Automatic Reconstruction of High-Accuracy 3D Roof Models from Orthophotos and Digital Surface Models

Yonghe Li[1], Masaya Shimasaki[2], Mitsuteru Sakamoto[2], Toshiaki Satoh[2], Tatsunori Sada[1]

[1]Nihon University, Chiba, Japan {ri.yonfu, sada.tatsunori}@nihon-u.ac.jp
[2]PASCO Corporation, Tokyo, Japan {miaksa4598, moittu9191, tuoost7017}@pasco.co.jp

Keywords: LoD2 building model, 3D roof model, Wireframe extraction, Aerial imagery, Orthophoto, DSM.

Abstract

In recent years, the demand for 3D city model development has grown, as demonstrated by initiatives such as Project PLATEAU in Japan. In the construction of LoD2 building models, which are an essential component of 3D city models, the reconstruction of 3D roof models still heavily depends on manual work. To enhance productivity through automation, this study proposes a novel method for automatically reconstructing high-accuracy 3D roof models using orthophotos and Digital Surface Models (DSMs) derived from aerial imagery. In the proposed method, a deep-learning-based model is first applied to orthophotos and DSMs to extract 2D rooflines. Then, the extracted 2D rooflines are refined and polygonised to assemble 2D roof models. Finally, planar fitting was performed on the point cloud generated from the DSM within each 2D roof plane to reconstruct 3D roof models. In this process, the horizontal alignment of rooflines and the continuity between adjacent roof planes were preserved. In the experiments, 3D roof models manually digitized by stereoscopic measurement were used as the ground truth, and the automatically reconstructed 3D roof models were evaluated by comparison with this reference. As a result, the recall values for 2D and 3D roof planes were 0.686 and 0.430, respectively, and increased to 0.723 and 0.455 for roof planes larger than 4 m².

1. Introduction

The reconstruction of digital twins critically depends on the availability of 3D city models as foundational data. In Japan, the development of such models has accelerated in recent years, driven in particular by Project PLATEAU (MLIT, 2020). In this study, we focus on reconstructing 3D building models, a core component of 3D city models.

CityGML classifies 3D building models into Levels of Detail (LoD) based on representational granularity. LoD1 comprises simple block models, LoD2 incorporates roof geometry, LoD3 adds façade details such as windows and doors, and LoD4 represents interior structures. While LoD1 building models are relatively easy to produce, they lack visual realism and are unsuitable for analyses or simulations that require roof geometry. Consequently, the reconstruction of LoD2 building models is essential for applications such as estimating rooftop photovoltaic energy yield, evaluating the effectiveness of rooftop billboards, analysing urban heat island effects, and modelling radio wave propagation. Despite these needs, the reconstruction of LoD2 building models—particularly 3D roof models— still relies heavily on manual digitization, which is labour-intensive, time-consuming, and costly, underscoring the need for automation.

Numerous studies have addressed the automatic reconstruction of LoD2 building models and 3D roof models. By input modality, prior work falls into three groups: (i) methods using orthophotos and Digital Surface Models (DSMs), with data derived from aerial imagery (Xu et al., 2024; Schuegraf et al., 2024; Gao et al., 2023) and from satellite imagery (Gui et al., 2024; Gui and Qin, 2021; Partovi et al., 2019); (ii) methods based on airborne LiDAR point cloud (Pađen et al., 2024; Peters et al., 2022; Awrangjeb et al., 2018); and (iii) methods utilizing single aerial or satellite images (Lussange et al., 2023; Alidoost et al., 2019).

The characteristics of three classes of automatic reconstruction methods can be summarized as follows. Airborne LiDAR-based methods offer very high positional accuracy, but they often suffer from lower spatial sampling than high-resolution aerial imagery, higher acquisition costs, and limited or low-resolution texture information. Single-image methods must infer height from imagery alone, which makes height estimation challenging and can reduce generalization across scenes and sensors. Methods using orthophotos and DSMs rely on products derived from aerial or satellite imagery. Satellite imagery affords low-cost, wide-area coverage but is limited by coarser ground sampling distance (GSD) and lower geolocation accuracy, whereas aerial imagery provides finer spatial resolution, higher positional accuracy, and richer texture at relatively low acquisition cost. Accordingly, in the automatic reconstruction of 3D roof models that require a specified level of surveying accuracy, aerial imagery is currently the most widely used data source.

Motivated by these considerations, we present a method for automatic reconstruction of high-accuracy 3D roof models from aerial-imagery-derived orthophotos and DSMs. The remainder of this paper is organized as follows: Section 2 reviews related work on reconstruction from orthophotos and DSMs; Section 3 describes the proposed method; Section 4 reports the experimental setup and results; Section 5 discusses the findings; and Section 6 concludes and outlines future work.

2. Related Works

2.1 3D Roof Model Reconstruction from Orthophotos and DSMs

In studies on the automatic reconstruction of 3D roof models from aerial-imagery–derived orthophotos and DSMs (hereafter, “orthophotos and DSMs” refers specifically to aerial imagery–derived products), a prevailing bottom-up pipeline proceeds in three stages: (i) extract 2D rooflines from the orthophoto; (ii) regularize and assemble the extracted rooflines into 2D roof planes; and (iii) fuse DSM-derived heights to reconstruct 3D roof models. Representative studies are reviewed below.

This contribution has been peer-reviewed.
https://doi.org/10.5194/isprs-archives-XLIX-B2-2026-459-2026

McClune et al. (2016) extracted roofline regions from the orthophoto using a Canny edge detector, suppressed false positives within those regions using nDSM data and a scan-line segmentation method, and vectorized the remaining regions to obtain 2D rooflines. Intersections of the 2D rooflines were used as roof vertices, and DSM-derived heights were assigned to reconstruct 3D roof models.

Schuegraf et al. (2024) used an instance-segmentation model to segment roof planes and two roofline classes from orthophotos and DSMs. Boundaries between building sections and adjacent roof planes were extracted via watershed and a tree search, the line sequences were vectorized and simplified with the Douglas–Peucker algorithm to obtain 2D rooflines, and these were polygonised into 2D roof planes. Subsequently, RANSAC-based plane fitting was applied to the DSM-derived point cloud (hereafter, the DSM point cloud) to reconstruct 3D roof planes. Near-horizontal planes were snapped to horizontal, and height consistency was enforced for planes sharing a roofline.

