Conjugate Feature-Guided Dense Stereo Matching for High-Precision Attribute-Enriched Urban Point Clouds

Yung-Ching Yang [1], Jen-Jer Jaw[2]

Dept. of Civil Engineering, National Taiwan University, Taipei, Taiwan – [1] ivy.ycyang@gmail.com, [2] jejaw@ntu.edu.tw

Keywords: Attribute-Enriched Point Clouds, Conjugate Feature-Guided Matching, Semantic-Geometric Constraints, Disparity Initialization, Information-Driven Modeling.

Abstract

This research develops an attribute-enriched dense image matching framework to generate high-precision 3D point cloud information characterized by enhanced geometric fidelity and semantic interpretability. Traditional dense reconstruction often suffers from radiometric ambiguities and structural distortions in complex urban scenes, such as those with repetitive industrial patterns and specular reflections. To overcome these limitations, the proposed pipeline integrates high-level semantic attributes with multi-scale geometric features, including keypoints, edges, and structural vertices.

These attributes serve as reliable conjugate feature seeds to initialize a sparse disparity map, which is subsequently densified via Weighted Inverse Distance Weighting and refined through adaptive Normalized Cross-Correlation. During the dense matching phase, a dual-layer constraint mechanism, comprising attribute-dependent searching ranges and class-consistency validation, is enforced to mitigate cross-object mismatches.

Experimental results demonstrate that the framework successfully transforms geometric coordinates into verifiable information assets. By inheriting 2D image metadata, the resulting point clouds provide an intuitive interface for operators to rapidly identify critical features. Furthermore, the system utilizes an Attribute Table to facilitate a reprojection verification process, effectively transforming redundant observations from multiple stereo pairs into a robust metric for cross-validation. This integration ensures a geometrically coherent and queryable foundation for automated feature extraction and high-precision mapping tasks.

1. Introduction

Accurate 3D reconstruction of urban scenes from multi-view images is critical for applications in city planning and monitoring, self-driving routing, and digital twin development. To support these applications, dense image matching algorithms are typically employed to recover geometric information and generate dense 3D representations. However, the traditional point cloud lacks the information for the user to interpret. Although there are many studies about point cloud classification, they do this process after generating the point cloud data. This data form occupies numerous storage spaces and requires more energy to process. Thus, if the information already contained in the 2D image domain could be extracted and transmitted to the 3D space domain through the dense image matching process. Then, an "Attribute-Enriched Point Clouds," where 3D data is expected to carry not only geometric coordinates but also semantic and structural labels, could be generated directly. This integration is essential for automated feature extraction and more sophisticated spatial analysis in complex urban environments.

Despite the success of traditional dense stereo reconstruction, standard algorithms generally rely on low-level image features such as intensity, texture, or gradients for cost aggregation and disparity estimation. These approaches often struggle in urban scenes, where repetitive façade patterns, low-texture regions, and complex occlusions are common. Consequently, the resulting point clouds can be noisy or erroneous. Specifically, the precise reconstruction of structural information remains highly challenging. The lack or distortion of these geometric details renders point clouds less effective for high-accuracy measurement and complete scene understanding, representing a significant research focus awaiting a robust solution.

This study proposes an attribute-enriched dense image matching framework, building upon our previous research (Yang and Jaw, 2024), to generate point cloud information with enhanced precision and interpretability. The core of this framework is the integration of high-level semantic attributes and geometric conjugate features to guide the dense matching process and the subsequent generation of the 3D point cloud information. By embedding these multi-layered attributes, the proposed pipeline ensures that structural saliency is preserved throughout the reconstruction.

The primary contributions of this research are as follows:

- Conjugate feature-guided disparity initialization: This framework establishes reliable cross-view correspondences by embedding semantic and geometric features extracted from multi-view images. These correspondences serve as initial disparity seeds to reduce matching errors and improve matching efficiency in challenging urban areas.
- Attribute-constrained matching mechanism: The framework utilizes a multi-layer attribute map (combining semantic labels from models like SAM and geometric cues) to adaptively narrow down disparity searching ranges and enforce attribute consistency between stereo pairs.
- Attribute-enriched 3D Reconstruction: The proposed pipeline ensures that semantic and structural labels are preserved and propagated from 2D images to the final 3D representation, producing a structured and interpretable point cloud.

