Geometry reconstruction and orthophoto generation from 3D Gaussian Splatting in architectures with thin elements

Mattia Previtali [1], Luigi Barazzetti [1], Fabio Roncoroni [2]

[1] Department of Architecture, Built environment and Construction engineering (ABC), Politecnico di Milano, Via Ponzio 31, Milan, Italy - (luigi.barazzetti, mattia.previtali)@polimi.it
[2] Politecnico di Milano, Polo territoriale di Lecco, via Previati 1/c, Lecco, 23900, Italy - fabio.roncoroni@polimi.it

Keywords: 3D Gaussian Splatting, Orthophoto, Photogrammetry, Façade Mapping, Thin Structures, Railings, Balconies.

Abstract

Orthorectified façade imagery and metric 3D models support architectural documentation, conservation, and further analysis. Orthophoto production from standard Structure-from-Motion and Multi-View Stereo (SfM–MVS) pipelines performs well on broad opaque surfaces, but can be challenging on thin, repetitive, and partly transparent elements (e.g., railings, balusters, grilles). Indeed, in the latter cases depth estimation becomes unstable, filtering removes some structures, and meshing priors thicken elements or bridge voids, compromising both geometry generation and orthophoto production. This paper evaluates 3D Gaussian Splatting (3DGS) as a surface-free alternative for metric façade representation and orthophoto generation in such conditions. We propose a compact façade-plane alignment and scale control procedure to render orthographic products. On a real façade dataset acquired under diffuse illumination, we compare a standard SfM–MVS true-orthophoto baseline with three 3DGS workflows: PostShot training with a custom orthographic renderer, Tortho-Gaussian for optimization and orthographic rendering, and Blender rendering of PostShot splats. Quality is assessed, using a laser scanning acquisition as a benchmark, via completeness, edge fidelity and topological preservation. Results indicate that 3DGS better preserves the topological pattern in railing regions, keeping members separated and apertures open, and enables rapid orthographic rendering once trained. SfM–MVS shows better results on large, well-textured wall areas, whereas 3DGS may introduce mild edge softening or halos at high-contrast boundaries.

1. Introduction

Orthorectified façade imagery and metrically reliable 3D models are key outputs in architectural surveying and documentation, supporting conservation planning, condition assessment, and downstream workflows such as HBIM. In particular, façade orthophotos provide a scale-consistent, map-like representation where measurements, annotations, and thematic layers (materials, decay, stratigraphy, interventions) can be created and shared with reduced ambiguity. The production of façade orthophotos is often embedded in a broader photogrammetric pipeline based on Structure-from-Motion (SfM) for camera orientation, followed by Multi-View Stereo (MVS) for dense depth estimation, surface generation, and true-orthophoto mosaicking. On large opaque surfaces such as masonry, plaster and stone cladding, this workflow performs well, provided that image networks are well planned and texturing is sufficient. However, façade scenes frequently include thin and repetitive elements like railings, balusters and grilles characterized by a strong view-dependent appearance. These components are visually and metrically important, both as architectural features and as sources of occlusion for the wall behind, yet they remain a recurring failure mode in dense reconstruction and true-orthophoto generation. In such regions, depth estimates can be noisy, incomplete, or locally inconsistent, and the subsequent stages that regularize and fuse depth may remove those elements or fill gaps in ways that distort the geometry. The resulting orthophoto may exhibit missing elements, bridged voids, merged members, thickened bars, fragmented details, or texture bleeding, compromising readability and metric interpretation exactly where fine detail matters most. Proper reconstruction of such objects requires carefully planned data acquisition that captures their complex structure and is integrated into the overall façade documentation project, increasing both the number of images acquired and the processing time.. Moreover, the repetitive structure of those elements may pose some issues in the image orientation stage.

The difficulty in document those elements is not merely a matter of required level of details. Indeed, MVS formulations implicitly assume that corresponding patches can be matched across views under approximately Lambertian reflectance and that the scene can be represented by a locally smooth surface. Thin structures challenge both assumptions simultaneously. Indeed, on the one side thin elements alternately reveal and hide the background wall across viewpoints, reducing the number of images that support the same depth hypothesis and making multi-view fusion unreliable. On the other hand, materials such as steel and metal often exhibit specular highlights and view-dependent shading that further weaken photoconsistency. In addition, thin elements can be often under-sampled since a bar or baluster may project to only a few pixels, making correspondences highly sensitive to noise and small exposure changes. Finally, repeated patterns create ambiguity and evenly spaced vertical members can generate multiple plausible matches.

As a result of those peculiarities, consistency checks, and depth-map fusion stages (e.g., visibility filtering and outlier removal) eliminates many valid but sparsely supported observations of thin parts. Ultimately, to suppress noise and fill gaps, MVS pipelines impose smoothness and surface continuity priors in depth estimation and in subsequent meshing (e.g., Poisson or screened Poisson). On alternating sequences of "material–void–material," these priors tend to bridge voids, thicken elements, or merge adjacent members, degrading topological fidelity. As previously mentioned, mitigation of those effects is possible by adopting stronger convergent coverage, diffuse illumination, and specialized multi-layer or edge-aware MVS, but these measures increase acquisition and processing complexity and

still leave thin, repetitive structures as a recognized failure mode for standard SfM–MVS pipelines.