Xu et al. (2024) first applied a semantic segmentation model to orthophotos and DSMs to segment regions corresponding to roof planes, rooflines, and vertices. Vectorization using connectivity learning was then performed to extract 2D rooflines, from which 2D roof planes were assembled. For 3D reconstruction, RANSAC-based plane fitting was applied to the DSM point cloud within each roof plane.

Gao et al. (2023) extracted 2D rooflines from orthophotos using the Keypoint-based Incremental Polygonal-line Parsing for Inference (KIPPI) algorithm, and, under geometric constraints, removed redundant or spurious rooflines and regularized the remaining rooflines. Building footprint polygons used as prior information were partitioned by the 2D rooflines to form 2D roof planes, and heights from the DSM point cloud were assigned to reconstruct 3D roof models.

Komori et al. (2023) first classified roofs into flat and non-flat types using orthophotos and processed them accordingly. Assuming building footprints as prior information, for non-flat roofs a deep-learning–based wireframe detection model was applied to orthophotos to extract 2D rooflines, and the regularized 2D rooflines together with the footprints were then used to assemble 2D roof planes. Finally, for 3D reconstruction, each roof plane was triangulated, and the vertex heights were estimated via linear programming with a bi-objective formulation that maximizes roof volume and minimizes the sum of angles between the normal vectors of adjacent triangles. DSM point cloud was not used directly as height inputs. Instead, lower and upper roof heights were derived and used as constraints.

2.2 Methods per Processing Stage: Advantages and Limitations

This subsection reviews alternatives for each stage of 3D roof model reconstruction from orthophotos and DSMs and summarizes advantages and limitations.

(i) 2D Roofline Extraction. Segmentation-then-vectorization applies semantic segmentation to obtain regions for roof planes or rooflines and then converts raster outputs to vector linework. Its limitations include the need for rule-based post-processing and the difficulty of end-to-end optimization across the two steps. Direct vector extraction encompasses rule-based and deep-learning-based methods. In general, deep-learning-based wireframe-detection models achieve higher accuracy, particularly for complex roof geometries and varied imaging conditions.

(ii) 2D Roof Plane Reconstruction. After 2D roofline extraction, geometric regularization removes spurious and redundant lines, merges fragmented segments, enforces orthogonality and parallelism, snaps nearly coincident segments, and repairs topology. Polygonization then produces closed 2D roof-plane polygons bounded by the regularized rooflines. Notably, the regularization is predominantly rule based and must be tailored to the characteristics of the chosen roofline extractor.

(iii) 3D Roof Plane Reconstruction. 3D roof planes are reconstructed from 2D roof planes using either the DSM or the DSM point cloud to provide elevations. Height assignment is treated at three granularities—vertices, rooflines, and roof planes. Because DSM accuracy degrades near building boundaries, direct assignment at vertices or along rooflines does not ensure sufficient accuracy. Accordingly, high-accuracy reconstruction favours plane-level height estimation, applying robust plane fitting to the DSM point cloud within each 2D roof plane to reduce noise sensitivity.

3. Method

3.1 Overview of the Proposed Method

Figure 1 outlines the proposed pipeline for automatic reconstruction of 3D roof models from orthophotos and DSMs. We first apply deep-learning–based semantic segmentation or object detection to the orthophotos to delineate per-building regions (masks). Each region is then expanded with a fixed-width buffer, and the corresponding patches are cropped from the orthophotos and DSMs. We next extract 2D rooflines from these inputs, assemble them into a 2D roof model, and finally assign elevations from the DSM to reconstruct the 3D roof model.

It should be noted that automatic methods for building region extraction are relatively mature, and numerous studies have been reported (e.g., Luo et al., 2021). Moreover, since a buffer is applied in the process of extracting individual buildings, the accuracy requirements for the extent of the extracted building regions are not particularly stringent. Accordingly, this study assumes that building regions are given and focuses on the downstream steps.

Figure 2 summarizes the processing pipeline for automatically reconstructing 3D roof models from per-building orthophoto and DSM patches. The subsequent sections detail each stage: (i) extraction of 2D rooflines, (ii) reconstruction of a 2D roof model, and (iii) reconstruction of a 3D roof model by assigning elevations from the DSM point cloud.

3.2 2D Roofline Extraction

In our prior work (Li et al., 2024), we identified HEAT (Chen et al., 2022) and HAWPv2 (Xue et al., 2022) as strong candidates for automatic 2D roofline extraction and evaluated their accuracy. HEAT is a Vision Transformer-based method, whereas HAWPv2 is CNN-based, and the two methods exhibit distinct characteristics. The evaluation results showed that HEAT achieved an F1-score 14% higher than HAWPv2 and that incorporating DSM data into orthophoto pipeline produced an additional 13% gain in accuracy.

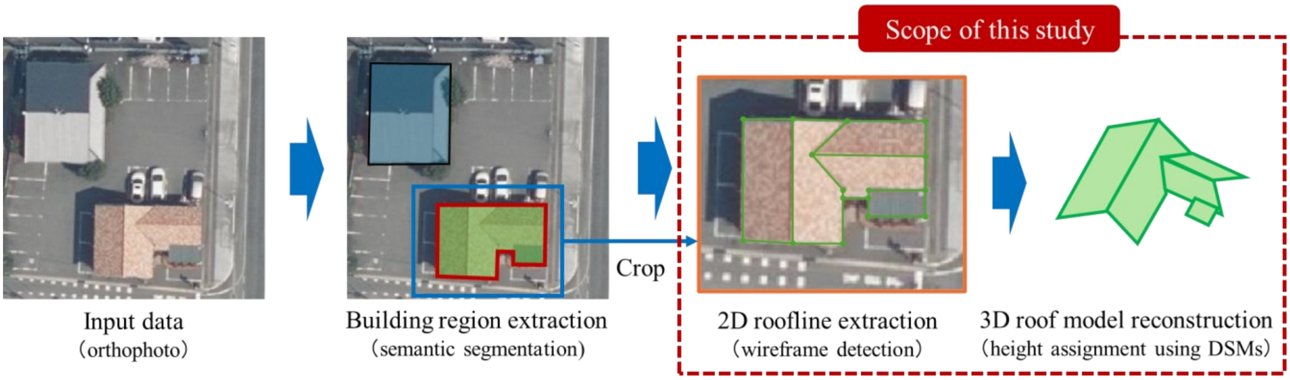


Figure 1. Overall pipeline for automatic 3D roof model reconstruction from orthophotos and DSMs.

