Omni2LOD3: Extracting Façade Geometry and Semantic Features from Omnidirectional Imagery and Drone Photogrammetry

Demi Julianna L. Gentiles[1], Khalil C. Torneros[1], Jarence David D. Casisirano[1], Alexis Richard C. Claridades[1]

[1]Department of Geodetic Engineering, University of the Philippines Diliman, Quezon City, Philippines
{dlgentiles, kctorneros, acclaridades, jdcasisirano}@up.edu.ph

Keywords: LOD3, Building façade extraction, CityGML, Drone photogrammetry, Omnidirectional imagery, 3D point cloud

Abstract

CityGML Level of Detail 3 (LOD3) building model generation requires accurate façade representation. Existing workflows often rely on LiDAR, dense multi-view imagery, or pre-existing lower-LOD models, limiting their applicability in data-scarce environments. Additionally, most methods are developed on buildings with simple façade geometry, focusing mainly on detecting openings while keeping façade features planar. This study presents Omni2LOD3, a semi-automated workflow that extracts LOD3-ready geometric and semantic façade components from UAV photogrammetry and a limited number of omnidirectional images, challenging the assumption that façade-level reconstruction requires dense multi-view stereo or LiDAR acquisition. The study integrates existing methods and operationalizes them within a unified workflow for façade data extraction. The workflow is demonstrated on a building with complex overhangs, a curved façade, and partial occlusions. An LOD2 base model is derived from nadir drone imagery, while façade information is captured using omnidirectional images. Façade geometry is reconstructed through monocular depth estimation, and semantic features are extracted using color-based segmentation, establishing a lightweight baseline for semantic extraction from sparse image-derived point clouds. Evaluation against a terrestrial SLAM LiDAR benchmark indicates that the reconstructed façade geometry adequately captures LOD3-relevant structures, while HSV-based semantic classification results in 70% accuracy, reaching 81% after manual refinement. Overall, the results demonstrate the feasibility of recovering meaningful façade geometry and semantics from sparse omnidirectional imagery. These outputs constitute the LOD2 base model and semantically segmented façade point cloud that form the foundation for LOD3 generation, with full reconstruction currently under active development within the broader Omni2LOD3 pipeline.

1. Introduction

3D city models are an important component of modern geospatial infrastructure and are widely used in applications such as urban planning, environmental simulation, and digital twins. The usefulness of these models depends on their Level of Detail (LOD), which determines their geometric and structural complexity (Biljecki et al., 2016). In CityGML 3.0, Levels of Detail describe geometric representation rather than semantic completeness, meaning that façade element recoverability contributes to LOD3 readiness even when semantic reconstruction remains partial. Recent updates further emphasize that LOD is defined in terms of spatial representation rather than semantic decomposition (Kutzner et al., 2020). Within this framework, LOD3 building models represent the highest level of exterior geometric detail typically used in façade-level reconstruction workflows (Open Geospatial Consortium (OGC), 2021), as illustrated in Figure 1.

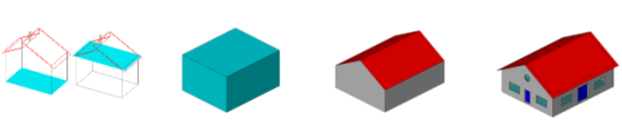

Figure 1. Representation of the same real-world building in different Levels of Detail. Adapted from OGC (2021)

Traditionally, LOD3 building reconstruction relies on LiDAR data, which provides accurate 3D measurements of building geometry but requires costly acquisition and specialized workflows (Wang, 2013). In addition, LiDAR-based pipelines are not always feasible in many regions, particularly where national-scale LiDAR surveys or detailed 3D building datasets are unavailable.

Image-based reconstruction methods have therefore gained traction as an alternative for generating 3D building models. Unmanned Aerial Vehicle (UAV) photogrammetry has emerged as a widely used technique for capturing building geometry due to its flexibility and relatively low cost, though deployment in urban environments may also be subject to permit requirements and airspace regulations. UAV imagery can provide high-resolution observations of roofs and building structures, enabling detailed reconstruction (Vacca et al., 2017). However, aerial imagery alone often struggles to capture façade information due to limited viewing angles, occlusions, or distortions, particularly in dense urban environments with surrounding structures or large canopies. As a result, ground-level details and recessed features are often incomplete. To address the limitations of aerial photogrammetry, close-range dense image capture offers an alternative, as photogrammetric geometry reconstruction techniques from multiple overlapping images reliably produce accurate point clouds (Pantoja-Rosero et al., 2022). However, the need for several overlapping image captures normally requires extensive field time.

To address these limitations, recent studies integrate complementary image sources that provide additional perspectives of building façades. Panoramic or omnidirectional imagery offers wide field-of-view observations of urban scenes and has shown potential for improving façade reconstruction (Ogawa et al., 2024; Tang et al., 2025). These perspectives complement aerial observations by providing direct views of vertical surfaces while reducing the number of required images. Their wide field of view also enables more complete façade coverage, even when captured from fewer or varying distances.

This study introduces Omni2LOD3, a semi-automated workflow for generating LOD3-ready geometric and semantic façade components from sparse omnidirectional imagery and UAV-derived base models. Three aspects constitute its methodological contribution. First, monocular depth estimation recovers the façade geometry, including recessed elements and building installations, rather than detecting openings alone. Second, the LOD2 model from nadir UAV imagery serves as the georeferencing reference frame, avoiding the need for ground control points. Third, HSV-based segmentation provides a lightweight baseline for semantic extraction where geometry-driven methods are less effective on sparse, irregular point clouds. Rather than presenting a complete end-to-end LOD3 pipeline, the proposed approach focuses on recovering structurally meaningful façade geometry and semantic elements under constrained acquisition conditions.