These limitations motivate the exploration of alternative scene representations that can better handle view-dependent radiance and layered visibility (Chen et al., 2025). In recent years, neural rendering approaches such as NeRF and, more recently, 3D Gaussian Splatting (3DGS) have demonstrated strong performance in synthesizing novel views from posed imagery, particularly in challenging scenes where classical stereo struggles (Murtiyoso and Macher, 2025). 3DGS (Kerbl et al., 2023) represents a scene as a set of anisotropic Gaussian primitives with optimizable position, covariance, opacity, and colour (often with view-dependent components), and it renders images through efficient rasterization of these "splats" without requiring an explicit watertight surface. This surface-free formulation is particularly appealing for thin and porous objects: instead of forcing observations into a single continuous mesh, the representation can maintain fine-scale detail and partially occluding structures in a way that remains consistent with the image evidence. If coupled with a metric reference frame and controlled scaling, 3DGS could therefore enable the generation of orthographic façade products that preserve thin-element topology while remaining compatible with surveying requirements.

Despite the strong visual results reported in the neural rendering literature (Mildenhall et al., 2021), some practical aspects of a 3DGS based orthophoto production remain open like the possibility to use 3DGS for metrically accurate façade orthophotos, the selection of well-defined ground sampling distance (GSD) and the alignment to a façade reference plane. To partially cope with those aspects in this paper, we investigate 3D Gaussian Splatting as an alternative representation for façade mapping in the presence of thin, repetitive structures. We propose a compact procedure for façade-frame definition and scale control that supports orthographic rendering at a prescribed GSD, and we benchmark three 3DGS-based workflows against a standard photogrammetric baseline. To fairly compare the prosed methods, all of them share the same image set, camera orientations, façade reference frame, crop window, and orthophoto output parameters.
The remainder of the paper is organized as follows. Section 2 reviews related work on façade orthophotos, SfM–MVS limitations on thin structures, and recent neural rendering approaches for geometric documentation. Section 3 describes the dataset, acquisition strategy, reference system definition, the photogrammetric baseline, and the tested 3DGS pipelines. Section 4 presents the results and comparative evaluation. Section 5 discusses observed trade-offs and implications for operational documentation workflows and outlines future research directions.

2. Related work

Orthophotos are widely adopted in architectural and built-environment documentation because they provide a scale-consistent representation suitable for metric annotation, thematic mapping, and condition assessment. When scenes contain significant depth discontinuities (e.g., balconies, frames, railings), conventional orthorectification may suffer from occlusions and relief displacement (Rau et al., 2002) for this reason more recent frameworks emphasize the central role of occlusion detection and the sensitivity of the product to DSM/surface quality (Gharibi and Habib, 2018).
Orthophotos are generally obtained by using an image-based 3D reconstruction strategy that typically follows a Structure-from-Motion (SfM) stage for camera pose estimation, followed by dense Multi-View Stereo (MVS) and surface reconstruction, enabling DSM/mesh-based orthophoto generation. Comparative studies show that dense matching is strongly affected by weak texture, specularities, repeated patterns, and occlusions, with error propagation into later regularization, fusion and meshing steps (Seitz et al., 2006; Remondino et al., 2014; Furukawa and Hernández, 2015).
Thin and repetitive façade elements (railings, balusters, grilles) combine multiple adverse conditions since they are frequently under-sampled, generate ambiguous correspondences due to periodic patterns, and create view-dependent occlusions of the background. Even when camera orientation is reliable, dense matching may be incomplete or locally inconsistent, and surface reconstruction tends to simplify the geometry (e.g., thickening members, bridging voids), which directly impacts the fidelity of true-orthophoto mosaics in these regions. More generally, stereo correspondence evaluations highlight how algorithmic choices in matching, smoothness, regularization, and outlier handling can trade off detail preservation against stability (Scharstein and Szeliski, 2002).
In the last few years Neural Radiance Fields (NeRF) introduced a differentiable volumetric representation optimized from posed images, enabling high-quality novel-view synthesis without explicit meshing (Mildenhall et al., 2020). Building on this paradigm, recent work has explored orthographic and true-orthophoto generation directly from radiance-field representations. Ortho-NeRF proposes tile-based reconstruction and "true-ortho" volume rendering for producing TDOMs from UAV imagery (Chen et al., 2024), while NeRFOrtho modifies the rendering rays to be parallel to generate orthographic projection images (with TDOM as a special case), reporting improvements in texture detail and reduced stitching/projection artifacts compared to mesh-projection approaches (Yue et al., 2025). Although most of these studies target aerial and UAV scenarios, their core insight, i.e., decoupling orthographic product generation from watertight meshing and DSM-dependent occlusion handling, is highly relevant to façade mapping where thin structures and layered visibility are common.
In particular, 3D Gaussian Splatting (3DGS) represents a scene as a set of anisotropic Gaussian primitives with optimized position, covariance, opacity, and appearance, rendered efficiently via splatting/rasterization. The approach demonstrated real-time rendering with competitive training times and strong view-synthesis quality (Kerbl et al., 2023). Because 3DGS does not require an explicit watertight surface, it is an appealing candidate for thin structures where mesh-based pipelines may bridge gaps or merge elements. Recent extensions explicitly target true-orthophoto products: TOrtho-Gaussian generates TDOMs by orthogonally splatting optimized Gaussians, aiming to bypass DSM generation and separate occlusion-detection steps while improving boundaries and performance on reflective or slender structures (Wang et al., 2024). Importantly for close-range documentation, initial photogrammetry-focused studies have started to assess Gaussian splatting for orthophoto generation in built environments: Previtali et al. (2024) investigate orthophoto production with Gaussian splatting and report reduced artifacts over reflective surfaces when using imagery from low-cost sensors, while Murtiyoso and Macher (2025) present a façade-oriented comparison between Gaussian splatting, MVS, and TLS for orthophoto generation, highlighting both the potential and the practical constraints of metric exploitation.
This paper evaluates whether 3DGS can be operationalized for metric façade orthophotos under controlled scale and façade-plane alignment, and how it compares to a conventional SfM–