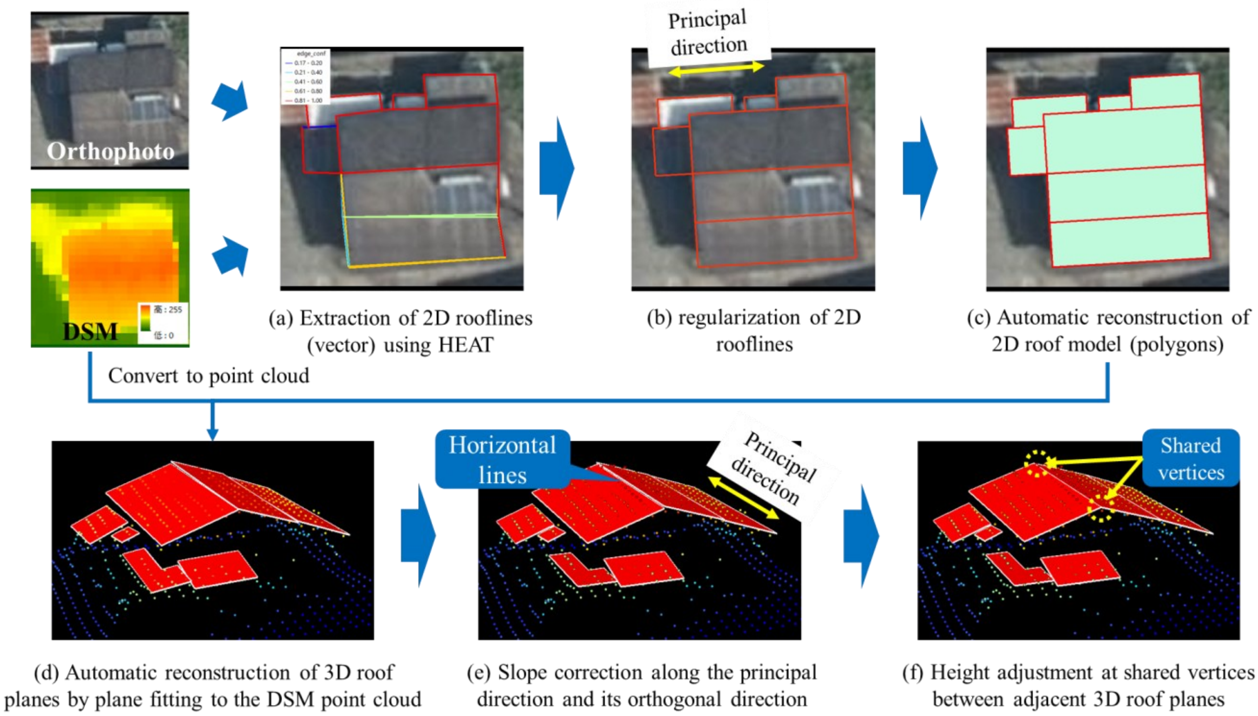


Figure 2. Per-building workflow for reconstructing a 3D roof model from an orthophoto and a DSM.

Based on these findings, this study uses HEAT with orthophotos and DSMs as inputs to extract 2D rooflines (Figure 2(a)). For each detected line segment, HEAT outputs the 2D coordinates of its two endpoints and a confidence score between 0.0 and 1.0.

3.3 2D Roof Model Reconstruction

In the reconstruction of 2D roof models, the 2D rooflines extracted by HEAT are first subjected to geometric regularization (Figure 2(b)), after which the smallest closed regions bounded by the regularized rooflines are extracted as 2D roof planes (polygons; Figure 2(c)). The specific processing procedure is described below.

3.3.1 Geometric Refinement of Rooflines: Roofline segments with confidence scores ≤ 0.1 are removed, redundant or coincident segments are discarded, and polylines are simplified by straightening near-collinear vertices. Vertices close to rooflines are snapped to their nearest roofline and endpoints are updated accordingly, ensuring geometric and topological consistency.

3.3.2 Principal Direction Estimation and Roofline Rectification: The dominant roofline direction is estimated by clustering roofline segments by orientation and computing a length-weighted mean direction over the largest clusters. Given this principal direction and its orthogonal counterpart, connected roofline segments are then rectified by projecting their vertices onto lines aligned with these two directions, as illustrated in Figure 2(b) and Figure 3.

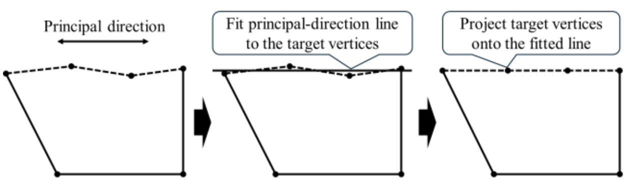


Figure 3. Orientation correction for 2D rooflines.

3.3.3 Reconstruction of 2D Roof Planes: After these regularization steps, the smallest closed regions bounded by the rooflines are extracted as 2D roof planes.

3.4 3D Roof Model Reconstruction

In the reconstruction of 3D roof models, we preserve the planimetric geometry of the 2D roof models and assign elevations using the DSM point cloud. This design reflects the premise that planimetry derived from orthophotos generally attains higher horizontal accuracy than that derived from DSM point cloud. Specifically, we first fit a plane to the DSM point cloud within each 2D roof plane to estimate the corresponding 3D roof plane (Figure 2(d)). Next, along the principal direction and its orthogonal direction, if the estimated slope magnitude is below a threshold (e.g., 10°), we set it to 0° to enforce horizontal alignment of rooflines in those directions (Figure 2(e) and Figure 4). Finally, we adjust elevations at shared vertices between adjacent 3D roof planes to ensure connectivity of the 3D roof model (Figure 2(f) and Figure 5).