2. Related Work

Research on 3D building reconstruction has produced diverse workflows across different Levels of Detail. The present study contributes a workflow for generating LOD3-ready façade geometry and semantics from sparse imagery through three methodological components: LOD3 façade enrichment from a UAV-derived base model, monocular depth estimation for façade geometry recovery, and HSV-based semantic segmentation of image-derived façade point clouds. Accordingly, the following sections review prior work related to each component.

2.1 LOD3 Façade Enrichment

Method	Input	Façade Geometry	Key Limitation
Tang et al. (2025)	Panoramic SVI, LOD base model	None (semantic enrichment only)	Requires pre-existing LOD2
Ogawa et al. (2024)	Omni-directional SVI, Building footprints + Aerial photographs	None (semantic/texture only)	No façade depth recovery; annotated training data required
Wysocki et al. (2024)	MLS point cloud, LOD1/LOD2 base model	Recess geometry	Requires dense LiDAR and existing LOD2
Pantoja-Rosero et al. (2022)	Multi-view images	Partial (flat openings only)	Limited to planar openings; no surface relief
Omni2LOD3 (this study)	Sparse omnidirectional imagery, Nadir UAV	Façade point cloud with geometric properties	LOD3 mesh generation is outside the scope of this study

Table 1. Comparison of LOD3 Façade Enrichment Method

Methods for generating LOD3 façade information generally follow one of two approaches: refining an existing lower-LOD model or reconstructing buildings from scratch. Refinement-based methods use an existing building model as a spatial reference. Tang et al. (2025) use LOD2 geometric constraints to orthorectify ground-level panoramic imagery for façade segmentation, while Ogawa et al. (2024) combine omnidirectional street images with GIS building footprints and aerial photographs to generate textured LOD2–3 models. Although both methods enrich façade appearance, neither recovers façade geometry such as depth, recesses, or surface relief. Wysocki et al. (2024) extend refinement by incorporating mobile laser scanning (MLS) point clouds to detect façade recesses but require both dense LiDAR data and a pre-existing LOD2 model.

From-scratch approaches first reconstruct the building geometry before recovering façade information. Pantoja-Rosero et al. (2022) generate an LOD2 model from multi-view imagery and subsequently derive LOD3 by projecting detected façade openings onto reconstructed surfaces. However, the approach is limited to flat, rectangular opening extraction and does not recover more complex façade geometry. Unlike refinement-based methods, Omni2LOD3 first generates an LOD2 base model from nadir UAV imagery before façade enrichment. Table 1 summarizes the methodological differences between previous approaches and the present study.

2.2 Monocular Depth Estimation for Façade Geometry

Traditional image-based reconstruction methods such as SfM and MVS require many overlapping images and known camera poses, making them impractical when building façades are visible in only one or two panoramas (Huang et al., 2025). Recent advances in zero-shot MDE address this by estimating relative depth from a single uncalibrated image without requiring camera intrinsics or pose estimates, making models such as *Depth Anything* well suited to sparse acquisition scenarios (Lin et al., 2025). Depth recovery from omnidirectional imagery has also been investigated. Saura-Herreros et al. (2021) estimate depth from spherical panoramas using cubemap projections and a user-guided optimization algorithm, demonstrating that depth can be recovered from omnidirectional imagery without specialized hardware. These studies show that omnidirectional imagery can support depth estimation, but do not focus specifically on façade geometry or integrate the output into a 3D building reconstruction workflow.

Omni2LOD3 applies MDE directly to sparse omnidirectional images to recover façade geometry, including recesses and building installations, without requiring known camera poses or dense image coverage. The UAV-derived LOD2 model serves as the georeferencing reference frame, eliminating the need for additional ground control points.

2.3 Color-Based Segmentation of Façade Point Clouds

Façade segmentation methods generally operate either on images or point clouds. Image-based approaches commonly use CNNs to detect façade components such as windows and doors (Tang et al., 2025; Wysocki et al., 2024), but CNN performance depends on training data robustness and these models can struggle with non-standard architectural elements outside the training distribution (Bello et al., 2020). Point-cloud-based methods typically rely on geometric features derived from dense LiDAR data (Gruen et al., 2019; Wysocki et al., 2024).

Image-derived point clouds differ from LiDAR in that they retain RGB color information but are typically sparse and irregular, limiting the effectiveness of geometry-based classification. Color-based segmentation has been proposed as an alternative that leverages this radiometric information directly (Zhan et al., 2009), and HSV thresholding provides a more robust basis for outdoor applications than RGB segmentation by isolating hue from brightness and saturation, making classification less sensitive to shadows and illumination changes (Kang et al., 2021; Smith, 1978). This approach has been applied to filter vegetation from photogrammetric point clouds and isolate structural elements in outdoor settings (Núñez-Andrés et al., 2021), motivating its use here as a baseline for semantic segmentation of sparse image-derived façade point clouds.

3. Methodology

The Omni2LOD3 workflow integrates nadir drone photogrammetry and a limited number of ground-level omnidirectional images to generate LOD3-ready geometric and semantic components. As illustrated in Figure 2, the pipeline first reconstructs an LOD2 base model from UAV imagery, then extracts and semantically segments façade points from omnidirectional images. Each stage is designed to operate with limited data, prioritizing accessible acquisition and user-guided processing over approaches that require dense inputs or extensive annotation. Together, these outputs establish the geometric and semantic foundation for CityGML LOD3 reconstruction within the broader Omni2LOD3 framework.