MVS true-orthophoto workflow specifically on challenging thin/repetitive elements.

3. Methodology

The aim of this study is to evaluate whether 3D Gaussian Splatting (3DGS) can improve façade orthophoto production in the presence of thin, repetitive, and partially occluding elements such as railings, balusters, and grilles where standard SfM–MVS pipelines frequently fail in good reconstruction due to missing members, fused bars, bridged voids, and loss of the "material–void–material" topology both on the 3D reconstruction and in the orthophoto production. To reach this goal we compare a conventional photogrammetric workflow based on SfM–MVS reconstruction and true-orthophoto generation implemented in Agisoft Metashape against three 3DGS routines for producing metrically consistent façade orthophotos. The goal is not to maximize view-synthesis realism, but to assess the suitability of outputs for architectural documentation tasks such as topological readability of thin elements and edge interpretability at defined GSD.

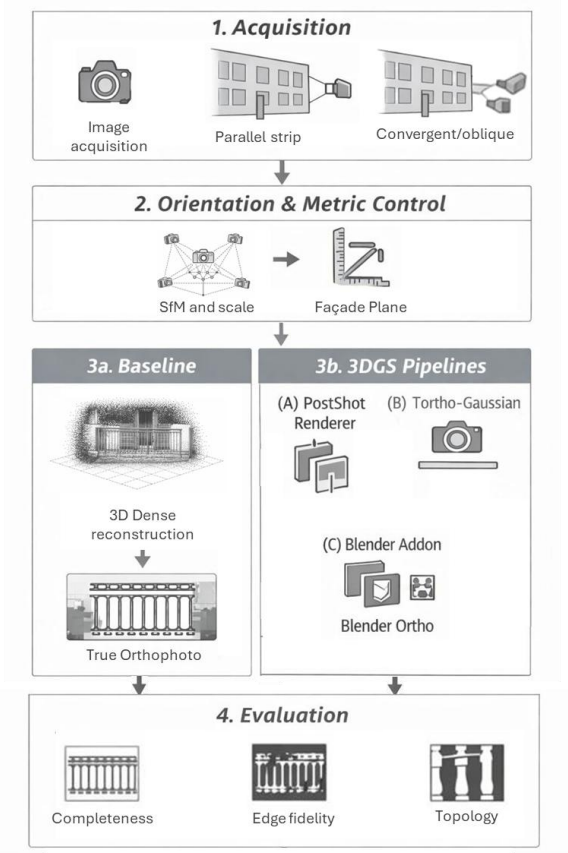

Figure 1. Overview of the methodology used to assess SfM-MVS and 3DGS workflows for orthophoto generation of thin elements.

A central requirement is to isolate the effect of scene representation and orthographic rendering therefore, all methods share: (i) the same image set, (ii) the same camera poses, (iii) the same façade reference frame (plane and local axes), and (iv) the same orthophoto specification (bounding box and pixel size). Under these controlled conditions, differences in orthophotos can be attributed primarily to the underlying 3D representation (mesh and Gaussian splats respectively) and the orthographic rendering strategy. In particular, Figure 1 summarizes the methodology adopted in this work. Starting from image acquisition, the dataset of façade is photographed using a combined network composed of a near-parallel strip to maximize orthographic fidelity on planar regions and convergent oblique views to increase parallax and visibility around thin, occluding elements such as railings (Section 3.1). The images are then processed in a standard SfM orientation stage, where image alignment provides camera poses and scale is imposed from field measurements; a façade plane is fitted to define the local reference frame. From this shared setup, two reconstruction and rendering branches are executed: (i) a baseline SfM–MVS workflow (Section 3.2) producing depth maps and a façade-plane true orthophoto at a target resolution of 2 mm/pixel, and (ii) a 3DGS branch (Section 3.3) where the same oriented imagery is used to train Gaussian splats and generate orthographic outputs via three workflows and configurations: (A) PostShot training and a custom orthographic splat renderer (Section 3.4), (B) TOrtho-Gaussian orthographic splatting (Section 3.5), and (C) Blender-based orthographic rendering through a dedicated addon (Section 3.6). All orthophotos are produced within the same façade-plane. Finally, the outputs are assessed through three criteria (Section 3.7) targeting the failure modes of thin structures: (i) completeness of reconstructed railing geometry, (ii) edge fidelity in the orthophotos, and (iii) preservation of topology (separated members and open apertures).

