Automated Scan-to-BIM Method for City-Scale Building Roof Reconstruction Using UAV Photogrammetry and Geospatial Building Footprints

Boan Tao[1], **Frédéric Bosché** [1] **and Jiajun Li**[1]

[1]School of Engineering, University of Edinburgh, Edinburgh, Scotland, UK

boan.tao@ed.ac.uk, f.bosche@ed.ac.uk, Jiajun.Li@ed.ac.uk

Abstract -
The digital transformation of existing building stock through Building Information Modelling (BIM) remains a significant goal and challenge in the Architecture, Engineering, and Construction (AEC) industry. This paper presents an automated Scan-to-BIM framework specifically designed for generating IFC-compliant roof model of buildings within a large urban environment from Unmanned Aerial Vehicle (UAV) photogrammetry data. The proposed method leverages geospatial building footprint data as semantic priors to intelligently filter the input large-scale 3D mesh and model the roofs of all individual buildings. A robust segmentation pipeline combining Taubin smoothing, region-growing algorithms, and adaptive height filtering successfully extracts planar roof surfaces. The extracted geometries are reconstructed into Industry Foundation Classes (IFC) compliant BIM objects using alpha-shape boundary regularisation, which preserves non-convex features such as conical spires and L-shaped roofs. Experimental validation is first conducted on a large building with a complex roof structure composed of 431 extracted roof patches, for which a patch-level correctness precision of 98.8% was achieved. The framework is subsequently demonstrated at the scale of a university campus to assess scalability and robustness.

Keywords -

Scan-to-BIM, Semantic Segmentation, Industry Foundation Classes (IFC), Digital Twin, Automated Roof Extraction, Geospatial Data Fusion.

1 Introduction

1.1 Background

A substantial proportion of the existing building stock lacks accurate and up-to-date digital representations. While new construction projects increasingly benefit from *born-digital* Building Information Models (BIM), many existing buildings—particularly those constructed prior to the widespread adoption of BIM—remain poorly documented in structured digital formats. This digital deficit limits the effectiveness of facility management, maintenance, retrofitting, and long-term asset monitoring.

Roofs constitute critical building elements that require regular inspection due to their continuous exposure to environmental stressors. Conventional inspection methods are labour-intensive, costly, and often pose safety risks. In recent years, Unmanned Aerial Vehicles (UAVs) equipped with high-resolution cameras have emerged as a practical alternative, enabling rapid and safe roof inspection at scale. Contemporary photogrammetry pipelines can transform UAV imagery into dense three-dimensional representations with centimetre-level geometric accuracy.

Despite these advances, a persistent *semantic gap* remains between raw geometric reconstructions—such as images, point clouds, and meshes—and structured, semantically-rich building models suitable for BIM-based workflows. Photogrammetric outputs are inherently unstructured and lack explicit semantic information, making it difficult to distinguish architectural elements such as roofs from surrounding terrain, vegetation, or urban clutter. As a result, manual interpretation and modelling remain common practice, leading to significant time and cost overheads.

Although extensive research has addressed large-scale processing and semantic labelling of aerial LiDAR and photogrammetric data, many Scan-to-BIM workflows still focus on individual processing stages, such as surface reconstruction or classification, rather than enabling end-to-end automated generation of IFC-compliant building components. In particular, approaches capable of robustly extracting roof geometry from dense photogrammetric meshes and directly translating it into interoperable BIM representations at scale remain limited. These challenges are further amplified by the size and geometric complexity of high-resolution UAV-derived datasets, underscoring the need for scalable and fully automated methods that bridge the gap between unstructured 3D data and structured digital building models. Importantly, the limitation is not the absence of individual algorithms, but the lack of an integrated and scalable framework that can bridge unstructured 3D data and structured smenatically-rich (BIM) representations in a fully automated manner.

2 Related Work

2.1 UAV Photogrammetry for Building Reconstruction

Unmanned Aerial Vehicle (UAV) technology is revolutionising building surveying and 3D reconstruction workflows. Besides, Structure-from-Motion (SfM) and Multi-View Stereo (MVS) algorithms first enabled fully automated 3D reconstruction pipelines from UAV imagery. Recently, depth-supervised neural radiation fields have also been growing rapidly to produce navigable 3D reconstructions from images [1]. Despite these advances, the resulting reconstructions—whether expressed as dense polygonal meshes, point-based representations, or implicit radiance fields—remain largely unstructured and lack explicit semantic information, limiting their direct integration into BIM-oriented workflows.