3.1. Study Area

The study was conducted within a university campus in the Philippines. A multi-story academic building was selected as the primary development site. The building contains both simple planar façades and complex architectural features such as curved wall segments, recessed elements, and partial occlusions from surrounding vegetation. These characteristics provide a realistic test case for façade extraction compared to many existing LOD3 studies, which often evaluate buildings with relatively simpler geometries, characterized by planar surfaces and minimal occlusions (Harshit et al., 2024; Pantoja-Rosero et al., 2022; Tang et al., 2025; Wysocki et al., 2024).

3.2. Data

UAV and omnidirectional imagery were the two primary datasets used in this study. Nadir images, captured using a DJI Mini 3 Pro drone, were used to reconstruct the initial building geometry. These images provide high-resolution observations of roof structures required for LOD2 generation. Ground-level omnidirectional images were captured using an Insta360 X4 camera mounted approximately 1.3 m above ground. Each image provides a full 360° field of view, allowing large portions of façades to be observed from relatively few locations. A total of 13 images were collected around the building, with positions selected based on visibility rather than fixed spacing. This corresponds to approximately one omnidirectional capture every 13–15 meters along the building perimeter. Given the building's approximate 176-meter perimeter, this represents a sparse acquisition setup compared to typical multiview stereo requirements. A SLAM-based terrestrial LiDAR dataset of the study building, collected by Ingles et al. (2024) using a Foxtech Simultaneous Localization and Mapping (SLAM) 100 scanner, was used for accuracy assessment. It was selected as the geometric reference due to its dense, high-accuracy 3D measurements, making it a well-established benchmark for evaluating image-derived point clouds (Wang, 2013). Although SLAM-based LiDAR provides slightly lower absolute precision than static TLS, it offers sufficient density and structural reliability for façade-scale benchmarking. Architectural plans were also used as supplementary references for façade dimensions, element counts, and structural features.

3.3 LOD2 Model Generation

Nadir Point cloud generation. Nadir UAV images were processed in Pix4Dmapper using Structure from Motion (SfM). This method reconstructs scene geometry by identifying tie points across overlapping images, estimating camera poses, and generating a dense point cloud through bundle adjustment (Westoby et al., 2012). SfM was selected for its well-documented reliability with UAV imagery, where high image overlap and consistent viewing angles support accurate tie-point matching and dense reconstruction. The resulting point cloud provides detailed observations of roof structures and upper building surfaces.

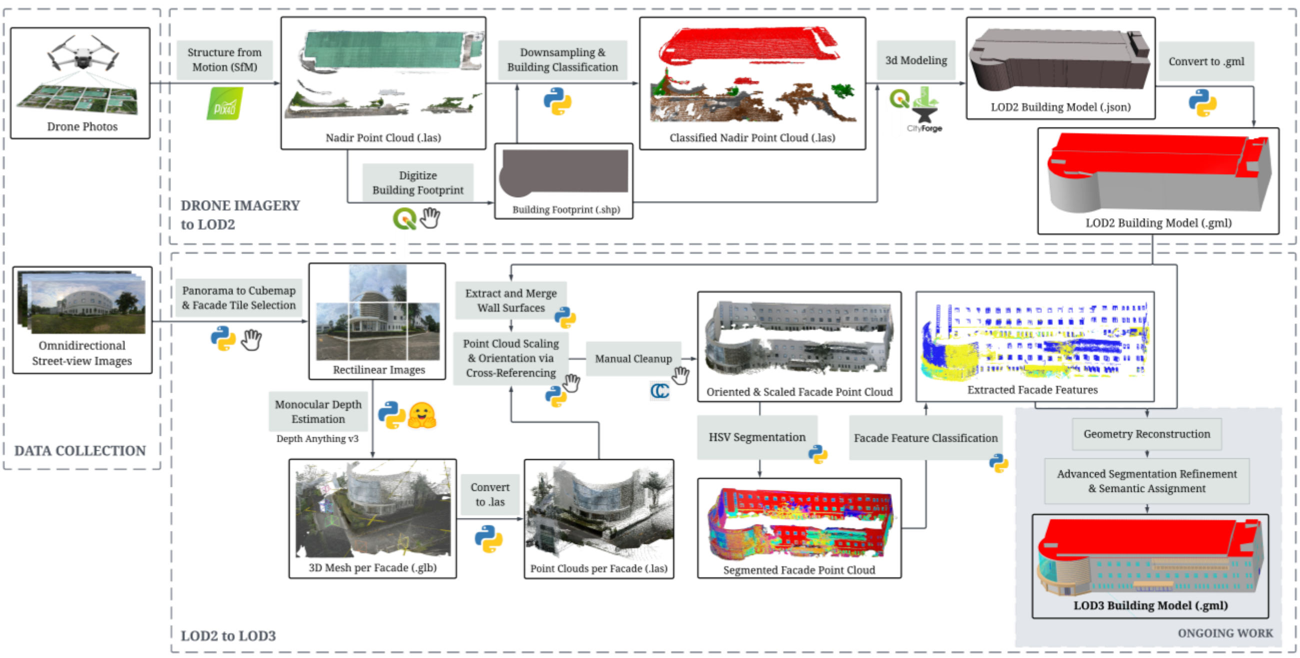

Figure 2. Overview of the Omni2LOD3 pipeline

3.3.1 Point cloud preprocessing. Since the dataset consists only of nadir imagery, automated building-point classification is unreliable. Existing classification methods rely on vertical surface detection and spatial relationships, which are difficult to infer from nadir-only data (Yastikli and Cetin, 2021; Zhang and Zhang, 2017). Manual digitization of the building footprint in QGIS was therefore adopted, using the roof point cloud and the UAV orthomosaic as a reference. This approach was preferred over automated methods given the nadir-only acquisition setup, where automated segmentation would introduce greater uncertainty than controlled user-guided extraction. The digitized polygon was then used as a spatial mask to isolate building points using a custom Python script. This served a dual purpose, enabling building-point classification and providing the base geometry for LOD2 reconstruction.