3.1 Dataset and acquisition protocol

The dataset consists of a near-range image block of a building façade including both extended planar regions (wall and shutters) and a balcony railing (Figure 2). The dataset is composed of 82 images captured under diffuse illumination and with constant exposure settings, reducing cast shadows.

Figure 2. Example of images in the dataset with railing and balcony used for MVS and 3DGS test.

The image network combines two complementary components:

1. a near-parallel strip approximately facing the façade, intended to maximise orthographic fidelity on planar regions. These images provide high-frequency texture on the wall and improve the stability of mapping-like products.
2. convergent oblique views, intended to increase angular diversity and parallax around protruding and occluding elements (railing/balustrade). This component supports multi-view consistency and reduces ambiguous matches in repetitive structures. For 3DGS, increased angular coverage is also beneficial to reduce view-dependent artefacts at edges

and improve the stability of opacity and colour estimates.

No image-specific masking was applied during processing; the same raw imagery is used in all pipelines. The same photo set is processed in Metashape and reused to train all 3DGS models, ensuring a direct comparison across methods.

3.2 Image orientation and photogrammetric baseline definition

The baseline follows a conventional SfM–MVS workflow in Agisoft Metashape 2.2.1, producing both dense 3D support and a true orthophoto on the façade plane.

Images are aligned using Metashape's SfM pipeline with estimation of intrinsic camera parameters and bundle adjustment to estimate exterior orientations. The resulting oriented camera model is used as the common pose set for all 3DGS experiments (Section 3.3). Reusing a single pose solution avoids attributing differences in orthophoto quality to pose discrepancies rather than to the reconstruction/rendering method. To define the scaling of the project a metric scale is imposed using an in-situ measurement, i.e. the length of the balcony (4.10 m). The façade reference system is constructed by a local orthonormal frame where the X–Y axes lie in the façade plane and the outward normal defines the Z axis. The resulting local frame is then used as the common orthophoto projection plane for every method.

Dense reconstruction is performed using depth maps at high quality, followed by depth-map fusion with Metashape default settings. This step provides dense point support and a surface representation used in true-orthophoto generation. Finally, the true orthophoto is generated by orthographic projection onto the façade plane at 2 mm/pixel.

3.3 3DGS test framework and controlled factors

To assess the viability of 3DGS for orthophoto production of thin elements experiments we test three practical configurations reflecting currently available tools and typical production workflows (Table 1):

- (A) 3DGS trained in PostShot and rendered with a custom orthographic renderer;
- (B) TOrtho-Gaussian optimisation and direct orthographic rendering (Wang et al., 2024); and
- (C) 3DGS trained in PostShot and rendered in Blender via 3dgs-render-blender-addon.

3DGS Orthophoto tests		
	Training	**Ortho rendering**
3DGS (A)	PostShot	gsplat-based custom orthographic projection
3DGS (B)	TOrtho-Gaussian integrated optimisation and orthographic splats rendering	
3DGS (C)	PostShot	Blender ortho camera + clipping slab; scene-linear sRGB export

Table 1. 3DFS Orthophoto configurations: training and ortho rendering strategy for the different tests.

The three workflows differ in the orthographic projection and rendering implementation (custom splat renderer, integrated orthographic splatting, Blender-based rendering), while sharing the same inputs and geometric constraints. This design helps separate "representation effects" (surface vs splats) from "toolchain effects" (renderer behaviour, clipping, compositing, colour handling). All 3DGS experiments are evaluated under a common framework to ensure comparability. Specifically, they use the same image set as the baseline and reuse the same camera poses obtained from the SfM solution. The metric scale and the façade reference frame are kept identical across configurations, and orthophotos are generated using the same façade-plane crop window at a fixed resolution of 2 mm/pixel. In addition, all 3DGS variants follow the same training schedule, with the number of optimization iterations set to 30,000.