2.2 Roof Extraction from Aerial 3D Data

When processing aerial 3D data (3D LiDAR or photogrammetrically-reconstruction point clouds and/or meshes) of built environments, region growing is a commonly adopted strategy for roof plane extraction, where planar surfaces are incrementally expanded from initial seed regions according to geometric similarity criteria. To overcome the sensitivity of traditional region-growing approaches to seed selection and boundary leakage, Li et al. [2] introduced a hierarchical roof segmentation framework for airborne LiDAR data that combines clustering-based initialisation with boundary relabelling. Their method refines roof plane boundaries through multi-level segmentation, improving robustness in separating adjacent roof facets and handling complex roof structures.

Chen et al. [3] proposed a contour extraction method for UAV-derived point clouds that combines multi-level region-growing plane segmentation and plane fusion, followed by neighbourhood geometric feature analysis on the fused planes to robustly extract both roof and façade (wall) contours, including sparse structural boundaries.

Overall, existing plane segmentation strategies primarily rely on iterative grouping mechanisms—such as region growing, RANSAC-based fitting, or clustering—to aggregate coplanar points based on local geometric similarity. While these approaches are effective for extracting dominant planar structures, their performance is often sensitive to parameter selection, point density variations, and surface completeness, which can lead to boundary leakage or over-segmentation in complex roof and façade geometries.

2.3 Geospatial Data Fusion and Building Footprint Utilisation

The integration of geospatial building footprint data with 3D point clouds provides an effective means for spatial filtering and semantic guidance in large-scale urban reconstruction tasks. While numerous studies focus on extracting building footprints from remote sensing data, many practical applications——particularly in urban environments——benefit from the availability of pre-existing footprint datasets derived from cadastral maps or open geospatial repositories. Previous studies have demonstrated that integrating aerial imagery with LiDAR-derived elevation information can significantly improve building footprint delineation in dense urban environments [4, 5].

In contrast to these extraction-oriented approaches, the present work aims to leverage readily-available accurate geospatial footprint data directly as semantic priors to process new aerial 3D data (UAV photogrammetry). By using footprints to constrain spatial queries and filter large-scale UAV-derived meshes, the proposed framework efficiently isolates each individual building and corresponding roof regions, providing a reliable foundation for subsequent segmentation and IFC-compliant BIM reconstruction. The benefits are: (1) reduced computational demand (building-by-building processing reduces computational overload and removes the need for merging processes in tiling-based processing strategies); (2) enhanced quality where roofs are not searched in areas outside building footprints, reducing the risk of false positives.

2.4 Scan-to-BIM Methodologies

Automated Scan-to-BIM pipelines have been widely investigated, particularly for converting terrestrial laser scanning or indoor point cloud data into IFC-compliant BIM models. For example, Cloud2BIM [6] introduced a fully automated pipeline for reconstructing building elements from large-scale point clouds, while BridgeTwin [7] proposed a semi-automated workflow tailored to bridge infrastructure. They are predominantly designed for ground-based acquisition scenarios and structured geometries, relying on interior layouts, predefined templates, or significant user interaction. As such, they are not directly applicable to UAV-derived photogrammetric meshes.

2.5 Research Gap and Contribution

While significant progress has been made in individual elements of the Scan-to-BIM pipeline, several critical gaps remain:

1. **Geometric Complexity**: Many existing roof modelling and Scan-to-BIM approaches operate on sparse or abstracted representations, such as planar segments derived from aerial LiDAR or simplified point clouds. As a result, these methods typically favour simplified parametric roof models (e.g., gable or hip roofs), limiting their ability to represent complex roof morphologies, including conical spires, L-shaped roofs, and non-convex boundaries. Leveraging dense UAV-derived photogrammetric meshes offers the potential to capture roof geometry at a higher level of detail, but scalable methods capable of exploiting this richness while remaining fully automated remain limited.

2. **BIM-compliant Output**: Few automated pipelines generate IFC-compliant models preserving non-convex geometries suitable for downstream digital twin applications, with most focusing on simple primitives.
3. **End-to-End Automation**: Most workflows require manual intervention for quality control, parameter tuning, or geometric editing, preventing deployment at scale across large building portfolios.

