Spherical Vision meets 3D Semantics: towards efficient LOD3 Model Generation for Smart Cities

Mohammad Saadatseresht[1,2], Hossein Arefi[2], and Qazaleh Askari[1]

[1] School of Engineering, Faculty of Surveying and Geospatial Engineering, University of Tehran, Tehran, Iran,
{msaadat,qazale.askari}@ut.ac.ir
[2] i3mainz – Institute for Spatial Information and Surveying Technology, Mainz University of Applied Sciences, Mainz, Germany,
{mohammad.saadatseresht, hossein.arefi}@hs-mainz.de

Keywords: LOD3 Reconstruction, Building Façade Modeling, Spherical Images, Multi-View Optimization, Mapillary, Open Data

Abstract

The generation of Level of Detail 3 (LoD3) building models is essential for applications such as urban digital twins, energy analysis, and smart city planning. However, conventional approaches based on terrestrial LiDAR or UAV photogrammetry remain costly, labor-intensive, and difficult to scale. This paper presents a scalable framework for transforming LoD1 building models into LoD3 façade representations using openly available urban data, including OpenStreetMap footprints, street-level spherical imagery, and weak point-cloud priors. The proposed method formulates the reconstruction problem as a facet-based modeling task, where each façade is processed independently in a local coordinate system derived from LoD1 geometry. A rectification strategy is introduced to generate fronto-parallel façade images directly from spherical panoramas, avoiding perspective distortions and facilitating image analysis. To address the challenges of unstructured data acquisition, a visibility-driven view selection scheme and a multi-view fusion framework are developed to construct robust façade evidence maps. The 3D geometry is estimated as a depth field through a multi-resolution optimization framework integrating ray consistency, appearance cues, point-cloud support, and structural regularization. Planar segmentation, polygonization, and geometric regularization are subsequently applied to derive structured façade elements. Openings such as windows and doors are detected using combined geometric and image-based evidence and further refined through architectural constraints. Experimental results demonstrate that the proposed framework enables reliable reconstruction of façade geometry and structural details using only open and low-cost data sources, providing a practical pathway for large-scale LoD3 generation in real urban environments.

1. Introduction

Three-dimensional building models at Level of Detail 3 (LoD3) are essential for a wide range of applications, including urban planning, energy simulation, and digital twins. Despite their importance, generating LoD3 models at scale remains challenging due to the high cost and operational complexity of traditional acquisition methods such as terrestrial LiDAR and UAV photogrammetry.

In contrast, large volumes of publicly available urban data—such as OpenStreetMap (OSM) building footprints and street-level panoramic imagery (e.g., Mapillary)—offer a promising alternative (Pokuciński and Mrozek, 2023). However, these data sources are inherently unstructured, lack controlled acquisition geometry, and are not designed for precise 3D reconstruction. This creates significant challenges in terms of geometric consistency, visibility reasoning, and depth estimation.

Recent works have explored image-based façade parsing and deep learning for semantic enrichment of LoD models (Chen et al., 2025; Wei et al., 2023). While effective in controlled scenarios, such approaches often rely on curated datasets, perspective projections, or dense multi-view geometry, limiting their applicability to opportunistic data (Tang et al., 2025).

This paper proposes a geometry-aware, image-driven framework for transforming LoD1 building models into LoD3 façade representations using unstructured and publicly available data. Unlike purely learning-based pipelines, the proposed method leverages LoD1 geometry as a structural prior and reformulates the problem as a facet-wise reconstruction task in rectified façade space.

The core contributions are: (i) A rectification-based façade reconstruction strategy from spherical imagery that avoids perspective conversion and simplifies image analysis, (ii) A visibility-driven view selection mechanism tailored for unstructured panoramic data, (iii) A 2.5D façade modeling framework combining image evidence, weak point cloud priors, and architectural constraints, and (iv) An efficient pipeline suitable for rapid LoD1-to-LoD3 conversion using open data.

2. Methodology

2.1 Overview

The proposed method follows a facet-based reconstruction paradigm, where each façade of a LoD1 building model is processed independently in a local coordinate system. The workflow integrates three complementary data sources:
(i) LoD1 building geometry derived from OSM,
(ii) spherical panoramic imagery with known exterior orientation parameters, and
(iii) a point cloud used as a weak geometric prior.
Unlike conventional multi-view stereo pipelines, the method operates in a rectified façade domain, significantly simplifying correspondence estimation and structural interpretation.