3.4 PostShot training and custom orthographic renderer (Configuration A)

In configuration (A), training is performed in a dedicated 3DGS environment, while orthographic rendering is executed with explicitly defined geometric parameters consistent with the photogrammetric baseline. Specifically, the 3DGS model is trained in PostShot - Jawset (https://www.jawset.com/) and then exported as a PLY containing the Gaussian primitives. The rendering to orthographic projection is carried out by using a custom Python script (Python 3.12.0). The renderer script reads the PLY and reconstructs, for each splat, its 3D position (x,y,z), anisotropic scale parameters (s_0, s_1, s_2), orientation stored as a quaternion, and opacity. Colour is not stored as plain RGB; instead, the file encodes spherical harmonics (SH) coefficients (DC terms plus the higher-order SH coefficients). The renderer can additionally restrict contributions to a narrow slab around the façade plane (near/far clipping) to suppress out-of-plane splats and reduce haze in the orthographic output.

Before rendering, the script applies the same parameter decoding used in typical 3DGS toolchains: the per-splat scales are converted from log-space via an exponential mapping ($s = exp(scale_{raw}) \cdot 0.02$), while opacities ($\alpha$) are mapped to [0-1] through a sigmoid function. In the current implementation, opacities are further biased to enforce a minimum contribution ($\alpha = sigmoid\ (\alpha_{raw}) \cdot 0.9 + 0.2$, followed by clamping), which increases splat visibility but may also exacerbate haloing or haze if not combined with slab clipping or opacity thresholding.

Orthographic image generation is then performed by using the gsplat library (https://github.com/nerfstudio-project/gsplat). The script sets a fixed view matrix to provide an orthogonal projection in the X–Y plane. The gsplat rasterizer projects the Gaussian primitives with an orthographic camera and performs the accumulation internally returning both an RGB image and an alpha map. To stabilize visualization and obtain cleaner visual results the rendered normalizes the model centring and scaling it to a unit bounding radius using (1):

$$x' = (x - c)/s \quad (1)$$

where:

x'= normalized coordinates,
x = original coordinates,
c = position of the centre of the bounding box,
s = scale factor to contain the model into a unit bounding radius

This determines that a scaling is performed in the model and the target GSD (2 mm/pixel) for the orthophoto needs to be rescaled. By computing the bounding box of the model, it is possible to determine the scaling of the model in the normalized reference system and compute a normalized GSN_{norm} ($GSN_{norm} = GDS_{target}/s$) to be used to obtain an orthophoto with the desired target GSD. Similarly, the inverse transform of the normalization ($x = s \cdot x' + c$) is then used to map the orthophoto origin and extent back to metric coordinates for georeferencing.

The rendered orthophoto is saved as a TIFF file with associated TFW containing positioning and scaling information.

3.5 TOrtho-Gaussian optimisation and orthographic rendering (Configuration B)

In configuration (B), we adopt the TOrtho-Gaussian (https://gwen233666.github.io/Ortho-Gaussian/) pipeline to optimize a 3DGS model and to generate orthographic products directly through orthogonal splatting. The method is explicitly designed for True Digital Orthophoto Map (TDOM) generation: instead of relying on a reconstructed DSM/mesh and separate occlusion-detection heuristics. Ortho-Gaussian produces orthophotos by projecting Gaussian ellipsoid kernels onto a 2D image plane in an orthographic manner.

The authors also describe a divide-and-conquer strategy for large scenes and an anisotropic Gaussian kernel formulation intended to improve accuracy and visual quality, particularly for challenging regions such as reflective surfaces and slender structures (Wang et al., 2004). From an implementation perspective, we follow the reference toolchain provided in the public Tortho_Gaussian repository (https://github.com/xwangSGG/Tortho_Gaussian), which separates perspective optimization and orthographic splatting into two environments. A first environment installs the standard differentiable Gaussian rasterizer for perspective training. As input the implementation is using Colmap results ad input, for this reason we converted the Metashape orientation results into a Colmap database and we performed the training using the repository's training scripts (train_vast.py) with the iteration budget fixed to 30,000 iterations to match the other 3DGS configurations. A second environment is a dedicated to orthographic rasterization using the provided splatting scripts (final_splat.py), which generate the orthographic outputs and store them in the project output directory.

Although in the current implementation the tool does not explicitly manage a GSD parameter, the effective GSD of the rendered orthographic image is determined implicitly by the combination of (i) the metric extent of the area covered in the output and (ii) the output image size (width and height in pixels). More specifically the amount of scene visible in the orthographic output is controlled mainly through two parameters: (i) fov_deg (horizontal field of view) and (ii) scale (camera distance/scaling).

Increasing either parameter enlarges the portion of the scene included in the final render, while decreasing it zooms in to a smaller area. In practice, only one of these parameters needs to be tuned: both act as a zoom/coverage control, affecting how much of the 3D scene is mapped into the 2D image. The output resolution sets the size (width and height) of the output image dimensions in pixels. Therefore, achieving a target GSD (e.g., 2 mm/pixel) requires selecting a consistent pair: the rendered footprint and the output resolution.