This work explores a fully automated Scan-to-BIM framework for large-scale reconstruction of building roofs from UAV photogrammetry. Its main contribution is the use of geospatial building footprints as semantic priors to guide instance-level extraction of roof structures from dense photogrammetric meshes. This footprint-guided design improves the efficiency and robustness of roof reconstruction in complex urban environments and enables the automatic generation of IFC-compliant roof models, including non-convex geometries, without manual intervention.

3 Method

The proposed framework is designed as an end-to-end automated Scan-to-BIM pipeline that processes dense UAV-derived photogrammetric meshes together with 2D geospatial building footprints as semantic priors, in order to directly generate IFC-compliant roof models without manual intervention.

By leveraging geospatial footprint data to guide building-level segmentation, the method enables efficient and scalable processing of large-scale urban scenes, bridging the gap between unstructured 3D reconstructions and structured BIM representations.

The method works in three stages:

1. Building Photogrammetric Data Segmentation,
2. Planar Primitive Segmentation, and
3. Automated IFC Reconstruction.

3.1 Building Photogrammetric Data Segmentation

The proposed framework operates on two heterogeneous but spatially consistent data sources: a dense 3D photogrammetric mesh generated from UAV imagery and a set of 2D geospatial building footprints. Both datasets are georeferenced within a common projected coordinate system, as UAV photogrammetry workflows typically employ RTK-GNSS positioning and Ground Control Points (GCPs) to ensure accurate global geolocation.

Given this shared spatial reference, no additional geometric alignment between the photogrammetric mesh and the footprint data is required. Instead, the building footprints are directly exploited as semantic priors to spatially filter the mesh, enabling the isolation of roof-level geometry corresponding to known building extents while discarding unrelated scene elements such as terrain, vegetation, and surrounding infrastructure.

The input photogrammetric mesh is denoted as

$$M = (V, F),$$

where $V = \{v_1, v_2, \ldots, v_n\}$ represents the set of mesh vertices and $F = \{t_1, t_2, \ldots, t_m\}$ denotes the set of triangular faces. Each vertex $v_i \in V$ is defined in a Cartesian coordinate system consistent with the georeferenced UAV reconstruction. Each face $t_i \in F$ is associated with a surface normal vector $\vec{n}_i$ and a local adjacency neighbourhood $\mathcal{N}(i)$ defined by shared edges.

The building footprint dataset consists of a set of polygonal geometries

$$P = \{P_b\},$$

where each polygon P_b represents the known footprint of a building of interest b and is defined in a global geodetic reference frame, reprojected as necessary to ensure coordinate consistency with the mesh.

For each building b, its corresponding footprint polygon P_b is used to segment the overall UAV-derived mesh and extract the subset of geometry associated with that building. A mesh vertex $v_i \in M$ is retained if and only if its orthogonal projection onto the horizontal plane lies within the footprint polygon P_b:

$$M_b = \left\{ v_i \in M \;\middle|\; \mathrm{Contains}\big(P_b, v_i^{xy}\big) \right\}.$$

This footprint-guided cropping step removes non-building geometry at an early stage, substantially reducing data volume and computational complexity prior to subsequent geometric segmentation. By constraining processing to semantically relevant regions, the approach improves robustness and scalability when applied to large-scale UAV photogrammetry datasets.

3.2 Planar Primitive Segmentation

For each building of interest b, individual roof panels are extracted from the cropped mesh M_b using a multi-stage segmentation pipeline designed to be robust to noise, irregular sampling, and complex roof morphologies.

3.2.1 Geometric Smoothing and Decimation

UAV photogrammetric meshes typically contain high-frequency noise that can adversely affect surface normal estimation and segmentation stability. To suppress high-frequency noise in the photogrammetric mesh, Taubin smoothing [8], a non-shrinking signal-processing–based filtering method, is applied ($\lambda = 0.5$, $\mu = -0.53$, $N = 20$).

The selected parameters were chosen to suppress high-frequency noise while maintaining sharp roof edges and planar integrity. The smoothing strength (λ) controls noise attenuation, while the negative scaling factor (μ) compensates for contraction effects introduced by the Laplacian operator. Performing $N = 20$ iterations was empirically found to provide sufficient noise reduction for stable surface normal estimation without over-smoothing architectural features critical for subsequent planar segmentation.