2.2 LoD1-Based Façade Parameterization

Each building footprint is extruded using its minimum and maximum elevation attributes to generate a LoD1 block model (Figure 1e). Each footprint edge defines a façade plane, parameterized by a local coordinate system such that horizontal axis along the edge, vertical axis aligned with gravity, and normal axis representing depth. This parameterization enables consistent mapping between image space and 3D geometry. The generated LoD1 geometry and integrated point cloud are illustrated in Figure 1.

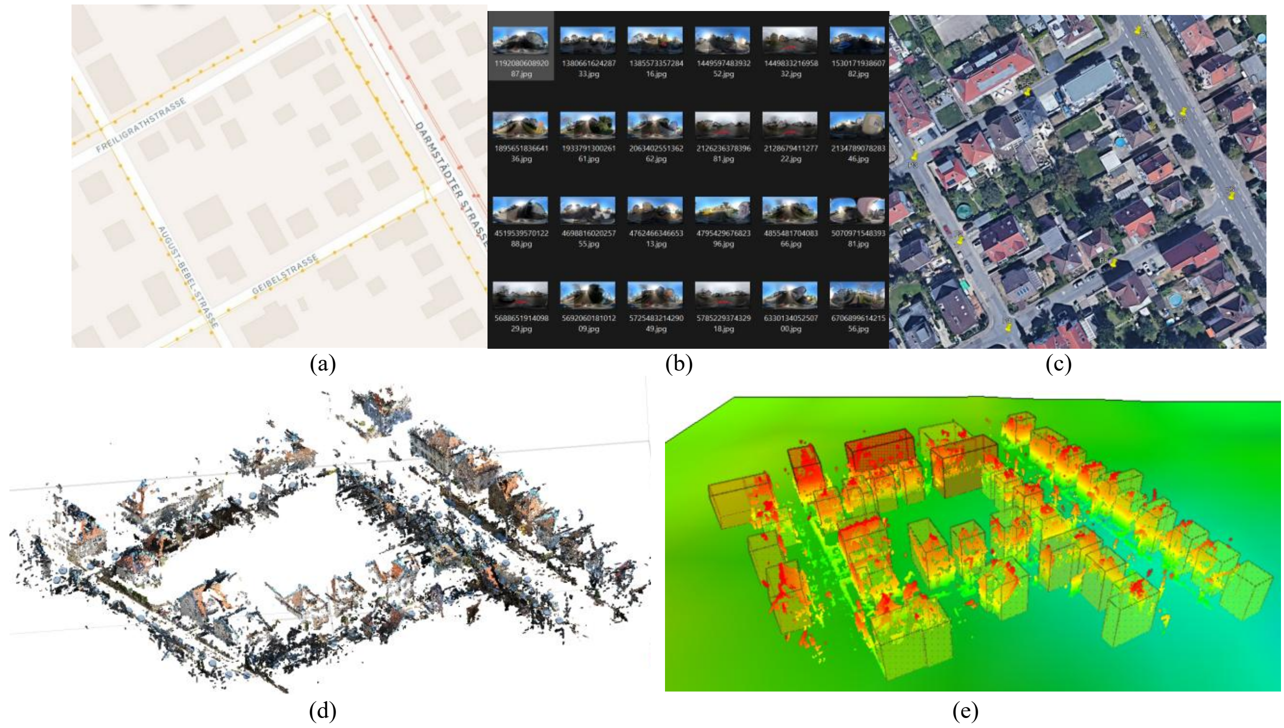


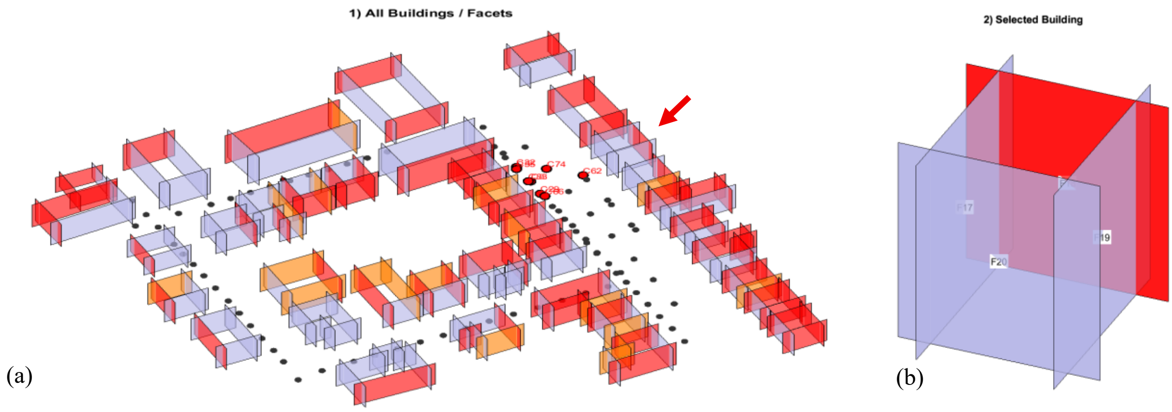
Figure 1. Input data acquisition and preprocessing for the study area. (a–b) Selection of Mapillary spherical imagery covering the urban block. (c) Eight ground control points (GCPs) extracted from Google Maps along the surrounding streets. (d) Dense 3D point cloud reconstructed from the spherical image dataset. (e) LoD1 building model derived from OSM footprints, integrated with the reconstructed point cloud, where ground elevation is obtained from the SRTM DEM and roof height is estimated from the highest points of the point cloud.

2.3 View Selection under Unstructured Acquisition

Given the unstructured nature of panoramic imagery, a two-stage view selection strategy is employed: (1) Geometric filtering including viewing angle constraint, distance constraint, and image footprint coverage; and (2) Visibility refinement including façade-based occlusion reasoning and point cloud–based occlusion analysis using a pyramidal volume. Each façade is assigned one primary and several auxiliary views, ensuring sufficient coverage while avoiding occlusions (Figure 2).