3.6 PostShot training and Blender rendering (Configuration C)

In configuration (C), the PostShot-trained model, stored as PLY file, is imported into Blender using the public addon 3dgs-render-blender-addon (https://github.com/Kiri-Innovation/3dgs-render-blender-addon). An orthographic camera is aligned with the façade normal, and its orthographic scale is derived so that the rendered pixel size matches 2 mm/pixel over the specified crop window in a way like the one used in TOrtho-Gaussian. However, Blender route is particularly use-friendly because the camera, framing, and output resolution can be verified interactively (viewport overlay, safe areas, and explicit pixel dimensions), reducing the risk of unintended scaling. To suppress splats outside the façade, near/far clipping is restricted to a thin slab around the façade plane. This prevents contributions from floating primitives or background elements that could otherwise appear as haze or ghosting in orthographic renderings. Rendering is performed in scene-linear sRGB, and outputs are exported in TIFF or JPG file. However, direct georeferencing of the file is not available.

3.7 Quality assessment and evaluation criteria

Evaluation considers both the reconstructed 3D geometry and the derived orthophotos, using a terrestrial/close-range laser scan as an external benchmark. The scan point cloud is first registered to the same façade reference frame adopted for the image-based workflows, and a reference scan-based orthophoto is generated by orthographic projection onto the façade plane at the same crop window and GSD. All metrics are then computed in this shared coordinate system and are designed to target the main failure modes observed on railings/balustrades. Three criteria are considered: (i) completeness, (ii) edge fidelity and (iii) topological preservation.

Completeness measures how much of the thin structure is actually reconstructed within a predefined railing region of interest (ROI). We compare each reconstruction (MVS dense cloud and 3DGS splat centres) against the scan by quantifying the fraction of scan points in the railing ROI that find a corresponding reconstructed element within a tolerance distance of ±5.0 mm. Equivalently, completeness can be reported as the percentage of the scan-supported railing area that is covered by the reconstruction, highlighting missing members or gaps, poor matching also results in mistakes of the final orthophoto.

Edge fidelity quantifies how well railing boundaries are preserved in the orthophoto with respect to the scan-based orthophoto. Railing edge maps are manually extracted on both the tested orthophoto and on the scan reference. The score is computed as the fraction of edges in the tested orthophoto that fall within a small tolerance band (1 pixels, i.e., 2.0 mm) around the reference edges (implemented as a dilation of the reference edge mask). This metric penalizes missing bars and edge shifts introduced by geometric or rendering inaccuracies.

Topological preservation assesses whether the orthophoto maintains the correct "material–void–material" structure of the railing. Using the scan-based orthophoto as reference, we evaluate whether vertical members remain separated and whether apertures (gaps between members) remain open. In practice, we count the number of distinct vertical members detectable in the railing ROI and the number of open gaps between them, and we compare these counts (and their spatial regularity) to the reference. In addition to counting separated members and open apertures, we assess the geometric regularity of the pattern by measuring the centre-to-centre spacing between adjacent vertical bars. Bar centrelines are extracted in the railing ROI and the distribution of inter-bar distances is compared to the scan-derived reference. The metric is reported as the mean absolute error (MAE) and relative error of spacing, as well as the percentage of spacing values within a tolerance (±2 pixels at the target GSD). This directly targets common failures such as merged members, bridged voids, or inconsistent bar thickness that compromises interpretability.

Overall, this scan-referenced evaluation framework links geometric fidelity and orthophoto readability in a single metric space, enabling a fair and repeatable comparison between surface-based SfM–MVS products and surface-free 3DGS orthographic renderings.

Figure 3. Laser scanner reference and results of the reconstruction of the railings for the different tests: 3D model (left) and detail of the orthophoto (right)

4. Results

This section reports the scan-referenced comparison between the SfM–MVS baseline and the three 3DGS configurations introduced in Section 3. A terrestrial close-range laser scan is used as an external benchmark for both geometry and orthophoto quality using metric introduced in Section 3.7. The benchmark is composed of 8 scans acquired with Leica RTC360. In particular, the railings is 4.12 m long and is composed of 38 vertical elements (6 of dimension 3x3 cm and 32 of dimension 1x1 cm) and 6 horizontal elements. The scan point cloud was registered to the same façade reference frame adopted by the image-based workflows, and a scan-based orthophoto was generated by orthographic projection onto the façade plane using the same crop window and sampling (2 mm/pixel). Figure 3 summarizes the results obtained for the different orthophoto tests and the laser scanner benchmark.

4.1 Qualitative comparison

The SfM–MVS baseline produces a visually coherent orthophoto on extended planar areas, where texture is strong and the geometry can be represented by a locally smooth surface model. In the railing region, however, the baseline exhibits the typical degradation expected from depth-map instability and subsequent fusion/regularization: several vertical members appear partially or completely missing, and apertures are not consistently preserved along the full span. In addition, when the orthophoto is generated, the railing is incorrectly projected onto the background, causing it to blend and merge with the underlying façade texture. These effects reduce the readability of the "material–void–material" pattern and can lead to ambiguous interpretation of which wall areas are truly visible behind the railing. From a documentation perspective, such failures are critical because the railing is not a secondary detail: it dominates occlusions and is often the target of mapping (e.g., bar spacing, deterioration, deformation).