Subsequently, mesh simplification is applied using Quadric Error Metric (QEM) decimation to reduce geometric complexity while preserving roof morphology. Rather than enforcing a fixed target triangle count, the level of decimation is adaptively determined based on the footprint area of each building. Specifically, the target triangle budget N_b for building b is defined proportionally to its footprint area A_b as

$$N_b = \lceil \alpha \cdot A_b \rceil ,$$

where α is a global density parameter controlling the number of retained triangles per unit area. This adaptive allocation ensures that larger buildings retain proportionally higher geometric resolution than smaller ones, while maintaining consistent surface fidelity across the dataset.

QEM decimation proceeds by iteratively collapsing edges while minimising the squared geometric error between the original and simplified surfaces, thereby preserving salient geometric features such as planar roof facets and sharp ridgelines [9]. By reducing mesh density in a controlled manner, this adaptive simplification significantly improves computational efficiency in subsequent processing stages, particularly during surface normal estimation and region-growing segmentation, without compromising the geometric fidelity required for accurate roof plane extraction.

3.2.2 Region Growing Segmentation

Planar roof segments are identified using a region-growing algorithm driven by surface normal consistency. Starting from an initial seed triangle t_i, a region is iteratively expanded to adjacent triangles $\{t_j \mid j \in \mathcal{N}(i)\}$ if the angular deviation between their surface normal vectors satisfies the following planarity criterion:

$$\arccos(\vec{n}_i \cdot \vec{n}_j) < \theta_{\text{th}},$$

where $\vec{n}_i$ and $\vec{n}_j$ denote the unit normal vectors of triangles t_i and t_j, respectively. The threshold θ_{th} is set to 10°, empirically selected to balance robustness to minor surface undulations against effective separation of adjacent roof facets with distinct orientations.

This process yields a set of planar roof regions

$$\mathcal{R}_b = \{\mathcal{R}_b^1, \mathcal{R}_b^2, \ldots, \mathcal{R}_b^K\},$$

where each region $\mathcal{R}_b^k$ consists of one or more connected triangles representing an individual roof panel of building b.

3.2.3 Adaptive Semantic Filtering via Height Analysis

To discriminate true roof surfaces from vertical façades, ground artifacts and secondary structures such as balconies, an adaptive height-based semantic filtering strategy is employed. Unlike fixed global height thresholds, which are unsuitable for scenes with varying terrain elevation, this approach dynamically adapts to the vertical extent of each individual building.

For each building b, a relative height threshold is defined as a fixed proportion of its vertical span:

$$z_b^{\text{th}} = z_b^{\max} - 0.3 \left(z_b^{\max} - z_b^{\min} \right),$$

where $z_b^{\max}$ and $z_b^{\min}$ denote the maximum and minimum elevations of all mesh vertices associated with building b, respectively.

Each segmented roof region $\mathcal{R}_b^k$ whose centroid elevation $z_c < z_b^{\text{th}}$ is classified as a non-roof element and discarded. This adaptive criterion ensures that only the primary roof structure is retained, even in the presence of multi-level buildings, terraces, or uneven terrain.

3.3 Automated IFC Reconstruction

The final stage converts the validated planar roof segments into georeferenced, IFC-compliant BIM entities without manual intervention.

3.3.1 Plane Estimation via Principal Component Analysis

For each retained roof patch P_b^k, an optimal planar representation is estimated using Principal Component Analysis (PCA). The covariance matrix of the patch vertices is computed, and the eigenvector corresponding to the smallest eigenvalue is selected as the surface normal $\vec{n}_k$. The patch vertices are subsequently projected onto the corresponding local 2D plane to facilitate robust boundary extraction.

3.3.2 Boundary Regularisation Using Alpha Shapes

To reconstruct complex and non-convex roof boundaries, such as L-shaped footprints or conical turrets, boundary regularisation is performed using the Alpha Shape algorithm. The algorithm operates on the Delaunay triangulation of the projected roof points, removing triangles whose circumradius exceeds a threshold α. This filtering yields polygonal boundaries that closely follow the true roof geometry while suppressing noise-induced artifacts [10].

In this work, the parameter α is set to 2.0 m, empirically selected to balance boundary smoothness and geometric fidelity. Smaller values preserve fine architectural details at the cost of boundary fragmentation, whereas larger values lead to over-smoothing and loss of concave features. The chosen value enables robust reconstruction of irregular roof boundaries while suppressing noise-induced artifacts, ensuring suitability for IFC-compliant BIM representation.