3.3.2 LOD2 model reconstruction. The classified point cloud and footprint were processed using the CityForge plugin for QGIS, which reconstructs building envelopes by detecting roof planes and generating wall surfaces (Oslandia, n.d). CityForge was selected for its streamlined, user-accessible workflow for LOD2 reconstruction within a GIS environment, consistent with the pipeline's low-cost design. The resulting model is generated in CityJSON format by default. To enable compatibility with downstream processing and CityGML-based reconstruction, the model was converted to CityGML using the citygml4j conversion toolkit (Nagel and Deininger, 2026).

3.4 Façade Geometry and Semantic Feature Extraction

The second stage of the workflow extracts façade-level geometric and semantic information from omnidirectional imagery, as a preliminary step toward LOD3 reconstruction. The output of this stage is a semantically segmented point cloud containing windows, doors, and building installations.

3.4.1 Façade extraction. Omnidirectional images were converted into planar images using the panorama-to-cubemap tool by Crane (2023), which applies Lanczos interpolation (Zhang et al., 2020) to preserve image quality during reprojection. Cubemap projection was adopted because monocular depth estimation models assume standard perspective projection. Converting to rectilinear tiles eliminates the radial distortions inherent in equirectangular formats that would otherwise degrade depth estimation accuracy. The program allows perspective tilting, ensuring each façade is captured from ground to roof in a single tile, improving depth stitching. Given the sparse dataset, manual tile selection was preferred to reduce uncertainty in façade boundary definition.

3.4.2 Façade point cloud generation. Façade geometry was reconstructed using Depth Anything v3: Recovering the Space from Any Views (Lin et al., 2025). This monocular depth estimation model was selected for its ability to infer scene geometry from images without requiring known camera poses or image overlap, making it well-suited to the sparse, uncalibrated acquisition setup. Each façade was processed individually to maintain geometric consistency within each tile, producing a sparse mesh in an arbitrary coordinate system. The mesh (GLB) is then converted to LAS format, followed by manual cleaning to remove non-façade points. This process produces point clouds representing the geometric structure observed in the omnidirectional imagery.

3.4.3 Point cloud georeferencing. Since the façade point clouds are generated in arbitrary coordinate systems, the LOD2 model was used to align them to a common coordinate system. This approach was adopted because the LOD2 model already occupies a consistent coordinate system derived from the nadir UAV reconstruction, avoiding the need for additional control points or external survey data and maintaining the low-cost design of our workflow. A GUI-based program allows users to select corresponding points between the LOD2 model and each façade point cloud, as visualized in Figure 3. A minimum of three points is required, with snapping to edges and corners to reduce user error.

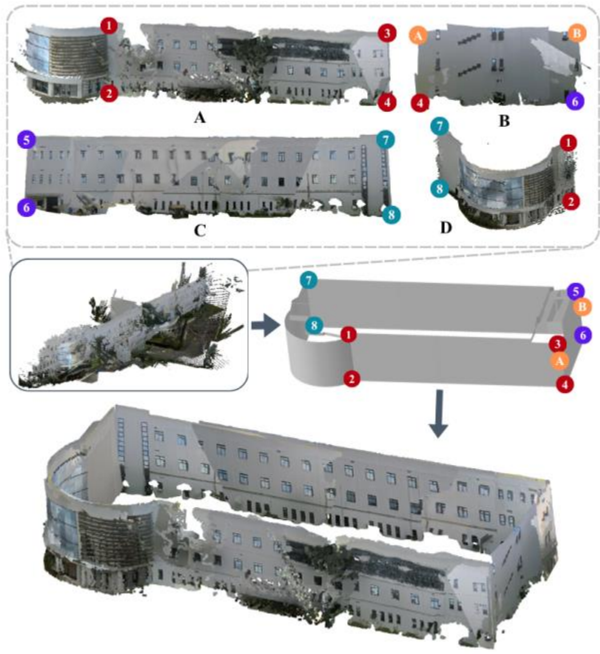

Figure 3. Georeferencing and alignment of façade point clouds

The transformation is computed via the Kubsch-Umeyama algorithm (Lawrence et al., 2019), which estimates scale (s), rotation matrix (R), and translation vector (T) parameters to minimize the root-mean-squared deviation between the ground coordinates (Y) and the arbitrary coordinates (X) as shown in equation 1.

$$Y = sRX + T \tag{1}$$

3.4.4 Semantic feature extraction. A Hue, Saturation, and Value (HSV)-based segmentation approach was applied to differentiate façade elements. Given the sparse and image-derived nature of the façade point cloud, conventional LiDAR-based or deep learning-based semantic segmentation approaches are not directly applicable. HSV-based clustering was therefore adopted as a lightweight and interpretable baseline for semantic extraction under constrained data conditions. Unlike LiDAR-based or learning-based approaches that rely on dense point clouds or annotated training data, this method operates directly on image-derived point clouds with minimal preprocessing. In this context, HSV-based clustering provides a geometry-independent semantic cue that remains stable even when surface normals and planar features are unreliable in sparse, monocular depth-derived façade point clouds.