By coupling top-down semantic information with bottom-up image matching, this method facilitates the generation of high-quality point clouds while enhancing the continuity of edges and reducing outliers, demonstrating the potential of a scalable approach for high-precision urban mapping.

This contribution has been peer-reviewed.
https://doi.org/10.5194/isprs-archives-XLIX-B2-2026-305-2026

2. Literature Review

2.1 Progress in Semi-Global Matching (SGM)

Dense stereo matching aims to generate pixel-wise correspondences between rectified image pairs. SGM remains a benchmark due to its efficient approximation of global energy minimization (Hirschmüller, 2005). However, SGM is sensitive to its penalty parameters (P_1, P_2), often leading to noise in uniform areas or blurred boundaries. Recent studies have sought to reduce this sensitivity; for instance, Chuang *et al.* (2018) integrated cross-based matching with support regions to generate an initial disparity map, using a confidence map to automatically tune parameters and improve matching stability.

2.2 Feature-Constrained Stereo Reconstruction

To improve the precision of building corners and edges, geometric features are increasingly integrated into the matching pipeline. Chuang *et al.* (2018) introduced edge-constrained penalty tuning, utilizing the Edge Drawing (ED) algorithm to identify depth discontinuities. By enforcing adaptive support regions that do not cross image edges, this method preserves sharp boundaries without increasing outliers. While these methods use geometric feature seeds to guide matching, they still rely on low-level radiometric cues, which may struggle with highly repetitive urban patterns.

2.3 Region Matching under Semantic-and-Topological-Constraint

Beyond low-level features, high-level contextual information offers robust constraints for correspondence. Yang and Jaw (2025) proposed a matching strategy that integrates semantic segmentation with spatial topological constraints. This approach models the adjacency and geometric configuration, such as relative azimuth angles and distance ratios, between regions. Because these spatial relations are invariant to viewpoint changes and local occlusions, they allow for the rejection of topologically implausible matches. Experimental results confirm that satisfying both semantic consistency and geometric logic increases accuracy in structured environments, providing a reliable foundation for scene understanding across image pairs.

2.4 Attribute-Enriched Point Cloud Modeling

The transition from "data-centric" to "information-driven" reconstruction is a critical shift in urban mapping. Traditional point clouds often lack the contextual cues necessary to identify structural features in dense scenes. Yang *et al.* (2025) emphasized the preservation and propagation of attribute information, from the multi-view image domain through the generation process into the final 3D representation. By embedding corner-like features and semantic labels as metadata, attribute-enriched point clouds combine spatial geometry with descriptive traces. This approach shifts the focus from passive 3D capture to active modeling, where preserved image-domain information directly influences feature selection and improves the reliability of urban corner and edge measurements.

2.5 Research Significance and Contribution: Unified Attribute-Guided Framework

While previous research has optimized SGM through initial seeds, edge constraints, and topological relations, these elements are rarely unified into a single, cohesive, dense matching framework. Consequently, there remains a critical need for a system that can simultaneously resolve repetitive pattern ambiguities through topological constraints, preserve structural saliency via geometric edge traces, and ensure both precision and interpretability in the final 3D product through end-to-end attribute propagation. By integrating these multifaceted requirements, this study proposes an attribute-constrained-and-enriched point cloud information generation framework, aiming to directly supply both geometric and semantic information within the generated point cloud, thus accelerating the mapping and scene understanding tasks in a more reliable and applicable manner.