3.3.3 IFC Entity Instantiation and Georeferencing

The regularised 2D roof boundaries are reprojected into 3D space and instantiated as `IfcRoof` entities. Geometry is represented using `IfcShellBasedSurfaceModel`, which supports complex and non-manifold roof surfaces and ensures compatibility with downstream BIM applications.

Georeferencing information is inherited directly from the UAV-derived photogrammetric mesh, which is originally reconstructed in a projected coordinate system based on onboard GNSS and flight metadata. To preserve numerical precision, model geometry is maintained in a local Cartesian coordinate system, while global positioning is encoded explicitly using the `IfcMapConversion` entity. The associated `IfcProjectedCRS` defines the transformation to the global reference system, enabling consistent spatial alignment with GIS datasets and ensuring interoperability across BIM and geospatial platforms.

4 Experimental Results

The proposed framework is evaluated using UAV-derived photogrammetric data wit <2 cm GSD acquired over the University of Edinburgh Central Campus, a dense urban environment with diverse building typologies and complex roof geometries (fig. 1). To assess both accuracy and scalability, a two-stage evaluation strategy is adopted, consisting of detailed validation on a representative individual building followed by deployment across the full campus.

For single-building evaluation, the Edinburgh Futures Institute (EFI) is selected due to its complex roof topology (fig. 2a). The corresponding photogrammetric mesh comprises 11,772,857 vertices and 23,527,453 triangular faces, providing a high-resolution geometric representation suitable for detailed geometric assessment.

Then, to evaluate the performance of our method in large scale environments, the complete campus dataset (fig. 1) is processed. This dataset contains approximately 57 million vertices and 114 million triangular faces, covering an area of 0.476 km^2. The campus-scale evaluation includes multiple building types and roof configurations, enabling a robust assessment of scalability and consistency.

Figure 1. UAV-derived photogrammetric reconstruction of the University of Edinburgh Central Campus.

Table 1. Quantitative results summary.

Metric	Value
Roof Segmentation (section 4.1)	
Extracted patches	431
Precision	98.8%
Geometric Alignment (section 4.2.1)	
RMSE to BIM surfaces	0.84 m
Mean deviation	0.68 m
Min deviation	0.002 m
Max deviation	~2.0 m
Matched area (τ = 2.0 m)	99.0%
Object Consistency(section 4.2.2)	
Matched entities	270 / 292
Matching rate	92.5%
Mean consistency	0.66
Scalability(section 4.2.3)	
Mesh size	~57M V, ~114M F
Spatial extent	0.476 km^2

4.1 Roof Segmentation Results

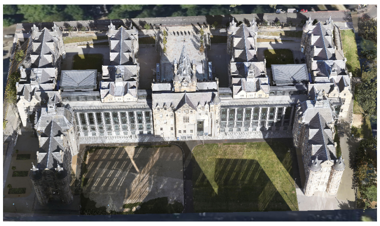
The proposed pipeline successfully processed the raw photogrammetric mesh and extracted individual roof surfaces in a fully automated manner. For the EFI building, a total of 431 unique roof patches were identified across the study area. Patch-level precision is computed by manually checking whether each extracted patch corresponds to a valid roof surface.

Qualitative inspection further indicates that the segmentation achieves a clean separation of complex roof structures into distinct planar elements. Non-roof components, including ground terrain, facades, balconies and surrounding vegetation, were effectively removed during the semantic filtering stage, leaving only architecturally relevant roof surfaces.

As illustrated in fig. 2b, the resulting segmentation provides a noise-free and topologically coherent representation of the complex building roof. Individual roof panels are clearly delineated, which forms a suitable basis for subsequent geometric regularisation and IFC-based BIM reconstruction.

4.2 Geometric Accuracy Evaluation

The evaluation in this study is conducted at three complementary levels. First, patch-level precision assesses whether each extracted roof patch corresponds to a valid roof surface after segmentation and filtering. Second, surface-level geometric alignment evaluates the Euclidean distance between sampled points on the generated IFC roof surfaces and the nearest reference BIM roof surfaces. Third, object-level consistency evaluates centroid-based



(a) UAV-derived photogrammetric reconstruction.