HSV color space was used instead of RGB to reduce sensitivity to illumination and shadow variations, since façade materials vary primarily in hue and saturation rather than brightness. Hue is defined as the angular position of a color on the color wheel. Saturation is derived from the normalized difference between the maximum and minimum RGB values, indicating the color intensity. Value represents brightness and is taken as the maximum of the RGB components. The formulas for HSV conversion are given by Equations 2 to 4.

$$H = 60°\text{C} \times \frac{G - B}{\max(R,G,B) - \min(R,G,B)} \times (mod6) \quad (2)^{1}$$

$$S = \frac{\max(R,G,B) - \min(R,G,B)}{\max(R,G,B)} \quad (3)$$

$$V = \max(R,G,B) \quad (4)$$

After conversion, K-Means clustering (Sinaga and Yang, 2020) groups points with similar HSV properties. The user then assigns CityGML semantic classes (WallSurface, Door, Window, and Building Installation) onto the resulting clusters. WallSurface points were discarded, as they are already represented in the LOD2 model.

4. Results and Discussion

The intermediate outputs of Omni2LOD3 were evaluated through two complementary assessments, both oriented toward establishing the suitability of these outputs as inputs to the subsequent LOD3 reconstruction stage. The first involved geometric comparison of the omnidirectional image-derived façade point cloud against a SLAM-based terrestrial LiDAR reference. The second involved semantic evaluation of the segmented façade elements against a manually prepared benchmark.

4.1 Reference Data

The LiDAR point cloud shown in Figure 4 served as the primary geometric reference for evaluating façade point cloud quality.

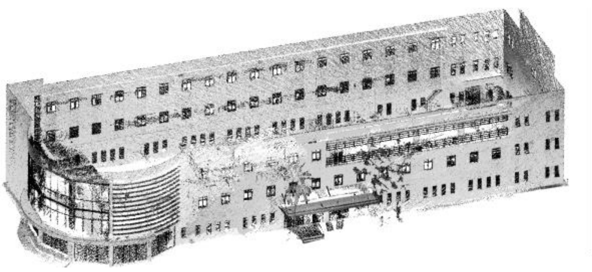

Figure 4. SLAM-based Terrestrial LiDAR point cloud of the study building.

For semantic evaluation, the benchmark model of the building was manually constructed using SketchUp and CityEditor, following a point-cloud-based LOD3 modeling workflow adopted from Technical University of Munich (2023). Manual construction ensured accurate representation of complex architectural features, including curved surfaces and recessed elements, that automated methods may not reliably capture. This benchmark, shown in Figure 5, served as the reference for identifying elements and checking the correspondence of extracted semantic classes. Architectural plans were additionally consulted to verify element counts and façade dimensions.

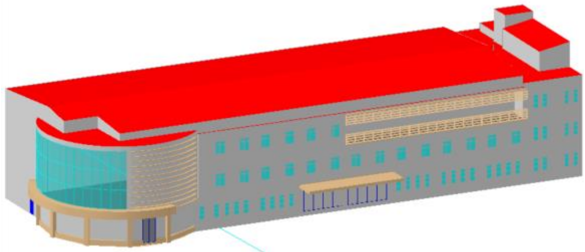

Figure 5. Manually constructed benchmark model

4.2 Geometric accuracy assessment

Geometric quality was evaluated using Cloud-to-Cloud (C2C) distance and Multiscale Model-to-Model Cloud Comparison (M3C2). The C2C metric provides a direct measure of point-wise proximity between the image-derived façade point cloud and the LiDAR reference, while M3C2 evaluates local surface agreement along estimated normals, enabling a more surface-aware comparison of the two point clouds (Kharroubi et al., 2022). The per-façade results of the C2C and M3C2 analyses are presented in Table 2.

Façade	C2C Mean distance	C2C Standard Deviation	M3C2 Distance
Front (South)	0.40 m	0.33 m	0.10 m
Back (North)	0.39 m	0.29 m	0.13 m
East	0.53 m	0.17 m	0.07 m
West	0.54 m	0.37 m	0.009 m
Overall	0.46 m	0.31 m	0.08 m

Table 2. C2C and M3C2 distances between omnidirectional and LiDAR-derived point clouds

The overall mean C2C distance of 0.46 m indicates moderate geometric agreement at the building scale. For LOD3 façade structuring, where planar surface consistency is more critical than millimeter-level precision, this level of deviation is sufficient to support geometric reconstruction of main façade planes, overhangs, and wall offsets. The mean M3C2 distance of 0.08 m further indicates that planar surfaces are generally well-captured, implying that most C2C discrepancies arise from fine-scale architectural detail rather than from systematic geometric errors. However, due to occlusions and limitations in camera placement, some areas were not fully reconstructed. For example, the upper portion of the east façade is missing in the image-derived reconstruction relative to the LiDAR reference due to occlusions from trees.

4.3 Semantic accuracy assessment

The K-Means-based HSV clustering semantic segmentation approach was evaluated through comparison with the benchmark model. Since the data is represented as point clouds, façade elements were quantified by clustering corresponding to individual detected features. Detection performance was measured by the correspondence between predicted and actual classes, while classification performance was evaluated using precision, recall, and F1 Score.

$$Precision = \frac{TP}{TP + FP}; \; Recall = \frac{TP}{TP + FN} \quad (5)$$

[1] The hue formula is if Red is the maximum value in R,G,B

$$F1\,Score = 2 * \frac{Precision * Recall}{Precision + Recall} \tag{6}$$

where TP represents correctly identified elements, FP represents incorrectly classified elements, and FN represents undetected reference elements.

Predicted point labels were compared with ground-truth classes derived from manually segmenting the Omni2LOD3-derived point cloud. The resulting confusion matrix is shown in Table 3, and the derived performance metrics are summarized in Table 4.