Across the 3DGS outputs, the railing is generally represented more continuously, with improved separation between members and gaps. This is consistent with a splat-based representation. Nevertheless, 3DGS orthophotos can show characteristic rendering artifacts when used for map-like outputs, including mild boundary softening, haloing around high-contrast edges, and haze/ghosting when out-of-plane splats contribute along the orthographic projection direction. The use of slab clipping is intended to mitigate these artifacts by limiting unwanted contributions outside the façade plane and stabilizing the orthographic projection.

4.2 Scan-referenced 3D completeness

Table 2 summarizes scan-referenced 3D completeness, computed as the fraction of scan points in the railing ROI that find a corresponding reconstructed element within a tolerance distance of 5 mm. The SfM–MVS baseline achieves a completeness of 17.4%, confirming that dense stereo and depth-map fusion tend to lose support on thin members where correspondences are unstable and occlusions are view-dependent. This loss of 3D support explains the corresponding orthophoto failure. In contrast, all 3DGS configurations improve completeness, ranging from 87.9% to 94.2% suggesting that the combination of splat-based representation and orthographic-oriented processing better retains geometry evidence along the railing. Configuration A and C share the same amount of completeness (87.9%) since both relies on the same splats training based on PostShot. From an operational point of view, the completeness gain of ~70 percentage points over SfM–MVS represents a meaningful improvement for façade documentation, because it reduces the probability that bars are entirely absent from the deliverables.

Method	Completeness C_{3D} [%]	Number of elements
Metashape (SfM–MVS)	17.4	9 vertical
3DGS-A (PostShot + custom)	87.9	38 vertical 9 horizontal
3DGS-B (TOrtho-Gaussian)	94.2	38 vertical 9 horizontal
3DGS-C (PostShot + Blender)	87.9	38 vertical 9 horizontal

Table 2. Scan-referenced 3D completeness in the railing ROI.

4.3 Orthophoto edge fidelity

Edge fidelity is evaluated against the scan-based orthophoto by measuring the fraction of railing edges in the tested orthophoto that fall within a 1-pixel tolerance band around the reference edges. As reported in Table 3, the SfM–MVS orthophoto achieves an edge fidelity of 18.2%. This reflects two complementary effects: while majority of railings bars are not 3D reconstructed the background façade is reconstructed and when orthophoto is generated the railing is wrongly projected on the background and fused with it.
All 3DGS routes yield higher edge fidelity than the baseline, with configuration all configurations exceeding 85%. Configuration A presents an accuracy 85.2%. The best performance is obtained by configuration B (88.3%), indicating that orthographic splatting can produce boundaries that are both more complete and more spatially consistent with the scan reference. Configuration C (87.1%) performs comparably, consistent with the practical benefit of controlling camera alignment and clipping in a graphical-based environment.

4.4 Topological preservation

Topological preservation is assessed through member and aperture counts and through the accuracy of centre-to-centre spacing between adjacent bars. In Table 3, ΔN and ΔA represent the difference in the number of detected members and apertures with respect to the reference (0 means all members/apertures are detected). The SfM–MVS baseline significantly underestimates both members and apertures (ΔN = -24, ΔA = -24), consistent with missing bars and local bridging. 3DGS configuration achieves the correct counting of railing members and apertures. Spacing regularity further differentiates the photogrammetric baseline and the 3DGS methods.

Method	Edge fidelity [%]	Members ΔN	Apertures ΔA	Spacing MAE [px]	Spacing within ±2 px [%]
Metashape (SfM–MVS)	18.2	-24	-24	46.5	0.0
3DGS-A (PostShot + custom)	85.2	0	0	1.8	77.9
3DGS-B (TOrtho-Gaussian)	88.3	0	0	1.6	81.4
3DGS-C (PostShot + Blender)	86.3	0	0	2.2	79.4

Table 3. Scan-referenced 3D completeness in the railing ROI.

The photogrammetric SfM–MVS baseline results confirms, both in terms of spacing MAE and share of spacing withing ±2 pixels, an unsuccessful reconstruction and orthophoto results. The 3DGS orthophotos reduce spacing error substantially: configuration B reaches 1.6 px MAE (81.4% within tolerance), while configurations A and C (using the same PhostShot reconstruction) achieve errors of 1.8 and 2.2 px MAE respectively and tares within tolerance of 77.9 and 79.4% respectively. These results are consistent with a more stable preservation of the repetitive "bar–gap–bar" pattern in splat-based orthographic outputs, which is crucial for façade documentation and metric interpretation.