(b) Automated roof detection and segmentation results.

Figure 2. UAV-derived photogrammetric reconstruction and automated roof detection and segmentation results for the EFI building.

association between extracted roof regions and BIM roof entities.

4.2.1 Evaluation of Global Geometric Alignment

To assess the geometric fidelity of the extracted roof surfaces, the spatial alignment between the generated IFC roof geometry and the corresponding roof surfaces in the reference BIM model was evaluated using a surface-based distance analysis. Specifically, geometric deviation was measured as the Euclidean distance between sampled points on the generated IFC roof surfaces and their nearest points on the reference BIM roof surfaces.

Minor geometric discrepancies between the reference BIM and the UAV-derived photogrammetric mesh are expected, as the two datasets were generated from different data sources and represent different abstraction levels (as-designed versus as-built geometry).

Within this context, the global alignment analysis between the generated IFC roof surfaces and the reference BIM roof geometry yielded a root mean square error (RMSE) of 0.84 m and a mean error of 0.68 m, with a minimum of 0.002 m and a maximum of approximately 2.0 m. It should be noted that a systematic vertical offset was observed between GPS-derived UAV elevations and the local coordinate reference system used in the BIM model, indicating a datum inconsistency of approximately 0.7 m.

To robustly identify valid extracted roof surface geometry under these conditions, a **detection threshold** of $\tau_{\text{match}} = 2.0$ m was adopted. A sampled point x_{gen} on a generated IFC roof surface is considered successfully matched if its distance to the nearest reference BIM roof surface S_{gt} satisfies

$$d(x_{\text{gen}}, S_{\text{gt}}) \leq 2.0 \text{ m}.$$

This threshold accommodates the observed systematic vertical offset, potential horizontal geolocation errors, and local geometric deviations at roof edges, while avoiding the exclusion of valid roof surfaces. Under this criterion, approximately 99.0% of the extracted roof surface area was successfully matched to roof elements in the reference BIM model, indicating that the derived roof geometry predominantly corresponds to genuine roof structures rather than noise.

4.2.2 Evaluation of Local Surface Fidelity and Higher-Level Object Consistency

Beyond patch-level precision and surface-based deviation analysis, we further assessed object-level consistency by associating extracted roof regions with reference BIM roof entities. Because the extracted roof regions are typically more fine-grained than the BIM representation, surface-level correspondence does not yield a one-to-one mapping. We therefore perform centroid-based association: each extracted roof region is matched to the nearest BIM roof panel entity within a search radius of $r_{\text{search}} = 5.0$ m. The larger tolerance (compared to the surface-distance threshold) accounts for the observed global offset and centroid displacement caused by region fragmentation.

For each matched pair, an object-level geometric consistency score is computed from area and slope agreement:

$$S_{\text{obj}} = 0.7 \cdot S_{\text{area}} + 0.3 \cdot S_{\text{slope}},$$

where the area similarity term is defined as

$$S_{\text{area}} = 1 - \frac{|A_{\text{gen}} - A_{\text{gt}}|}{\max(A_{\text{gen}}, A_{\text{gt}})},$$

penalising over-fragmentation, and the slope similarity term is defined as

$$S_{\text{slope}} = 1 - \frac{|\theta_{\text{gen}} - \theta_{\text{gt}}|}{45^\circ},$$

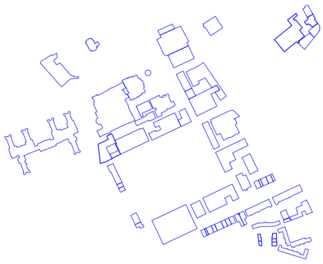
Using this evaluation, 270 out of 292 extracted roof panel entities were successfully associated with a BIM roof panel element (92.5%), with an average S_{obj} of 0.66. Slope agreement was consistently high, supporting the accuracy of PCA-based plane estimation, whereas lower area agreement mainly reflects differences in modelling granularity (a BIM roof panel versus multiple extracted planar regions).