		Predicted		
		Window	Door	Installation
Actual	Window	171507	5367	11154
	Door	10878	2153	11244
	Installation	34612	23567	53866

Table 3. Confusion matrix of façade element classification.

Class	TP	FN[2]	Precision	Recall	F1-score
Window	171507	16521	0.79	0.91	0.85
Door	2153	22122	0.07	0.09	0.08
Installation	53866	58179	0.71	0.48	0.57

Table 4. Detection and classification performance metrics for façade elements

The overall accuracy of the HSV-based semantic segmentation is 70% with the highest accuracy for windows and lowest for doors. Doors are located at ground level, where occlusions are most present. Further, while point-based accuracy for each class is low to moderate, instance-based accuracy is high, as shown in Figure 6, where façade elements are differentiated with windows, building installations, and doors represented as distinct point clusters. Thirty-two windows and two doors were not reconstructed in the point cloud, primarily due to occlusions and limited façade visibility.

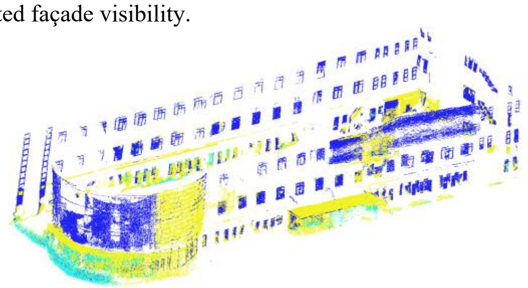

Figure 6. HSV-based semantic segmentation results, showing extracted façade elements including windows (blue), building installations (yellow), and doors (cyan).

Semantic performance is strongly influenced by visual distinctiveness and color separability. Window elements, which exhibit consistent color contrast and regular spatial patterns, achieved high classification performance despite moderate recall. Building installations showed strong detection performance but lower precision, indicating partial confusion with adjacent façade elements. Door elements were more difficult to recover due to limited visual contrast and their location near ground level, where occlusions, vegetation, and reduced point density affect reconstruction. Furthermore, the building design makes it difficult to distinguish between doors and building installations, since the two are metal grills following the same pattern. Following minor manual refinement to reassign misclassified points, the completeness of instance-based feature detection reached 81%, with the remainder being undetected elements. Overall, the results show that HSV-based segmentation can recover meaningful façade elements, although performance remains class-dependent and influenced by visibility, geometry, and color similarity.

4.4 Discussion

The results support the central premise of this study. A limited number of omnidirectional images can already provide meaningful façade geometry and semantic information, even prior to full LOD3 reconstruction. This is significant because omnidirectional imagery offers a wide field of view, enabling large portions of a building façade to be captured from relatively few acquisition points. Such efficiency is advantageous not only for field acquisition but also for future workflows that rely on sparse or publicly available street-level imagery. Further, we find that a limited number of omnidirectional images provide sufficient data for a realistic test case involving a building with complex geometry and partial occlusions.

At the same time, several limitations remain. Image-derived façade point clouds remain less geometrically precise than LiDAR references, particularly at fine scales. Semantic performance is class-dependent, with stronger results for visually distinct elements such as windows and weaker performance for classes with similar color appearances. These limitations are consistent with early-stage workflows that rely on sparse, image-derived geometry.

The choice of HSV-based segmentation should be interpreted in the context of the input data. Unlike LiDAR-derived façade point clouds, which support geometry-driven segmentation through surface-normal and planar constraints, point clouds reconstructed from sparse omnidirectional imagery exhibit uneven density and local geometric distortions due to monocular depth estimation. Under these conditions, a simple and interpretable color-based approach provides a practical baseline for semantic extraction. This approach remains a minimum viable option for reducing manual work and can be extended with more advanced classification strategies tailored to image-derived façade point clouds. Future work can build on this foundation by incorporating more advanced classifiers, contextual rules, or learning-based approaches tailored to image-derived façade point clouds.

Most importantly, the outputs presented in this study constitute the LOD3-ready geometric and semantic components required for full CityGML reconstruction. Together, the LOD2 base model and the semantically segmented façade point cloud provide the geometric and semantic foundation for CityGML-compliant LOD3 reconstruction. Semantic readiness, rather than complete semantic reconstruction, is proposed as an appropriate and achievable target for early-stage workflows that operate under constrained data acquisition, with the full reconstruction pipeline actively under development. These results suggest that semantic readiness represents a practical intermediate stage between LOD2 reconstruction and full LOD3 completion in image-based urban modeling workflows operating under constrained acquisition conditions.

Several stages of the workflow currently require user input, including building footprint digitization, façade tile selection, point cloud cleaning, and georeferencing. While these manual

[2] False Negatives include undetected features due to occlusions.

steps introduce some user dependency, they also provide flexibility when handling irregular geometries and sparse façade reconstructions that remain difficult to fully automate. This semi-automated nature is nonetheless a design characteristic rather than a shortcoming, the pipeline balances automation and user control, remaining significantly less resource-intensive than LiDAR-based or fully dense multiview approaches. Future work may focus on increasing automation in alignment and segmentation, particularly as methods for handling sparse, image-derived point clouds continue to develop.

This positions Omni2LOD3 as a viable early-stage approach for LOD3 reconstruction in data-scarce environments, where traditional dense acquisition workflows are not feasible. Rather than treating sparse acquisition as a limitation, Omni2LOD3 reframes it as a design constraint that still allows LOD3-ready reconstruction.

5. Conclusions and Ongoing Work

We demonstrate that LOD3-ready façade geometry and semantic information can be derived from sparse omnidirectional imagery without LiDAR or dense multi-view image capture. Omni2LOD3 is a replicable strategy that generates LOD3-ready components from accessible image-based inputs, specifically an LOD2 base model and a semantically segmented, geometrically representative façade point cloud, which can support subsequent LOD3 refinement. In doing so, the study challenges the conventional assumption that dense acquisition is required for LOD3 reconstruction.

