Qu Y, Zheng X, Ji J, Zhang S, Cao L. FastVGGT: Fast visual geometry transformer. In: The fourteenth international conference on learning representations. 2026, URL <https://openreview.net/forum?id=asl8NJIIMe>.
- [40] Chen H, Cai R, Marschner S, Snavely N. ArchSym: Detecting 3D-grounded architectural symmetries in the wild. 2026, <http://dx.doi.org/10.48550/ARXIV.2604.22202>, CoRR arXiv:2604.22202.
- [41] Kerr J, Kim CM, Goldberg K, Kanazawa A, Tancik M. LERF: Language embedded radiance fields. In: 2023 IEEE/CVF international conference on computer vision. ICCV, 2023, p. 19672–82. <http://dx.doi.org/10.1109/ICCV51070.2023.01807>.
- [42] Paolini G, Tortorici C, Berretti S, Youssef AH, Benhabiles H, Cabani A, Ruiwen H, Hammoudi K, Ganapathi I, Ali S, Velayudhan D, Assefa M, Werghi N. SHREC'25 track on multiple relief patterns: Report and analysis. 2025, <http://dx.doi.org/10.48550/arXiv.2508.09909>.
- [43] Blender Online Community. Blender - a 3D modelling and rendering package. Stichting Blender Foundation, Amsterdam: Blender Foundation; 2018, URL <http://www.blender.org>.
- [44] Broekman A, Gräbe PJ. PASMVS: A perfectly accurate, synthetic, path-traced dataset featuring specular material properties for multi-view stereopsis training and reconstruction applications. ScienceDirect 2020;32:106219. <http://dx.doi.org/10.1016/j.dib.2020.106219>, URL <https://www.sciencedirect.com/science/article/pii/S2352340920311136>.
- [45] Jacobson A, Panozzo D, et al. libigl: A simple C++ geometry processing library. 2018, URL <https://libigl.github.io/>.
- [46] Rousseeuw P. Least median of squares regression. J Amer Statist Assoc 1984;79:871–80. <http://dx.doi.org/10.2307/2288718>.
- [47] Umeyama S. Least-squares estimation of transformation parameters between two point patterns. IEEE Trans Pattern Anal Mach Intell 1991;13(4):376–80. <http://dx.doi.org/10.1109/34.88573>.
- [48] Ravi N, Gabeur V, Hu Y-T, Hu R, Ryali C, Ma T, Khedr H, Rädle R, Rolland C, Gustafson L, Mintun E, Pan J, Alwala KV, Carion N, Wu C-Y, Girshick R, Dollár P, Feichtenhofer C. SAM 2: Segment anything in images and videos. 2024, arXiv preprint arXiv:2408.00714, URL <https://arxiv.org/abs/2408.00714>.
- [49] Kazhdan M, Bolitho M, Hoppe H. Poisson surface reconstruction. In: Proceedings of the fourth eurographics symposium on geometry processing. SGP '06, 7, (4):Goslar, DEU: Eurographics Association; 2006, p. 61–70, URL <https://dl.acm.org/doi/10.5555/1281957.1281965>.
- [50] Tancik M, Weber E, Ng E, Li R, Yi B, Wang T, Kristoffersen A, Austin J, Salahi K, Ahuja A, Mcallister D, Kerr J, Kanazawa A. Nerfstudio: A modular framework for neural radiance field development. In: ACM SIGGRAPH 2023 conference proceedings. SIGGRAPH '23, New York, NY, USA: Association for Computing Machinery; 2023, <http://dx.doi.org/10.1145/3588432.3591516>.
- [51] Yariv L, Gu J, Kasten Y, Lipman Y. Volume rendering of neural implicit surfaces. In: NIPS '21, Red Hook, NY, USA: Curran Associates Inc.; 2021, URL <https://dl.acm.org/doi/10.5555/3540261.3540628>.
- [52] Tukey JW. Some graphic and semigraphic displays. Stat Pap Honor Georg W. Snedecor 1972;5:293–316.
- [53] Dodge Y. Exploratory data analysis. In: The concise encyclopedia of statistics. New York, NY: Springer New York; 2008, p. 192–4. http://dx.doi.org/10.1007/978-0-387-32833-1_136.