2.4 Façade Rectification from Spherical Imagery

To eliminate perspective distortions, façade regions are extracted via ray-based projection from spherical images as illustrated in Figure 3 (Rezaei and Arefi, 2025). For each façade grid point, a 3D point is generated in world coordinates, a ray is cast toward the camera, spherical coordinates are computed, and corresponding image pixels are sampled. This produces a fronto-parallel rectified façade image, where architectural elements exhibit regular geometry, significantly facilitating subsequent processing, as also demonstrated in spherical-image-based 3D reconstruction approaches (Askari et al., 2024).



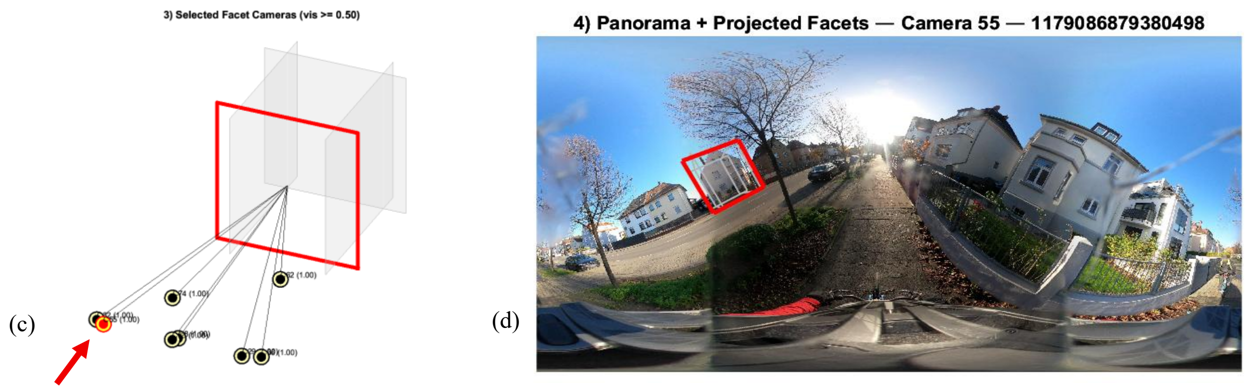


Figure 2. Input data preparation and initial geometric setup. (a) OSM building footprints enriched with elevation attributes as Figure 1e. (b) Generation of façade facets as vertically extruded planes from footprint edges, with slight extensions applied to compensate for positional uncertainties. (c) Selection of candidate spherical image stations located in front of and within a suitable distance from the target façade (highlighted in red). (d) Projection of the selected façade onto the equirectangular image corresponding to the chosen camera station (indicated by the red arrow in (c)).