3.4.1 3D Roof-Plane Fitting and Elevation Assignment: For each 2D roof plane, we extract the DSM point cloud within its boundary and apply RANSAC-based least-squares plane fitting (Figure 2(d)). For RANSAC-based plane fitting, the residual threshold is set to 20 cm and the minimum sample size is 3 when the number of points is fewer than 20 and 5 otherwise. The horizontal coordinates of each 3D roof-plane vertex are inherited from the corresponding 2D roof plane, and the elevation is computed using the equation of the fitted plane. Because DSM heights tend to degrade near building boundaries, when sufficient points are available, we fit using only points located at least a prescribed distance (e.g., 50 cm) from the polygon boundary.

3.4.2 Slope Correction of 3D Roof Planes: For each 3D roof plane, slope angles are computed along the principal and orthogonal directions, and any angle below a threshold (e.g., 10°) is set to 0° (Figure 2(e)). As illustrated in Figure 4, the DSM point cloud belonging to each roof plane is projected onto a vertical plane whose normal corresponds to either the principal or orthogonal direction. A line is then fitted to the projected points using RANSAC-based least squares. The corrected 3D roof plane is the plane determined by the fitted line and the corresponding normal vector. When both directions are set to 0°, a horizontal plane is directly estimated using RANSAC-based least squares.

3.4.3 Height Adjustment at Shared Vertices between Adjacent 3D Roof Planes: As illustrated in Figure 2(f) and Figure 5, we enforce height consistency at vertices shared between adjacent 3D roof planes. First, for each vertex shared by a pair of adjacent 2D roof planes, we compute its height on each of the corresponding 3D roof planes. If the inter-plane height difference is within a prescribed tolerance, we designate the vertex as a 3D shared vertex (hereafter, 3D shared vertex). Next, any two 3D roof planes that possess a 3D shared vertex are judged to be in an "adjacent" relation, and 3D roof planes that are mutually adjacent are grouped. For all 3D roof planes within a group, we estimate the optimal planes jointly by performing simultaneous multi-plane fitting to the DSM point cloud while enforcing height equality at the 3D shared vertices. The point-cloud samples used in this step are the same as those employed during the individual plane-fitting stage. The estimation minimizes the objective function defined in Equation (1), and the optimum is obtained via nonlinear least squares. The total objective function E is defined as the sum of the plane fitting term E_1 and the shared-vertex elevation consistency term E_2. Figure 6 shows examples before and after the height adjustment at 3D shared vertices. Enforcing height consistency at these vertices effectively eliminates the gaps that were previously visible between adjacent planes.

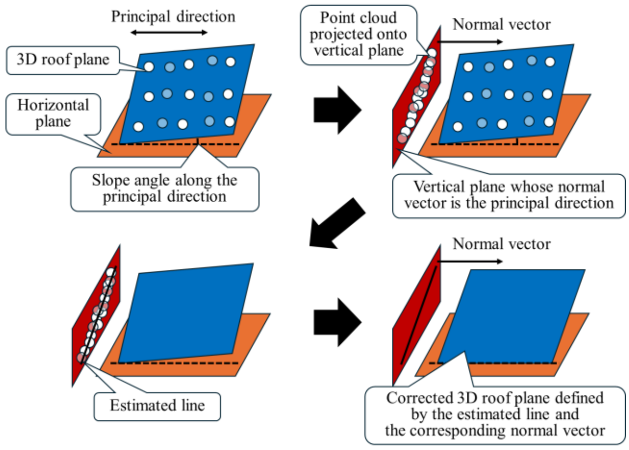
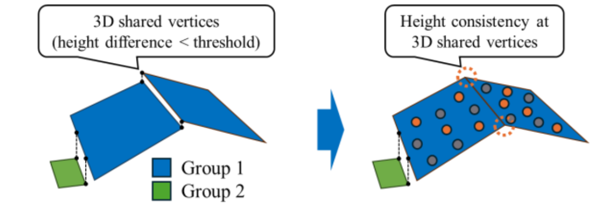


Figure 4. Slope correction for 3D roof planes.



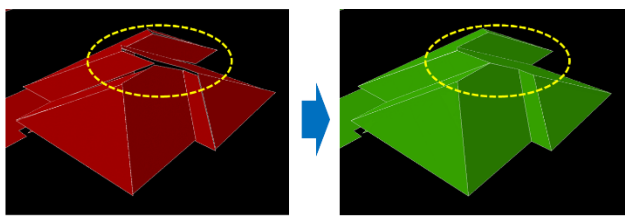
(a) Identification of 3D shared vertices and grouping of 3D roof planes

(b) Height adjustment at 3D shared vertices in Group 1

Figure 5. Height adjustment at 3D shared vertices.

$$\begin{cases} E = E_1 + E_2 \\ E_1 = \sum_{k=1}^{P} \sum_{i=1}^{N_k} \frac{(a_k \cdot xp_{ki} + b_k \cdot yp_{ki} - zp_{ki} + d_k)^2}{a_k^2 + b_k^2 + 1} \\ E_2 = \sum_{k1=1}^{P} \sum_{k2=1}^{P} \sum_{j=1}^{M} \lambda_{k1,k2,j} \left(|a_{k1} \cdot xc_j + b_{k1} \cdot yc_j + d_{k1}| - |a_{k2} \cdot xc_j + b_{k2} \cdot yc_j + d_{k2}| \right)^2 \end{cases}, \quad (1)$$

where a_k, b_k, d_k = plane parameters of the k-th 3D roof plane (hereafter, plane k)
$xp_{ki}, yp_{ki}, zp_{ki}$ = 3D coordinates of the i-th DSM point belonging to plane k
xc_j, yc_j = 2D coordinates of the j-th 2D shared vertices
$\lambda_{k1,k2,j}$ = weighting coefficient (large value, e.g. 10^4, if vertex j is 3D shared vertex by planes k_1 and k_2; otherwise, 0)
P = number of 3D roof planes within a group
N_k = number of DSM points assigned to plane k
M = number of 2D shared vertices



Before height adjustment

After height adjustment

Figure 6. Comparison of results before and after height adjustment at 3D shared vertices.

4. Experiments

To verify the effectiveness of the proposed method, we automatically reconstructed 3D roof models using orthophotos and DSMs derived from aerial imagery, and evaluated accuracy by comparing them with manually constructed 3D roof models.