3. Methodology

The proposed framework aims to bridge the gap between 2D image analysis and 3D spatial reconstruction by embedding multi-layered attributes into the dense matching process. The overall workflow of this research is illustrated in Fig. 1. The methodology is systematically organized into three primary stages: First, a top-down semantic segmentation and bottom-up geometric feature extraction are performed in the image domain to identify structural anchors. Second, these attributes are utilized as conjugate feature seeds to initialize and constrain the dense matching searching range, effectively mitigating ambiguities in repetitive industrial textures. Finally, an attribute-enriched 3D reconstruction pipeline is executed to generate high-quality point clouds where every point carries coordinated semantic and structural information. The following sections provide a detailed description of each component within this integrated framework.

3.1 Attribute Acquisition and Initialization Pipeline

To retain both geometric saliency and scene-level contextual cues, the input multi-view images are processed through two parallel pipelines: semantic attribute acquisition and geometric feature acquisition.

3.1.1 Semantic Attribute Acquisition

Semantic attributes provide the high-level contextual cues necessary to resolve geometric ambiguities in dense urban scenes. This framework utilizes the Segment Anything Model (SAM), a high-performance foundation model for class-agnostic and promptable segmentation (Kirillov *et al.*, 2023), to generate a classified image.

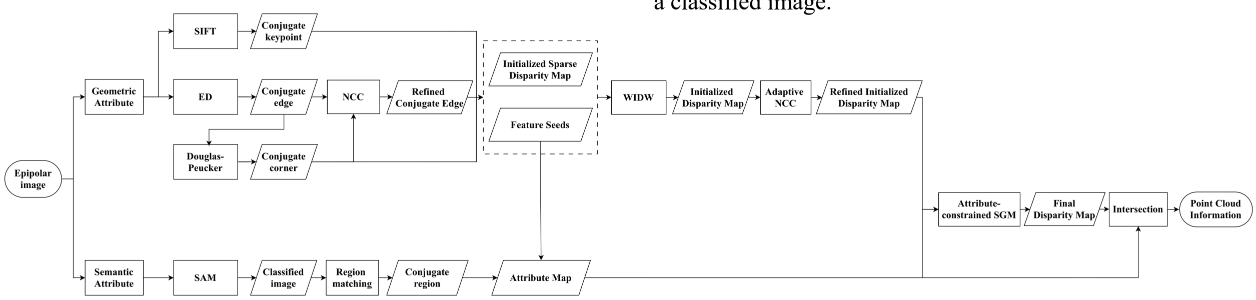


Fig. 1 Workflow

Following the initial segmentation, a region matching (Yang and Jaw, 2025) process is executed to establish conjugate correspondences between the segmented regions in the left and right stereo images. By identifying these corresponding regions across the image pair, the framework creates a spatial link between semantic entities, which is essential for guiding the subsequent dense matching algorithms and disparity searching range optimization.

Unlike traditional radiometric-based matching, this semantic layer assigns a categorical context to every pixel. These labels are instrumental in not only guiding disparity searching constraints but also preventing semantic mismatches, which eliminates false correspondences where pixels from distinct object classes are incorrectly matched due to high radiometric similarity.

By integrating these semantic labels, the matching process is no longer purely data-driven but is instead governed by the logical structure of the urban environment, ensuring that the final 3D reconstruction maintains both thematic and geometric consistency.

3.1.2 Geometric Feature Attributes

The extraction of geometric and structural attributes is fundamental to capturing the structural skeleton of the urban environment. In this framework, geometric associations are established through the integration of three distinct feature types, points, edges, and vertices, which collectively function as reliable conjugate features and geometric feature seeds for the subsequent matching stages.

The process begins with the identification of feature points using the Scale-Invariant Feature Transform (SIFT) to ensure radiometric saliency and robust correspondence across stereo pairs (Lowe, 2004). These keypoints provide point-based feature seeds that remain invariant to the scaling and rotation differences often encountered in multi-view urban imagery. To complement these sparse points with linear continuity, the Edge-Drawing (ED) algorithm is employed to extract contiguous and high-quality edge segments (Akinlar and Topal, 2011). Unlike conventional edge detection methods, ED generates clean structural boundaries that accurately represent potential depth discontinuities.