4.5 Discussion of results

Overall, the scan-referenced evaluation indicates that 3DGS-based pipelines provide a clear advantage in railing-dominated areas, increasing 3D completeness by roughly 70 percentage points compared to SfM–MVS and improving both boundary agreement and topological correctness in the orthophoto. The baseline remains effective on broad planar wall regions, where dense stereo produces sharp textures, but its reconstruction support degrades on thin repetitive members, leading to missing bars and reduced aperture visibility. Among the 3DGS routes, configuration B yields the most consistent performance across metrics. Configuration A and C perform similarly. These findings confirm that, when controlled for camera geometry and orthophoto specification, 3DGS can better preserve the structural readability of thin façade elements while maintaining metric consistency against an external scan reference

5. Conclusions

From our experiments, we observe that in railing/balustrade zones, 3DGS better preserves the alternating "material–void–material" pattern in the reconstructed geometry, delivering higher topological fidelity: bars remain separated, and apertures stay open more consistently. This also translates into more visually coherent orthophotos, especially at moderate target GSDs. In addition, once trained, 3DGS renders orthographic views very quickly (milliseconds to seconds), which is advantageous for survey teams iterating multiple orthophotos without re-running longer image orthorectification processes. By contrast, on broad, textured plaster or stone, classical MVS performs strongly; 3DGS offers little to no benefit and can even reduce quality. Orthographic compositing of anisotropic splats can introduce edge softening at high-contrast boundaries and faint halos around thin members. If training views under sample certain directions (e.g., only façade parallel), 3DGS may bake view-dependent appearance into the radiance, causing subtle ghosting or tone shifts in the orthophoto; oblique images generally mitigate this. Strong glass panels and areas can induce view-dependent colour that looks plausible in novel views but is not faithful as a true orthophoto. In such cases, both MVS and 3DGS can fail, even if for different reasons and in different modes.

References

Chen, X., Zhan, Z., Wang, X., Yuan, Q., Xie, H., 2024: Ortho-NeRF: generating a true digital orthophoto map using the neural radiance field from unmanned aerial vehicle images. *Geo-spatial Information Science*.

Chen, S., Yan, Q., Qu, Y., Gao, W., Yang, J., Deng, F., 2025. Ortho-NeRF: generating a true digital orthophoto map using the neural radiance field from unmanned aerial vehicle images. Geo-spatial Information Science, 28(2), 741-760.

Furukawa, Y., Curless, B., Seitz, S.M., Szeliski, R., 2010: Towards Internet-scale multi-view stereo. In: *Proc. CVPR 2010*, pp. 1434–1441. https://doi.org/10.1109/CVPR.2010.5539802

Furukawa, Y., Hernández, C., 2015: Multi-View Stereo: A Tutorial. *Foundations and Trends® in Computer Graphics and Vision*, 9(1–2), 1–148. https://doi.org/10.1561/0600000052

Gharibi, H., Habib, A.F., 2018: True Orthophoto Generation from Aerial Frame Images and LiDAR Data: An Update. *Remote Sensing*, 10(4), 581. https://doi.org/10.3390/rs10040581

Kerbl, B., Kopanas, G., Leimkühler, T., Drettakis, G., 2023. 3D Gaussian splatting for real-time radiance field rendering. ACM Trans. Graph., 42(4), 139-1.

Mildenhall, B., Srinivasan, P. P., Tancik, M., Barron, J. T., Ramamoorthi, R., Ng, R., 2021. Nerf: Representing scenes as neural radiance fields for view synthesis. Communications of the ACM, 65(1), 99-106.

Murtiyoso, A., Macher, H., 2025. Gaussian Splatting for Facade Orthophoto Generation–Comparison with MVS and TLS. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *48*, 1059-1064.

Previtali, M., Barazzetti, L., Roncoroni, F., 2024. Orthophoto generation with gaussian splatting: mitigating reflective surface artifacts in imagery from low-cost sensors. International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 48, 371-378.

Rau, J.-Y., Chen, N.-Y., Chen, L.-C., 2002: True Orthophoto Generation of Built-Up Areas Using Multi-View Images. *Photogrammetric Engineering & Remote Sensing*, 68(6), 581–588.

Remondino, F., Spera, M.G., Nocerino, E., Menna, F., Nex, F., 2014: State of the art in high density image matching. *The Photogrammetric Record*, 29(146), 144–166. https://doi.org/10.1111/phor.12063

Scharstein, D., Szeliski, R., 2002: A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms. *International Journal of Computer Vision*, 47, 7–42. https://doi.org/10.1023/A:1014573219977

Seitz, S.M., Curless, B., Diebel, J., Scharstein, D., Szeliski, R., 2006: A Comparison and Evaluation of Multi-View Stereo Reconstruction Algorithms. In: *Proc. CVPR 2006*, pp. 519–528.

Wang, X., Zhang, W., Xie, H., Ai, H., Yuan, Q., Zhan, Z., 2024. Tortho-gaussian: Splatting true digital orthophoto maps. arXiv preprint arXiv:2411.19594.

Yue, D., Liu, X., Wan, Y., Zhang, Y., Zheng, M., Fan, W., Zhong, J., 2025: NeRFOrtho: Orthographic Projection Images Generation based on Neural Radiance Fields. *International Journal of Applied Earth Observation and Geoinformation*, 136, 104378. https://doi.org/10.1016/j.jag.2025.104378