4.1 Experimental Data

As shown in Figure 7, the study area is a residential district in a regional city in Gifu Prefecture, Japan, characterized by a high concentration of detached houses. The area of interest, delineated by the red boundary in Figure 7, is approximately 0.25 km². A Leica Geosystems CityMapper-2 sensor was mounted on a fixed-wing aircraft to simultaneously acquired oblique aerial imagery and LiDAR point-cloud data. The forward overlap and sidelap were 80% and 51%, respectively. The ground sampling distance (GSD) of the aerial imagery was 10 cm, and the point-cloud density was 6.4 points/m². The orthophoto and DSM were automatically generated using PhotoMesh (Skyline Software Systems) with a GSD of 10 cm. The DSM was derived from oblique aerial imagery and LiDAR point cloud, and was subsequently cleaned manually to remove noise and outliers. In generation of the DSM, LiDAR point-cloud data were used in addition to aerial imagery as supplementary information. Although this may modestly improve the quality of the DSM, it should be noted that this has no essential impact on the proposed method for reconstructing 3D roof models using orthophotos and DSMs in this study.

The study area comprises 759 detached houses, which we manually digitized in 3D to build the ground-truth dataset. The 3D coordinates of vertices on roof planes were compiled by stereoscopic compilation from aerial stereo imagery, providing higher 3D positional accuracy than automatic extraction from orthophotos and DSMs. Accordingly, these data were adopted as ground truth, resulting in a dataset comprising 3,496 roof planes and 18,415 rooflines.

4.2 Criteria for Quantitative Evaluation

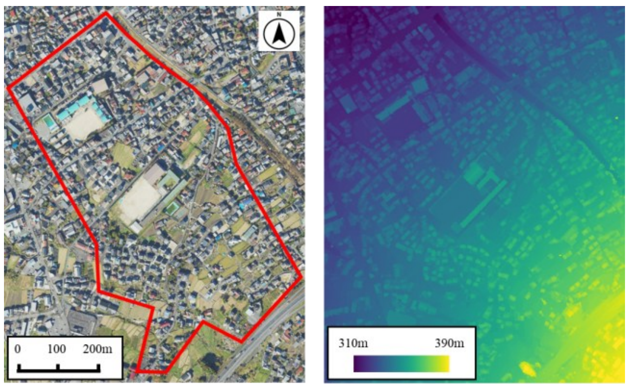
3D roof models are required to satisfy Map Information Level 2500, a Japan-specific geospatial accuracy standard that stipulates a horizontal positional standard deviation ≤ 1.75 m and a vertical standard deviation ≤ 0.66 m. To accommodate potential errors in the manually compiled ground-truth data, stricter evaluation criteria were adopted. For each vertex on rooflines and roof planes, reconstruction was counted as successful only when both the maximum horizontal deviation from the reference was ≤ 1.75 m and the maximum vertical deviation was ≤ 0.66 m (thresholds defined as maximum deviations, not standard deviations). Quantitative results under these criteria are summarized in Table 1 and discussed below.

4.3 Evaluation of 2D Roofline Extraction

For training and testing in 2D roofline extraction, five-fold cross-validation was employed. Eighty percent of the data were used for training and the remaining 20% for testing, and five experiments were conducted with different test sets in each fold.

In preparing the training and test datasets, a bounding box was computed from the ground-truth building polygons for each building, expanded by a 10% buffer, and used to crop the corresponding regions from both the orthophoto and DSM. Each cropped window was rescaled to 256 × 256 pixels while preserving aspect ratio, with out-of-bounds regions zero-padded (black) to form a square image patch. For the DSM, heights within each window were linearly normalized to the 0–255 range. As ground truth, 2D rooflines were derived from the manually compiled 3D roof-plane polygons and mapped to the image coordinates of the corresponding patch.

Training was performed for 400 epochs until the loss converged. The model was initialized from publicly available pretrained weights and trained using the default hyperparameters. For evaluation, the nearest automatically reconstructed 2D roofline was associated with each manually delineated 2D roofline, and a detection was counted as correct only if the distances between the corresponding endpoints at both ends were ≤ 1.75 m. Averaging over five inference runs, the test-set metrics were: recall = 0.794, precision = 0.751, and F1 score = 0.772. False negatives were primarily attributable to cast shadows and roof-plane boundaries that were indistinct in the orthophotos, with additional failures arising from occlusions by trees and other objects. By contrast, false positives were mainly caused by redundant detections of 2D rooflines, along with misidentifications of rooftop equipment such as solar panels and parked vehicles near buildings.

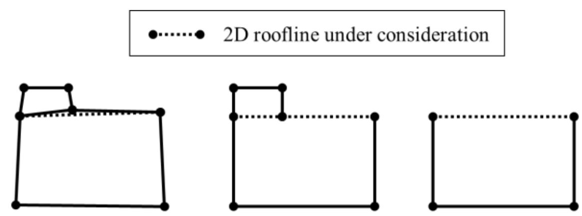


(a) Orthophoto with study-area boundary
(b) DSM (Elevation, m)

Figure 7. Study area overview—orthophoto and DSM.

	Recall	Precision	F1-score
HEAT-detected 2D roofline	0.794	0.751	0.772
Regularized 2D roofline	0.760	0.826	0.791
2D roof plane	0.686	0.657	0.671
3D roof plane	0.430	0.411	0.421

Table 1. Accuracy evaluation of 2D rooflines, 2D roof planes, and 3D roof planes



(a) HEAT-detected 2D rooflines
(b) Regularized 2D rooflines
(c) Ground-truth 2D rooflines

Figure 8. Example of recall degradation in 2D roofline extraction after geometric regularization.