Furthermore, to refine these raw linear features into high-level structural nodes, the Douglas-Peucker algorithm is applied for recursive polyline simplification (Douglas and Peucker, 1973). By systematically filtering out redundant points while preserving the essential geometric trajectory, this step identifies critical vertices that correspond to building corners and structural intersections. The synergy between SIFT-based point saliency, ED-based linear continuity, and Douglas-Peucker vertex abstraction ensures that the matching process is guided by a comprehensive geometric logic, providing the necessary high-confidence feature seeds to stabilize the dense point cloud reconstruction.

3.1.3 Integration of Attribute Information

The extracted geometric and semantic data are structured into several intermediate products that serve as the computational foundation for the subsequent dense matching stages. This organizational phase transitions sparse image-domain features into a continuous spatial prior.

Feature Seeds and Initialized Sparse Disparity Map
The initial reconstruction stage is anchored by feature seeds, which are integrated from the previously established positions of edges, keypoints, and structural vertices. By leveraging the conjugate relationships of these seeds, the framework generates an initialized sparse disparity map. Despite their sparsity, these initial disparity values are highly trustworthy and represent a reliable geometric skeleton of the scene. This skeleton serves as a high-confidence reference baseline, providing the necessary geometric constraints to stabilize the subsequent dense matching.

Classified Image
The classified image is a pixel-wise semantic label map generated from the SAM-based segmentation process. It assigns a discrete semantic category to every pixel in the stereo pair. In the dense matching phase, this product acts as a rigorous semantic filter, ensuring that correspondence candidates are only considered valid if they share identical class labels, thereby eliminating cross-object mismatches.

Attribute Map
The attribute map functions as a fused, pixel-wise constraint layer that integrates high-level semantic labels from SAM with geometric structural traces. It serves as a constraint layer for dense image matching, enabling adaptive disparity searching ranges, structure-aware smoothing, and the preservation of sharp edges and corners. Besides, the attribute map provides the 3D points with their associated attributes.

Refined Initialized Disparity Map
The initialized disparity map is generated through a multi-step densification process of the initial sparse feature seeds. This process produces a continuous positional prior that serves as a critical reference for the subsequent SGM algorithm, providing the necessary guidance for precise matching and localization. The transformation from a sparse distribution to a dense Initialized Disparity Map is achieved using Weighted Inverse Distance Weighting (WIDW) interpolation to cover the entire image extent.

To account for the heterogeneous reliability of different feature types, the interpolation incorporates both category-based weights and spatial proximity (as illustrated in Fig. 2). The disparity value $D(p)$ for a target pixel p is calculated as a weighted average of its K nearest feature seeds, following the mathematical framework defined in Eq. (1).

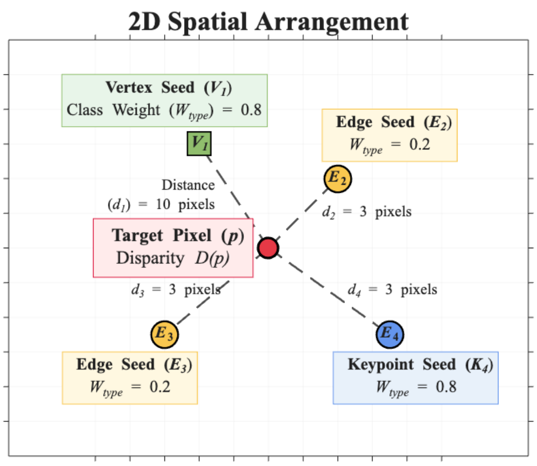


Fig. 2 Illustration of the WIDW interpolation mechanism considering spatial proximity and category-based reliability.

$$D(p) = \frac{\sum_{i=1}^{K}\left(W_{type,i} \cdot d_i^{-2} \cdot V_i\right)}{\sum_{i=1}^{K}\left(W_{type,i} \cdot d_i^{-2}\right)} \quad (1)$$

where V_i : the disparity of the i-th seed
d_i : Euclidean distance between the target pixel and the seed

W_{type} : a predefined weight corresponding to the feature category (Edge, Point, or Vertex)

This weighting scheme ensures that structural vertices and high-confidence edges exert a stronger influence on the surrounding disparity estimation.