Geometric evaluation against SLAM terrestrial LiDAR confirms that the reconstructed façade point clouds capture the overall building form, including overhangs and balconies, even when generated from a limited number of omnidirectional images through monocular depth estimation. Semantic evaluation shows that HSV-based segmentation achieves an overall accuracy of 70%, and minor manual refinement raises this to 81% detection completeness. Semantic performance was class-dependent: windows were reliably detected due to their consistent color contrast, whereas doors were difficult to recover because of their visual similarity to surrounding building installations and ground-level occlusions.

The results presented in this paper support semantic readiness as a practical target for early-stage, low-cost LOD3 workflows. In this context, the HSV-based method should be understood as a lightweight baseline for semantic extraction from sparse image-derived façade point clouds. More advanced image-based or learning-based classification methods may later be integrated into the broader Omni2LOD3 workflow to improve performance on visually ambiguous or architecturally complex façade elements. These findings open the possibility for scalable urban modeling workflows in settings where dense acquisition or LiDAR remains inaccessible.

5.1 Ongoing Work

The following describes work currently in progress and not yet evaluated in this paper, as illustrated in Figure 7. Current work focuses on completing the reconstruction pipeline by combining the LOD2 model with the semantically segmented façade point cloud to produce a CityGML 3.0-compliant LOD3 model. This involves deriving semantic surfaces and planar features from the generated point clouds, clustering semantically segmented points into façade elements, generating geometric proxies for openings and installations, and integrating these with the LOD2 envelope.

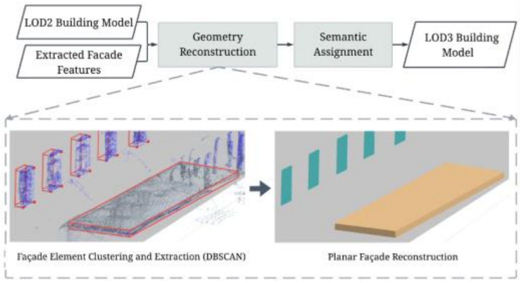

Figure 7. Visualization of the ongoing Omni2LOD3 reconstruction pipeline

A particular proposed innovation in Omni2LOD3 is the handling of curved façade regions through tessellation or segmented approximation prior to feature reconstruction. In this way, the present study serves as a foundation for a more complete end-to-end image-based workflow that progresses from accessible data acquisition toward more detailed and semantically meaningful 3D building models.

Building on this, ongoing work also focuses on improving semantic extraction through more advanced image-based and learning-based segmentation methods. In this context, the HSV-based approach serves as an initial step for semantic extraction, providing a simple and interpretable basis for façade segmentation. More advanced methods are expected to improve performance for visually ambiguous and architecturally complex façade elements.

Acknowledgements

This research was done as part of the research project KLIMA: A Coupled Digital Twin and Urban Climate Modeling System for Smart and Resilient Urban Environment implemented by the University of the Philippines Training Center for Applied Geodesy and Photogrammetry (UP TCAGP), through the support and funding of the Department of Science and Technology (DOST) of the Republic of the Philippines and the Philippine Council for Industry, Energy, and Emerging Research and Development (PCIEERD) with Project No. 1213106, 2024. This work is also supported by the PhD Incentive Award (252515 PhDIA Y2) from the Office of the Vice Chancellor for Research and Development (OVCRD), University of the Philippines Diliman.

References

Bello, S.A., Yu, S., Wang, C., Adam, J.M., Li, J., 2020. Review: Deep Learning on 3D Point Clouds. Remote Sensing 12. https://doi.org/10.3390/rs12111729

Biljecki, F., Ledoux, H., Stoter, J., 2016. An improved LOD specification for 3D building models. Computers, environment and urban systems 59, 25–37. https://doi.org/10.1016/j.compenvurbsys.2016.04.005

Crane, L., 2023. Panorama To Cubemap. GitHub.

Gruen, A., Schubiger, S., Qin, R., Schrotter, G., Xiong, B., Li, J., Ling, X., Xiao, C., Yao, S., Nuesch, F., 2019. Semantically enriched high resolution LoD 3 building model generation, in: International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives. International Society for Photogrammetry and Remote Sensing, pp. 11–18. https://doi.org/10.5194/isprs-archives-XLII-4-W15-11-2019

Harshit, Chaurasia, P., Zlatanova, S., Jain, K., 2024. Low-Cost Data, High-Quality Models: A Semi-Automated Approach to LOD3 Creation. ISPRS International Journal of Geo-Information 13. https://doi.org/10.3390/ijgi13040119

Huang, Z., Ali, M.A.H., Nukman, Y., Xu, H.L., Zhang, S., Chen, H., Alkhedher, M., 2025. A systematic review of monocular depth estimation for autonomous driving: Methods and dataset benchmarking. Results in Engineering 26, 105359. https://doi.org/10.1016/j.rineng.2025.105359

Ingles, H.A., Legaspi, A.M.M., Sarmiento, C.J.S., Claridades, A.R.C., 2024. ACCURACY ASSESSMENT OF A SLAM-ACQUIRED POINT CLOUD DATA USING A VARIETY OF CLASSIFICATION APPLICATIONS. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences 48, 307–312.