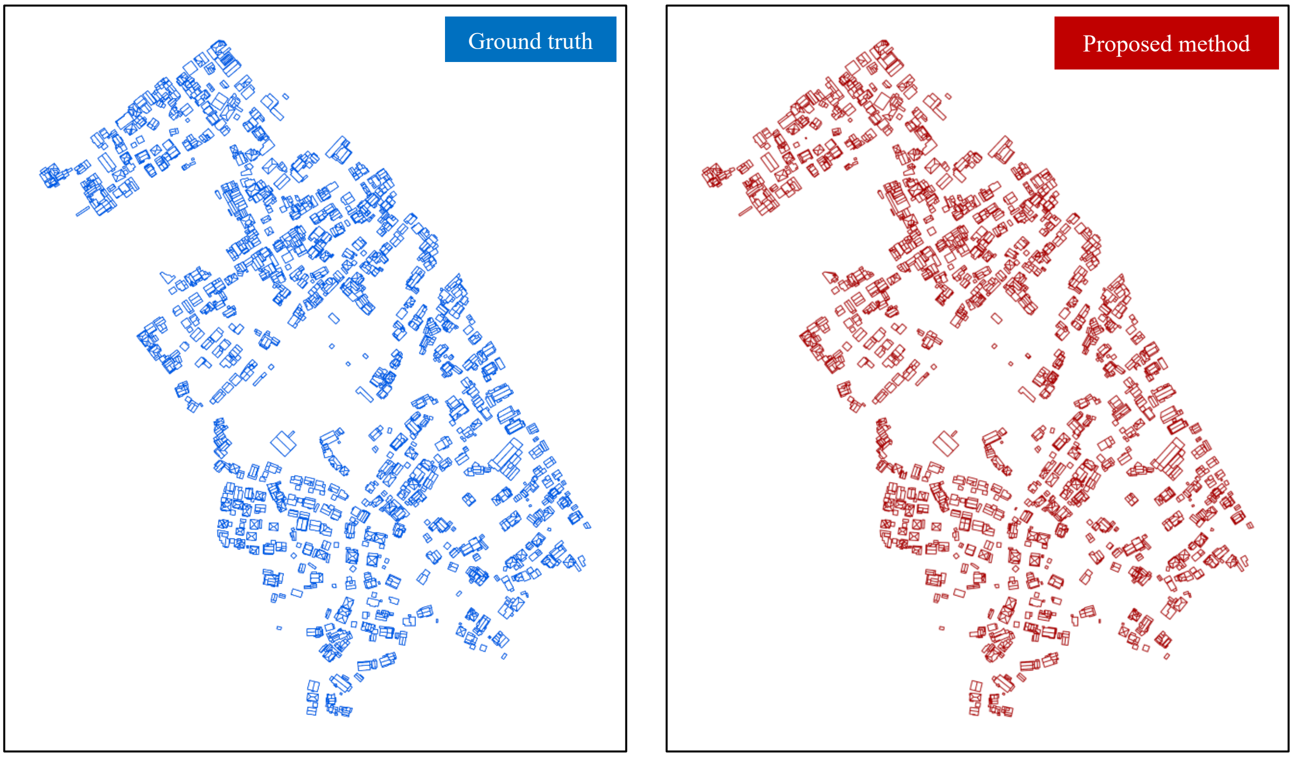


Figure 9. Qualitative evaluation results for 2D roof planes.

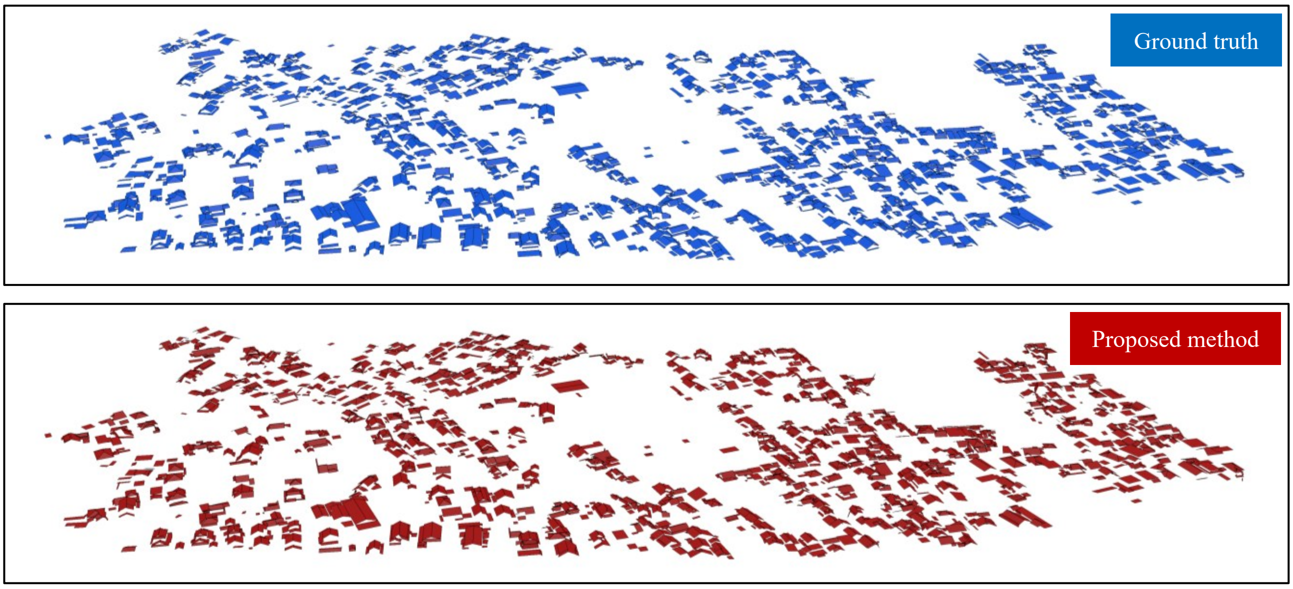


Figure 10. Qualitative evaluation results for 3D roof planes.

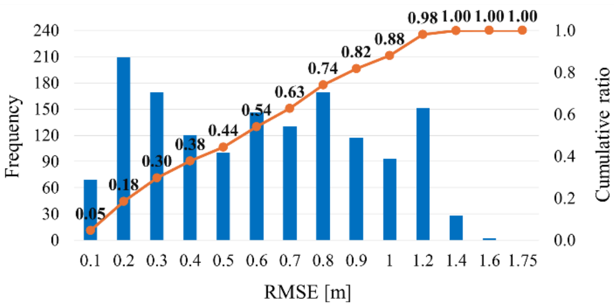


Figure 11. Distribution of horizontal accuracy (RMSE) of vertices on successfully reconstructed 3D roof planes.