Following interpolation, the disparity map undergoes a preliminary optimization using Normalized Cross-Correlation (NCC). To ensure that the target windows contain sufficient discriminative information for reliable matching, the NCC target window size is dynamically adjusted based on the distance to the nearest feature seeds, as visualized in Fig. 3. The adaptive window radius R is defined in Eq. (2):

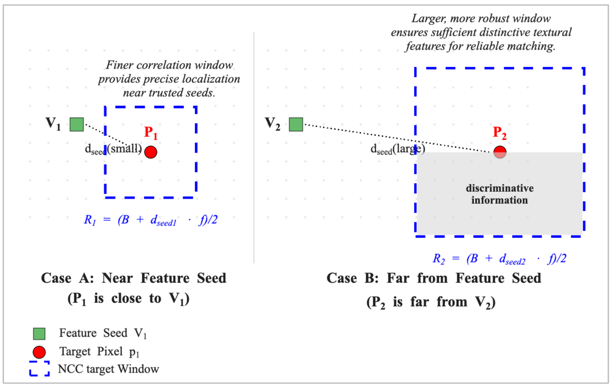


Fig. 3 Illustration of the adaptive NCC target window size selection based on the distance to feature seeds.

$$R(p) = \frac{(B + d_{seed} \cdot f)}{2} \tag{2}$$

where B: the base window size
d_{seed}: the distance to the nearest feature seed
f: a distance scaling factor

This dynamic adjustment mechanism maintains finer correlation windows in proximity to high-confidence feature seeds, while employing larger, more robust windows in sparse regions to encompass enough distinctive textural features. The resulting refined initialized disparity map provides a critical approximation for the subsequent dense matching stage, effectively narrowing the disparity searching range and significantly reducing computational overhead in challenging urban environments.

3.2 Attribute-Constrained SGM

Building upon the intermediate products, the proposed algorithm extends the "Class-aided Matching" concept (Yang and Jaw, 2024) by introducing a dual-layer constraint mechanism. It integrates the refined initialized disparity map as a positional prior and the attribute map and classified image as semantic-geometric filters to restrict the correspondence searching range.

3.2.1 Initialization-based Positional Priors and Adaptive Searching Windows

The refined initialized disparity map provides $d_{(init)(p)}$, which serves as the central anchor for the dense matching process. To suppress large deviations caused by repetitive façade patterns or occlusions, common failure points in standard SGM, the searching range for a corresponding pixel is restricted to a localized window around $d_{(init)(p)}$, as illustrated in Fig. 4. This constraint ensures that the matching cost is only evaluated within a high-probability interval, defined by Eq. (3):

$$d(p) \in \left[d_{(init)(p)} - \epsilon_a, d_{(init)(p)} + \epsilon_a\right] \tag{3}$$

where ϵ_a: an attribute-dependent searching tolerance

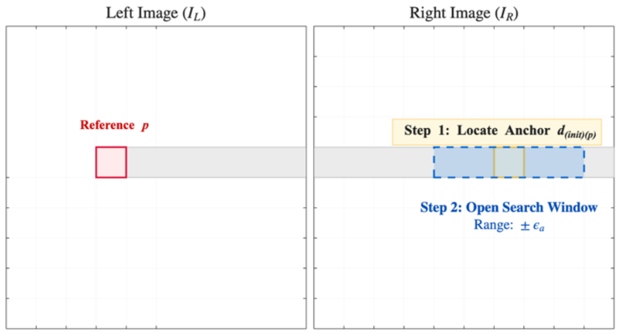


Fig. 4 Illustration of the constrained searching range based on the Refined Initialized Disparity Map.

Unlike traditional methods that employ a global constant threshold, ϵ_a is dynamically assigned based on the feature type defined in the attribute map. For instance, pixels associated with stable planar surfaces, such as roads or simple rooftops, are assigned a narrower ϵ_a to maximize computational efficiency and reliability. In contrast, pixels identified as complex building geometries or vegetation are granted a wider tolerance to accommodate higher local depth variance. This adaptive mechanism ensures that the searching range is precisely tailored to the expected geometric complexity of the urban feature.