4.2.3 Scalability and Campus-Scale Deployment

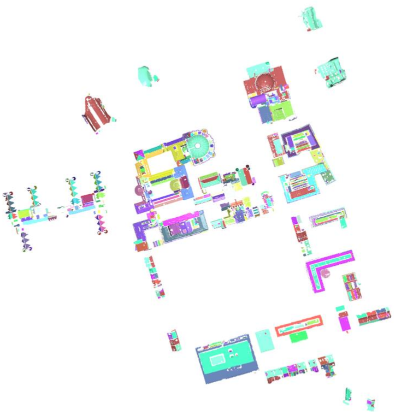
To evaluate scalability, the proposed framework was applied to the complete UAV-derived photogrammetric reconstruction of the University of Edinburgh Central Campus. Geospatial building footprints covering the entire campus were used as semantic priors to guide building-level mesh filtering and automated roof extraction.

Figure 3 shows (a) the spatial distribution of building footprints and (b) the corresponding campus-scale roof extraction results. The framework successfully processed all buildings in a fully automated manner, handling diverse architectural typologies without manual intervention or parameter adjustment.

These results demonstrate that the footprint-guided processing strategy enables robust and efficient scaling from individual buildings to large urban areas, confirming the feasibility of the proposed workflow for campus- and city-scale roof modelling applications.



(a) Geospatial building footprints.



(b) Automated roof extraction results.

Figure 3. Geospatial building footprints and automated roof extraction results for the University of Edinburgh Central Campus.

To provide a concise overview of the quantitative findings reported across Sections 4.1–4.2.3, Table 1 summarises the key performance metrics of the proposed framework.

5 Discussion and Future Work

The proposed method demonstrates robust performance in automated geometric roof reconstruction. Building on this, several directions for future research can be identified to further enhance the method's performance as well as the semantic richness, generality, and applicability of the resulting BIM models.

5.1 Automated Defect Mapping via Texture Projection

The current framework produces geometrically accurate but textureless roof representations, effectively generating 'white-box' models. A natural extension is the retexturing of those dense meshes and then evaluate and record the condition of each roof panel. Using state-of-the-art computer vision methods (e.g., Mask R-CNN or YOLO-based detectors), the panels could be analysed to identify sub-components [11], individual slates [12], and their defects. As a result, defects could be spatially mapped back to their precise 3D locations and tracked over time (over multiple inspections), thereby transforming the static geometric BIM into a dynamic *digital twin* capable of supporting condition monitoring and predictive maintenance workflows.

5.2 Integration with Deep Learning–Based Segmentation

The current framework adopts a geometry-driven segmentation strategy based on footprint-guided filtering, surface-normal consistency, and adaptive height analysis. This design offers several practical advantages for large-scale Scan-to-BIM deployment. First, it avoids the need for large annotated training datasets, which remain scarce for UAV-derived roof meshes with IFC-oriented labels. Second, the method is computationally lightweight and interpretable, making it easier to deploy across new urban areas without retraining. Third, the reliance on explicit geometric constraints is well suited to planar and piecewise-planar roof typologies, which remain common in many existing building stocks.

Nevertheless, the present approach also has limitations. Because it relies on planarity assumptions and empirically selected thresholds, its performance may degrade for curved, free-form, or highly irregular roof structures, as well as in areas affected by noise, incomplete reconstruction, or weak geometric boundaries. Under these conditions, the method may produce over-segmentation of smoothly curved surfaces into multiple planar patches, incomplete recovery of roof boundaries, or simplified geometric approximations that do not fully preserve the original roof form. These limitations are inherent to the current geometry-driven design and should be interpreted within the intended application domain of predominantly planar and piecewise-planar roof reconstruction. In such cases, deep learning-based approaches may provide complementary benefits by learning higher-level geometric and contextual features directly from data.

Future work will investigate hybrid strategies in which deep learning is used to support semantic pre-classification and boundary refinement, while the current geometry-based reconstruction and IFC generation stages are retained for interpretable and BIM-compliant model generation. In particular, multimodal learning on textured photogrammetric meshes could combine 3D geometric information with radiometric appearance cues, improving robustness for challenging roof typologies such as domes, vaults, spires, and curved shells.

6 Conclusion

This paper presented a fully automated Scan-to-BIM framework for large-scale roof modelling from UAV-based photogrammetric data. By exploiting readily available geospatial building footprints as semantic priors, the proposed approach enables efficient isolation of roof geometry from dense urban reconstructions and supports robust processing of complex roof structures.

The framework directly generates georeferenced, IFC4-compliant roof models attributed to a known set of buildings from unstructured photogrammetric meshes, preserving non-convex roof geometries through alpha-shape–based boundary regularisation and avoiding manual intervention. This establishes a practical link between reality capture outputs and interoperable BIM representations suitable for downstream digital twin and GIS applications.