Figure 3. Rectified (normalized) façade images for all selected camera stations shown in Figure 2c. Perspective distortions are removed through façade-based rectification, resulting in fronto-parallel views. Some images exhibit occlusions (e.g., trees or street objects) and motion blur, reflecting the challenges of unstructured image acquisition.

2.5 Multi-View Fusion and Point-Cloud Support

To integrate information from multiple observations, all normalized façade images are combined into a set of fusion maps defined on the façade domain. These include a fused intensity image, a validity mask, per-pixel view support, radiometric consistency indicators, and edge confidence measures. The fusion process is designed to be robust against outliers and variations across views, ensuring that only consistent observations contribute to the reconstruction. In parallel, the point cloud is projected into the local façade coordinate system. Rather than being used directly for reconstruction, it provides statistical depth support in each region of the façade. For each grid cell, aggregated measures such as median depth, dispersion, and point density are computed. This point-cloud information is incorporated as a soft geometric prior, guiding the reconstruction without enforcing strict constraints.

2.6 Multi-View Raster Optimization

The façade surface is modeled as a depth field defined over a regular grid in the local façade coordinate system. The optimal depth configuration is estimated by minimizing a composite objective that integrates multiple sources of information, including geometric consistency, image evidence, and prior knowledge. For each facet, the depth raster W is estimated by minimizing a composite energy in relation (1):

$$E(W) = \lambda_{\text{ray}} E_{\text{ray}} + \lambda_{\text{app}} E_{\text{app}} + \lambda_{\text{pc}} E_{\text{pc}} + \lambda_{\text{smooth}} E_{\text{smooth}} + \lambda_{\text{planar}} E_{\text{planar}} \quad (1)$$

where each term encodes a different source of information. Examples of the generated fusion maps are presented in Figure 4.

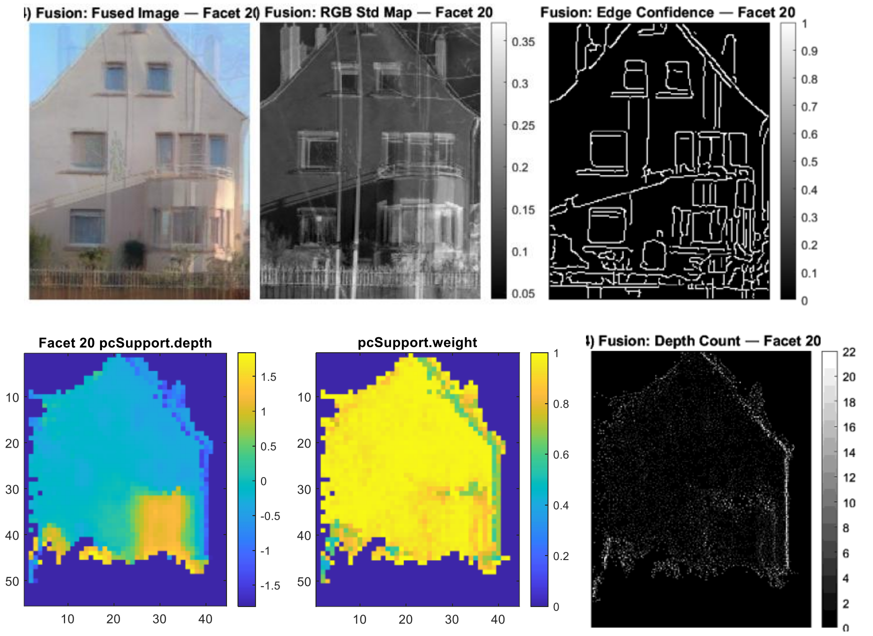


Figure 4. Fusion maps in the façade domain. The upper row shows image-based fusion maps generated by integrating normalized façade images from Figure 3, including intensity, validity, and edge confidence measures. The lower row presents point cloud–derived maps providing statistical depth support, such as aggregated depth values, dispersion, and point density.