3.2.2 Class Consistency Validation

To further enhance matching reliability, a strict filtering mechanism based on the classified image is enforced. Although within the restricted searching range defined in section 3.2.1, a candidate pixel must satisfy a semantic consistency check to be considered a valid match, as illustrated in Fig. 5. This requirement is mathematically expressed in Eq. (4):

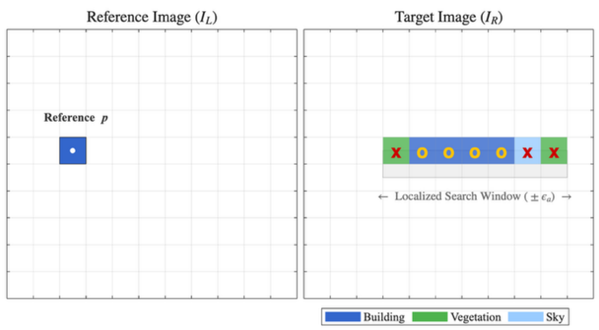


Fig. 5 Illustration of the semantic consistency filter for enhancing matching reliability based on the Classified Image.

$$Valid(d) \Leftrightarrow Class_{L(p)} = Class_{R(p+d)} \tag{4}$$

where $Class_L, Class_R$: denotes the semantic categories assigned to the reference and target pixels

This validation layer ensures that the matching process remains semantically logical, effectively preventing a pixel from being matched to a different object class even if their radiometric signatures are nearly identical.

By jointly enforcing these initialization-based positional priors, attribute-adaptive searching ranges, and semantic class verification, the proposed method effectively eliminates cross-object mismatches. This integrated constraint framework ensures that structural boundaries and urban object edges are reconstructed with high geometric precision and semantic awareness.

3.3 Generation of Point Cloud Information

The final stage of the proposed framework focuses on the transformation of the attribute-constrained matching results into a structured 3D product, emphasizing the seamless transition from image-domain perception to spatial measurement.

3.3.1 Dense 3D Reconstruction with Attribute Preservation

This stage utilizes the refined dense disparity map, where each pixel has been endowed with specific meaning through the hierarchical attributes obtained in previous processing steps. By leveraging the constraints expressed via semantic-geometric attributes, the final disparity map maintains strict geometric consistency while simultaneously functioning as a semantics-enriched data layer. This attribute retention not only enhances the reliability of depth estimates along object edges and discontinuities, where traditional methods often fail, but also provides a structured and interpretable representation of the entire scene.

3.3.2 Attribute-Enriched Point Cloud Information

The dense disparity map is subsequently converted into 3D point cloud information, where the primary objective is the preservation and propagation of the established attributes. In this 3D reconstruction, every individual point carries its geometric coordinates *(X, Y, Z)* alongside its associated semantic-structural metadata inherited from the classified image and attribute map.

Unlike traditional geometry-only point cloud data, the resulting attribute-enriched point cloud information maintains the full lineage of the image-domain information. By embedding this information into a structured metadata table for each point, the point cloud becomes an active data source for downstream applications. This enables higher-level tasks, such as precise urban measurement, object recognition, and automated scene analysis, to be performed with significantly greater reliability. The proposed pipeline thus demonstrates a complete transition from passive 3D capture to an information-driven modeling approach that effectively bridges the gap between perception and measurement.

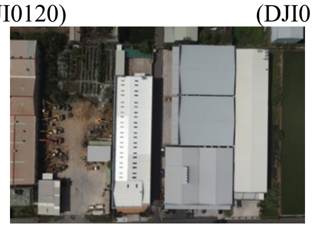
4. Experiment Result

4.1 Experimental Dataset and Environment

The experimental image datasets were provided by the Chinese Society of Photogrammetry and Remote Sensing (CSPRS). The study area is situated in an industrial district in Wuri, Taichung City, Taiwan. The original aerial images selected for this research are displayed in Fig. 6.