Experimental results on a university campus dataset demonstrate both accuracy and scalability. For a representative complex building, 431 roof panels were correctly extracted with a precision of 98.8%, while campus-scale experiments confirmed the feasibility of applying the workflow to large urban environments.

Overall, the proposed method provides a scalable and fully automated Scan-to-BIM workflow for reconstructing planar and complex non-convex roof geometries from UAV photogrammetric meshes using geospatial building footprints as semantic priors. The contribution of this study lies not in replacing learning-based roof understanding methods, but in demonstrating that a geometry-driven and IFC-oriented pipeline can already provide robust and interpretable performance for large-scale deployment without training data or manual modelling. At the same time, the framework is not intended to address all roof typologies equally well. Curved and free-form structures remain a clear limitation and motivate future hybrid approaches that combine learned semantic understanding with the current geometric reconstruction and BIM generation stages.

Acknowledgement

This work was supported by the Centre for Net Zero High Density Buildings funded by UKRI (UKRI238), as well as a donation by Bentley Systems.

References

[1] Y. Ge, B. Guo, P. Zha, S. Jiang, Z. Jiang, and D. Li. 3d reconstruction of ancient buildings using uav images and neural radiation field with depth supervision. *Remote Sensing*, 16(3), 2024. doi:10.3390/rs16030473.

[2] L. Li, J. Yao, J. Tu, X. Liu, Y. Li, and L. Guo. Roof plane segmentation from airborne lidar data using hierarchical clustering and boundary relabeling. *Remote Sensing*, 12 (9), 2020. doi:10.3390/rs12091363.

[3] X. Chen, Q. An, B. Zhao, W. Tao, T. Lu, H. Zhang, X. Han, and E. Ozdemir. Contour extraction of uav point cloud based on neighborhood geometric features of multi-level growth plane. *Drones*, 8(6), 2024. doi:10.3390/drones8060239.

[4] S. Mola Ebrahimi, H. Arefi, and H. Rasti Veis. Aerial imagery and lidar data fusion for unambiguous extraction of adjacent level-buildings' footprints. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLII-4/W4:179–183, 2017. doi:10.5194/isprs-archives-XLII-4-W4-179-2017.

[5] P. Dabove, M. Daud, and L. Olivotto. Revolutionizing urban mapping: deep learning and data fusion strategies for accurate building footprint segmentation. *Scientific Reports*, 14:13510, June 2024. doi:10.1038/s41598-024-64231-0.

[6] S. Zbirovský and V. Nežerka. Open-source automatic pipeline for efficient conversion of large-scale point clouds to ifc format. *Automation in Construction*, 177:106303, 2025. doi:10.1016/j.autcon.2025.106303.

[7] Y. Schatz and B. Domer. Semi-automated creation of ifc bridge models from point clouds for maintenance applications. *Frontiers in Built Environment*, 10:1375873, June 2024. doi:10.3389/fbuil.2024.1375873.

[8] G. Taubin. Curve and surface smoothing without shrinkage. In *Proceedings of IEEE International Conference on Computer Vision*, pages 852–857, 1995. doi:10.1109/ICCV.1995.466848.

[9] M. Garland and P.S. Heckbert. Simplifying surfaces with color and texture using quadric error metrics. In *Proceedings Visualization '98 (Cat. No.98CB36276)*, pages 263–269, 1998. doi:10.1109/VISUAL.1998.745312.

[10] H. Edelsbrunner and E. P. Mücke. Three-dimensional alpha shapes. *ACM Trans. Graph.*, 13(1):43–72, January 1994. doi:10.1145/174462.156635.

[11] J. Li, B. Tao, F. Bosché, C. X. Lu, and L. Wilson. Extracting roof sub-components from orthophotos using deep-learning -based semantic segmentation. In *Proceedings of the 41st ISARC*, pages 675–682, June 2024. doi:10.22260/ISARC2024/0088.

[12] J. Li, F. Bosché, C. X. Lu, L. Wilson, and B. Tao. Slate detection in orthophotos of building roof panels. In *Proceedings of the 42nd ISARC*, pages 49–56, July 2025. ISBN 978-0-6458322-2-8. doi:10.22260/ISARC2025/0008.