- **Ray Consistency:** Each normalized image provides a set of viewing rays toward the façade. The reconstruction enforces that the estimated 3D surface remains consistent with these viewing rays across all supporting views. A weighting strategy is used to prioritize views that observe the façade more frontally and are not affected by occlusions. This ensures that reliable observations have a stronger influence on the solution, while oblique or partially occluded views are down-weighted.
- **Appearance Consistency**: Although rectification reduces geometric distortions, differences between views remain due to depth variations and non-planar structures. To address this, appearance-based features such as intensity and gradient information are extracted from each normalized image. The method encourages consistency between these features across multiple views, providing an additional cue for stabilizing the reconstruction, particularly in regions where geometric information is weak.
- **Point-Cloud Support**: Where point-cloud data are available, they provide valuable depth cues. The reconstruction is encouraged to remain close to the statistical depth estimates derived from nearby points, particularly in regions with strong support. However, this influence is applied in a soft manner, allowing the solution to deviate where image evidence suggests more accurate geometry.
- **Regularization**: To ensure stable and physically plausible surfaces, regularization terms are introduced (i) a smoothness constraint to reduce noise and enforce local continuity, and (ii) a planarity constraint to encourage piecewise planar structures. Together, these constraints produce surfaces that are smooth in homogeneous regions while preserving sharp transitions at structural boundaries.
- **Optimization Strategy**: The optimization is performed in a multi-resolution framework, starting from a coarse representation and progressively refining the solution. At each level, the depth field is updated iteratively using gradient-based optimization. This hierarchical strategy improves convergence and helps avoid local minima, especially in complex façade regions.

An example of the optimized depth raster is shown in Figure 5.

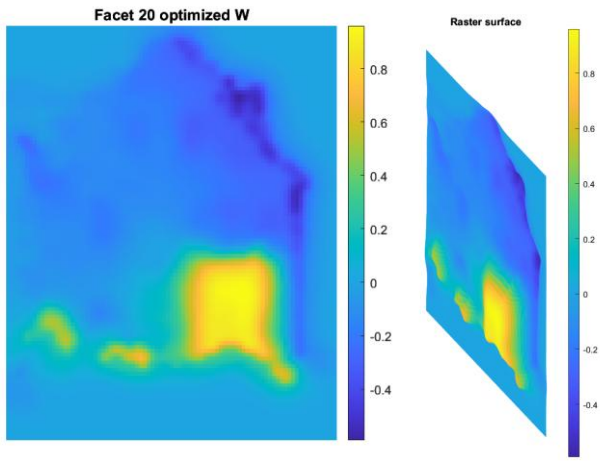


Figure 5. Result of multi-resolution raster optimization for facet-based depth modeling. As can be observed in yellow area, the protrusion of the building entrance is clearly visible in the depth map.

2.7 Planar Patch Segmentation

After optimization, A multi-view image-based fusion stage is employed to aggregate edge and line evidence from normalized façade views into a unified representation. For each view, edge maps are extracted using gradient-based detectors, and line segments are identified to capture strong structural features. These cues are combined into per-view confidence maps and subsequently fused across all views through pixel-wise accumulation and normalization, taking into account visibility and support consistency. The resulting fusion map enhances persistent structural boundaries while suppressing noise and view-specific artifacts, providing a robust basis for subsequent segmentation and detection of façade elements such as windows and doors. The resulting fused structural boundary maps are illustrated in Figure 6.

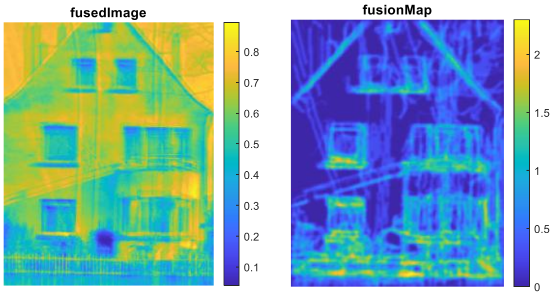


Figure 6. Persistent structural boundaries derived from multiple normalized views. Left: fused intensity image. Right: fused line map obtained from multi-view edge extraction.