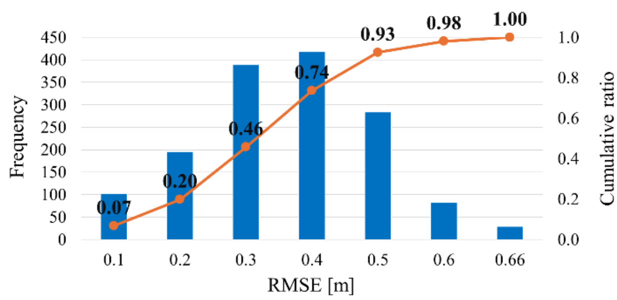


Figure 12. Distribution of vertical accuracy (RMSE) of vertices on successfully reconstructed 3D roof planes.

4.4 Evaluation of 2D Roof Plane Reconstruction

Based on the procedure described in Section 3.3, the 2D rooflines extracted by HEAT were refined, and 2D roof planes were automatically generated. As shown in Table 1, the refined 2D rooflines achieved a recall of 0.760, a precision of 0.826, and an F1-score of 0.791. Compared with the original HEAT extraction, precision and F1-score increased by 0.075 and 0.019, respectively, while recall decreased by 0.034. These results indicate that the overall accuracy of the 2D roofline extraction improved after refinement, particularly due to a substantial reduction in over-detections. The decrease in recall suggests that some 2D rooflines accurately extracted by HEAT were displaced during refinement.

However, even when the refinement procedure itself was appropriate, cases occurred in which over-detected rooflines from HEAT caused a decline in recall. For instance, in the example shown in Figure 8, the automatically extracted segment indicated by the dashed line in Figure 8(a) was originally judged as correctly extracted relative to the manually digitized roofline (dashed line in Figure 8(c)). Because excessive rooflines were present in the upper left of Figure 8(a), refinement caused the originally single roofline (dashed line) to be split into two segments, as shown in Figure 8(b), leading to a false negative. Nevertheless, most refinements were applied appropriately, and the impact of erroneous refinements on recall was confirmed to be limited.

The accuracy of 2D roof-plane reconstruction was measured at recall, precision, and F1-score of 0.686, 0.657, and 0.671, respectively. The recall for 2D roof planes was approximately 0.074 lower than that for the refined 2D rooflines. This decline is presumed to result from the larger number of vertices forming each roof plane, which increases the likelihood that some vertices fail to meet the positional-accuracy threshold, thereby reducing recall. Figure 9 shows a comparison between the automatically reconstructed 2D roof planes and the manually created ground-truth data. Although a high level of overall agreement was achieved, local over-extraction and omission of 2D roof planes were observed.

4.5 Evaluation of 3D Roof Plane Reconstruction

Based on the 2D roof planes and DSM point cloud, 3D roof planes were reconstructed using the method described in Section 3.4. The automatically reconstructed 3D roof planes achieved recall, precision, and F1-score values of 0.430, 0.411, and 0.421, respectively (Table 1). The accuracy of the 3D roof planes was evaluated using the same criteria as for the 2D roof planes, but the assessment additionally incorporated vertical discrepancies within 0.66 m as well as horizontal ones. Although the recall for 3D roof planes decreased substantially—by 0.256 compared with that of the 2D roof planes—this reduction can be attributed to insufficient height accuracy at some vertices, since their horizontal positions remained fixed. Furthermore, Figures 11 and 12 show the frequency distributions and cumulative ratios of the horizontal and vertical positional accuracies (RMSE) for the successfully reconstructed 3D roof planes. The results show that, among the reconstructed 3D roof planes, 88% had a horizontal RMSE below 1.0 m and 93% had a vertical RMSE below 0.5 m, demonstrating high positional accuracy in both directions for approximately 90% of the roof planes.

Figure 10 shows a comparison between the automatically reconstructed 3D roof planes and the manually created ground-truth data. The shapes of the 3D roof planes were generally well reproduced, and the reconstruction success rate was particularly high for larger roof planes. However, in some cases, roof planes that should be separated vertically were reconstructed as continuous surfaces, while the opposite situation also occurred.

5. Discussion

To identify factors affecting the accuracy of 2D and 3D roof planes, Figure 13 presents recall values classified by roof-plane area. As shown in Figure 13, the trends for 2D and 3D roof planes are similar, with both exhibiting a marked decline in recall when the area is smaller than 4 m². For roof planes with an area of 4 m² or larger, the recall values for 2D and 3D roof planes are 0.723 and 0.455, respectively. Furthermore, the differences in recall between 2D and 3D roof planes are 0.186 and 0.145 for areas larger and smaller than 4 m², respectively, indicating that the recall varies considerably depending on roof-plane size. The lower recall observed for small roof planes is presumed to result from frequent roofline extraction failures by HEAT in the 2D case and from insufficient DSM point density that hinders stable plane fitting in the 3D case. Therefore, future work should focus on improving extraction accuracy for small roof planes.

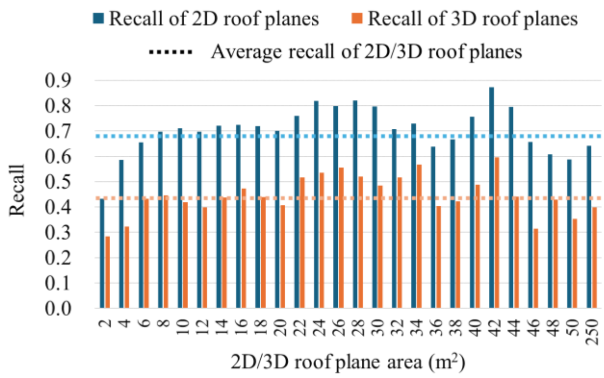


Figure 13. Recall by area for automatically reconstructed 2D and 3D roof planes.

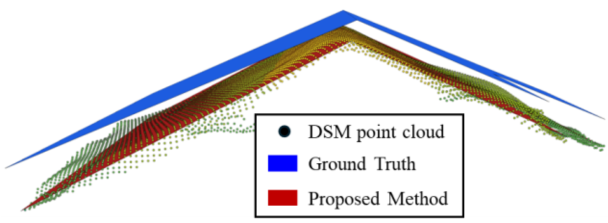


Figure 14. Comparison of the automatically reconstructed 3D roof model, the ground-truth model, and the DSM point cloud.