Kang, H.-C., Han, H.-N., Bae, H.-C., Kim, M.-G., Son, J.-Y., Kim, Y.-K., 2021. HSV Color-Space-Based Automated Object Localization for Robot Grasping without Prior Knowledge. Applied Sciences 11. https://doi.org/10.3390/app11167593

Kharroubi, A., Poux, F., Ballouch, Z., Hajji, R., Billen, R., 2022. Three Dimensional Change Detection Using Point Clouds: A Review. Geomatics 2, 457–485. https://doi.org/10.3390/geomatics2040025

Kutzner, T., Chaturvedi, K., Kolbe, T.H., 2020. CityGML 3.0: New Functions Open Up New Applications. PFG – Journal of Photogrammetry, Remote Sensing and Geoinformation Science 88, 43–61. https://doi.org/10.1007/s41064-020-00095-z

Lawrence, J., Bernal, J., Witzgall, C., 2019. A Purely Algebraic Justification of the Kabsch-Umeyama Algorithm. J. RES. NATL. INST. STAN. 124, 124028. https://doi.org/10.6028/jres.124.028

Lin, H., Chen, S., Liew, J., Chen, D.Y., Li, Z., Shi, G., Feng, J., Kang, B., 2025. Depth Anything 3: Recovering the Visual Space from Any Views. https://doi.org/10.48550/arXiv.2511.10647

Nagel, C., Deininger, K., 2026. citygml-tools. citygml4j.

Núñez-Andrés, M., Prades, A., Buill, F., 2021. Vegetation Filtering using Colour for Monitoring Applications from Photogrammetric Data:, in: Proceedings of the 7th International Conference on Geographical Information Systems Theory, Applications and Management. Presented at the 7th International Conference on Geographical Information Systems Theory, Applications and Management, SCITEPRESS - Science and Technology Publications, Online Streaming, --- Select a Country ---, pp. 98–104. https://doi.org/10.5220/0010523300980104

Ogawa, Y., Nakamura, R., Sato, G., Maeda, H., Sekimoto, Y., 2024. End-to-End Framework for the Automatic Matching of Omnidirectional Street Images and Building Data and the Creation of 3D Building Models. Remote Sensing 16. https://doi.org/10.3390/rs16111858

Open Geospatial Consortium (OGC), 2021. City Geography Markup Language (CityGML) Part 1: Conceptual Model Standard (OGC Doc. 20-010). Open Geospatial Consortium.

Oslandia, n.d. CityForge documentation.

Pantoja-Rosero, B.G., Achanta, R., Kozinski, M., Fua, P., Perez-Cruz, F., Beyer, K., 2022. Generating LOD3 building models from structure-from-motion and semantic segmentation. Automation in Construction 141, 104430. https://doi.org/10.1016/J.AUTCON.2022.104430

Saura-Herreros, M., Lopez, A., Ribelles, J., 2021. Spherical panorama compositing through depth estimation. Visual Computer 37, 2809–2821. https://doi.org/10.1007/s00371-021-02239-7

Sinaga, K.P., Yang, M.-S., 2020. Unsupervised K-Means Clustering Algorithm. IEEE Access 8, 80716–80727. https://doi.org/10.1109/ACCESS.2020.2988796

Smith, A.R., 1978. Color gamut transform pairs. SIGGRAPH Comput. Graph. 12, 12–19. https://doi.org/10.1145/965139.807361

Tang, W., Li, W., Liang, X., Wysocki, O., Biljecki, F., Holst, C., Jutzi, B., 2025. Texture2LoD3: Enabling LoD3 Building Reconstruction With Panoramic Images, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops. pp. 2016–2026. https://doi.org/10.48550/arXiv.2504.05249

Technical University of Munich, C. of G., 2023. LoD3 Modeling based on Mobile Laser Scans [WWW Document]. Creating CityGML Datasets. URL https://creating-citygml-datasets.readthedocs.io/en/latest/creation-guidelines/lod3-models-based-on-point-clouds.html

Vacca, G., Dessì, A., Sacco, A., 2017. The Use of Nadir and Oblique UAV Images for Building Knowledge. ISPRS International Journal of Geo-Information 6. https://doi.org/10.3390/ijgi6120393

Wang, R., 2013. 3D building modeling using images and LiDAR: a review. International Journal of Image and Data Fusion 4, 273–292. https://doi.org/10.1080/19479832.2013.811124

Westoby, M.J., Brasington, J., Glasser, N.F., Hambrey, M.J., Reynolds, J.M., 2012. 'Structure-from-Motion'photogrammetry: A low-cost, effective tool for geoscience applications. Geomorphology 179, 300–314.

Wysocki, O., Hoegner, L., Stilla, U., 2024. MLS2LoD3: Refining Low LoDs Building Models with MLS Point Clouds to Reconstruct Semantic LoD3 Building Models, in: Andreas, H, B.C.K.T., Donaubauer (Eds.), Recent Advances in 3D Geoinformation Science. Springer Nature Switzerland, pp. 367–380.

Yastikli, N., Cetin, Z., 2021. Classification of raw LiDAR point cloud using point-based methods with spatial features for 3D building reconstruction. Arabian Journal of Geosciences 14, 146.

Zhan, Q., Liang, Y., Xiao, Y., 2009. Color-based segmentation of point clouds. Laser scanning 38, 155–161.

Zhang, Liqiang, Zhang, Liang, 2017. Deep learning-based classification and reconstruction of residential scenes from large-scale point clouds. IEEE Transactions on Geoscience and Remote Sensing 56, 1887–1897.

Zhang, S., Yang, F., Wan, S., Di, P., 2020. Spherical Lanczos Interpolation in Planar Projection or Format Conversions of Panoramic Videos. IEEE Access 8, 9655–9667. https://doi.org/10.1109/ACCESS.2020.2964789