2.8 Polygonization and Geometric Regularization

Each planar patch is converted into a polygonal representation by extracting its boundary in the façade domain and mapping it back to 3D space.

To improve geometric quality, the resulting polygons are simplified and regularized. In particular:

- edges are aligned with dominant façade directions
- parallelism and orthogonality constraints are enforced
- approximately rectangular regions are adjusted accordingly

This step ensures that the reconstructed geometry is not only accurate but also consistent with architectural structures.

2.9 Opening Candidate Detection

Openings such as windows and doors are detected within suitable façade regions. The method combines two complementary cues:

- **geometric evidence**, based on depth differences relative to the surrounding surface (e.g., recessed regions)
- **image evidence,** derived from edge strength and intensity patterns (Chen et al., 2025; Wei et al., 2023)

Candidate regions are identified and approximated by rectangles aligned with façade directions. Each candidate is evaluated based on geometric consistency, image support, and shape plausibility. Low-confidence detections are removed, and overlapping candidates are filtered to retain only the most reliable ones.

2.10 Opening Regularization

Detected openings are further refined using architectural regularities.

Openings are grouped into rows and columns based on their spatial arrangement. Within each group:

- positions are aligned
- sizes are normalized
- spacing is regularized

This step significantly improves the structural consistency of the façade and reduces detection noise, resulting in more realistic configurations.

2.11 Building-Level Fusion

Although each façade is reconstructed independently, consistency must be ensured at the building level. Adjacent façade patches are aligned along shared edges, such as corners or roof-wall intersections. Geometric inconsistencies are resolved through snapping and intersection constraints.

Openings located too close to shared boundaries are optionally filtered to avoid conflicts during model integration.

2.12 Final LoD3 Model Assembly

The final LoD3 model is constructed by assembling all validated façade patches and their associated openings. Each façade element is represented as a polygon with possible inner holes corresponding to openings. This structured representation preserves topology and enables flexible export to formats such as OBJ or CityJSON.

In addition, each element is associated with a confidence measure reflecting the reliability of the reconstruction.

3. Conclusion

This paper presents the initial development of an efficient framework for LoD1-to-LoD3 façade reconstruction using unstructured and publicly available data. By combining geometric priors from LoD1 models with rectified spherical imagery and weak point cloud constraints, the method demonstrates a pathway for scalable and cost-effective generation of detailed 3D building models.

The proposed rectification-based formulation simplifies the reconstruction problem and provides a practical alternative to traditional data-intensive approaches. As this is ongoing research, future work will focus on completing the building-level fusion, integrating learning-based semantic cues, and extending the framework to roof structures and full building reconstruction.

References

Askari, Q., H. Arefi and M. Maboudi (2024). "3D Reconstruction of Building Blocks Based on Extraction of Exterior Wall Lines Using Point Cloud Density Generated from Spherical Camera Images." Remote Sensing 16(23): 4377.

Chen, Y., T. Wang, G. Chen, K. Zhu, X. Tan, J. Wang, W. Guo, Q. Wang, X. Luo and X. Zhang (2025). "BFA-YOLO: A balanced multiscale object detection network for building façade elements detection." Advanced Engineering Informatics 65: 103289.

Pokuciński, S. and D. Mrozek (2023). "Object detection with YOLOv5 in indoor equirectangular panoramas." Procedia Computer Science 225: 2420-2428.

Rezaei, S. and H. Arefi (2025). "Evaluation of Network Design and Solutions of Fisheye Camera Calibration for 3D Reconstruction." Sensors 25(6): 1789.

Tang, W., W. Li, X. Liang, O. Wysocki, F. Biljecki, C. Holst and B. Jutzi (2025). Texture2LoD3: Enabling LoD3 Building

Reconstruction With Panoramic Images. Proceedings of the Computer Vision and Pattern Recognition Conference.

Wei, G., F. Wan, W. Zhou, C. Xu, Z. Ye, W. Liu, G. Lei and L. Xu (2023). "Bfd-Yolo: A Yolov7-based detection method for building façade defects." Electronics 12(17): 3612.