Figure 14 presents a comparison of the automatically reconstructed 3D roof model, the ground-truth model, and the DSM point cloud. As shown in the figure, the automatically reconstructed 3D roof model exhibits high consistency with the DSM point cloud. However, because the point cloud has a negative height bias of approximately 20–50 cm, a substantial discrepancy arises when compared with the ground-truth model. Therefore, to improve the accuracy of 3D roof-plane reconstruction, it is necessary not only to refine the proposed method but also to enhance the accuracy of the DSM point cloud or develop height-estimation approaches that do not depend on the DSM.

6. Conclusion

In this study, we proposed a new method for automatically reconstructing high-accuracy 3D roof models using orthophotos and DSMs generated from aerial imagery. To evaluate the effectiveness of the proposed method, experiments were conducted using real data, and the recall for 3D roof planes was confirmed to be 0.430. These results demonstrate that by incorporating manual corrections based on the automatically reconstructed 3D roof models, the proposed workflow can contribute to improving the efficiency of LoD2 building model production.

The main contributions of this study are summarized as follows.
(1) Based on a review of previous studies, optimal methods were examined for achieving high-accuracy results at each processing stage, and a new processing workflow was proposed (Figures 2).
(2) At each processing stage, original improvements enhanced performance over conventional methods: (i) construction of a high-accuracy 2D roofline extraction method using HEAT with orthophotos and DSMs; (ii) development of a roofline refinement procedure tailored to HEAT's extraction characteristics; and (iii) design of a plane-fitting–based 3D roof-plane reconstruction method that guarantees roofline horizontality and continuity across adjacent planes.
(3) The proposed method does not require prior information such as building footprints or roof templates, enabling flexible application to a wide variety of roof shapes.

To achieve further improvements in accuracy and efficiency, several important challenges must be addressed: (i) advancing methods for 2D roofline extraction and regularization—particularly for small roof surfaces; (ii) leveraging higher-accuracy point-cloud data; and (iii) developing height-estimation techniques based on aerial photogrammetry that do not rely on the DSM.

References

Alidoost, F., Arefi, H., Tombari, F., 2019: 2D image-to-3D model: Knowledge-based 3D building reconstruction (3DBR) using single aerial images and convolutional neural networks (CNNs). *Remote Sensing*, 11(19), 2219.

Awrangjeb, M., Gilani, S.A.N., Siddiqui, F.U., 2018. An effective data-driven method for 3-D building roof reconstruction and robust change detection. *Remote Sensing*, 10(10), 1512.

Chen, J., Qian, Y., Furukawa, Y., 2022. HEAT: Holistic Edge Attention Transformer for Structured Reconstruction. *Computer Vision and Pattern Recognition*, DOI: 10.1109/CVPR52688.2022.00384.

Gao, W., Peters, R., Stoter, J., 2023. Unsupervised roofline extraction from true orthophotos for LoD2 building model reconstruction. *International 3D GeoInfo Conference*, 425–436.

Gui, S., Qin, R., 2021. Automated LoD-2 model reconstruction from very-high-resolution satellite-derived digital surface model and orthophoto. *ISPRS Journal of Photogrammetry and Remote Sensing*, 181, 1-19.

Gui, S., Schuegraf, P., Bittner, K., Qin, R., 2024. Unit-level LoD2 Building Reconstruction from Satellite-derived Digital Surface Model and Orthophoto, ISPRS Annals of the Photogrammetry, *Remote Sensing and Spatial Information Sciences*, X-2-2024, 81-88.

Komori, J., Yu, C., Hattori, S., Nonaka, H., 2023. Development of an automatic reconstruction tool for LoD2 building models. *Proceedings of the Japan Society of Photogrammetry and Remote Sensing, Autumn Conference 2023*, 61–64. [in Japanese]

Li, Y., Sakamoto, M., Satoh, T., 2024. Accuracy evaluation of a wireframe detection method for automatic reconstruction of LOD2 building models. *Proceedings of the Japan Society of Photogrammetry and Remote Sensing, Autumn Conference* 2024, 101–102. [in Japanese]

Luo, L., Li, P., Yan, X., 2021. Deep Learning-Based Building Extraction from Remote Sensing Images: A Comprehensive Review, *Energies*, 14(23), 7982.

Lussange, J., Yu, M., Tarabalka, Y., Lafarge, F., 2023. 3D detection of roof sections from a single satellite image and application to LoD2-building reconstruction. *arXiv:2306.17139*.

McClune, A.P., Mills, J.P., Miller, P.E., Holland, D.A., 2016. Automatic 3D building reconstruction from a dense image matching dataset. The International Archives of the Photogrammetry, *Remote Sensing and Spatial Information Sciences*, XLI-B3, 12-19.

Ministry of Land, Infrastructure, Transport and Tourism (MLIT), 2020. Project PLATEAU website. https://www.mlit.go.jp/plateau (accessed 15 Oct 2025).

Pađen, I., Peters, R., García-Sánchez, C., Ledoux, H., 2024.Automatic high-detailed building reconstruction workflow for urban microscale simulations. *Building and Environment*, 265, 111978.

Partovi, T., Fraundorfer, F., Bahmanyar, R., Huang, H., Reinartz, P., 2019. Automatic 3-D building model reconstruction from very high resolution stereo satellite imagery, *Remote Sensing*, 11(14), 1660.

Peters, R., Dukai, B., Vitalis, S., van Liempt, J., Stoter, J., 2022. Automated 3D reconstruction of LoD2 and LoD1 models for all 10 million buildings of the Netherlands. ISPRS Annals of the Photogrammetry, *Remote Sensing and Spatial Information Sciences*, V-4-2022, 171–178.

Schuegraf, L., Beil, C., Dall'Osso, S., 2024. PLANES4LOD2: Reconstruction of LoD-2 building models using a depth attention-based fully convolutional neural network. *ISPRS Journal of Photogrammetry and Remote Sensing*, 211, 425-437.

Xu, Y., Jubanski, J., Bittner, K., Siegert, F., 2024. Roof plane parsing towards LoD-2.2 building reconstruction based on joint learning using remote sensing images. *International Journal of Applied Earth Observation and Geoinformation*, 133, 104096.

Xue, N., Wu, T., Bai, S., Wang, F., Xia, G., Zhang, L., Torr, P., 2022. Holistically-Attracted Wireframe Parsing: From Supervised to Self-Supervised Learning. *arXiv:2210.12971*.