(DJI0099) (DJI0100)

(DJI0120) (DJI0122)



(DJI0124)

Fig. 6 Original image

4.1.1 Study Area and Scene Challenges

This study area is characterized by a high density of sheet metal factories and industrial sheds, which present significant challenges for traditional dense matching algorithms. The metallic corrugated rooftops exhibit highly repetitive patterns and specular reflections, often leading to severe matching ambiguities and salt-and-pepper noise in standard SGM-based reconstructions. Furthermore, the sharp, linear boundaries of these industrial structures require high geometric fidelity to avoid the error of depth information between the building eaves and the ground. The presence of factory buildings provides a rigorous environment for testing the effectiveness of attribute-enriched constraints in resolving radiometric ambiguities.

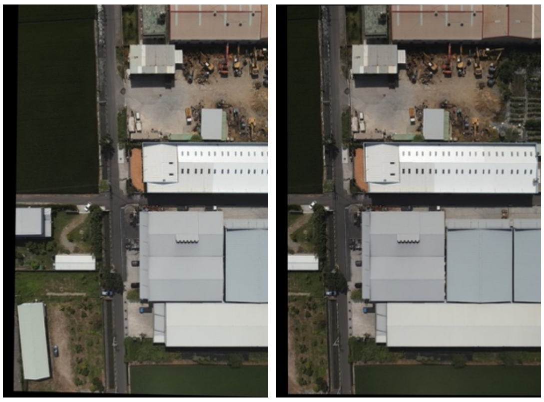
4.1.2 Data Specifications and Image Configuration

To evaluate the stability and robustness of the proposed algorithm, three sets of nadir-view stereo pairs were utilized. The detailed configurations of these pairs are summarized in Table 1, and the corresponding epipolar images used for dense matching are illustrated in Fig. 7. To ensure computational efficiency while maintaining critical structural features, the original images underwent a down-sampling pre-process, resulting in an operational resolution of approximately 400×600 pixels, and the average Ground Sample Distance (GSD) for this dataset is 32 cm.

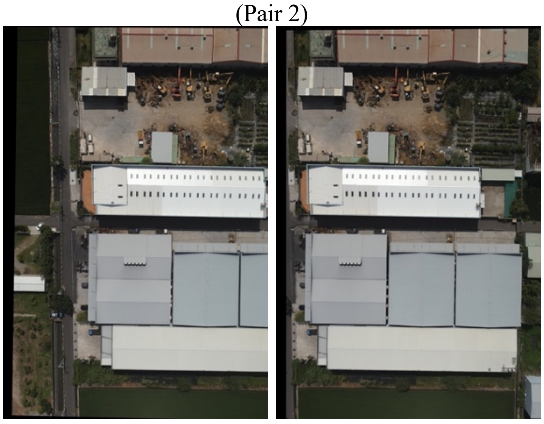
Name	Left image	Right image
Pair 1	DJI0099	DJI0100
Pair 2	DJI0120	DJI0122
Pair 3	DJI0122	DJI0124

Table 1. Experiment configuration

(Pair 1)



(Pair 2)



(Pair 3)

Fig. 7 Epipolar image pair

4.2 Attribute-Constrained SGM

The primary contribution of the proposed framework is the generation of point clouds that inherit multi-layer attributes from the image domain. To ensure a robust transition from 2D perception to 3D reconstruction, the framework systematically produces a series of intermediate attribute products that bridge the semantic and geometric domains. The following workflow is demonstrated using a representative stereo image pair to illustrate the evolution of these attributes.

The process begins with the classified image (Fig. 8) and the attribute map (Fig. 9), which capture the semantic categories and structural essence of the urban scene. These image-domain attributes are then propagated into the spatial domain through Feature Seeds, forming the initialized sparse disparity map (Fig. 10). This geometric skeleton represents the foundational 3D structure, which is subsequently densified into the initialized disparity map (Fig. 11). Finally, the map undergoes optimization to produce the refined initialized disparity map (Fig. 12), providing a high-precision positional prior to guide the final dense matching process.

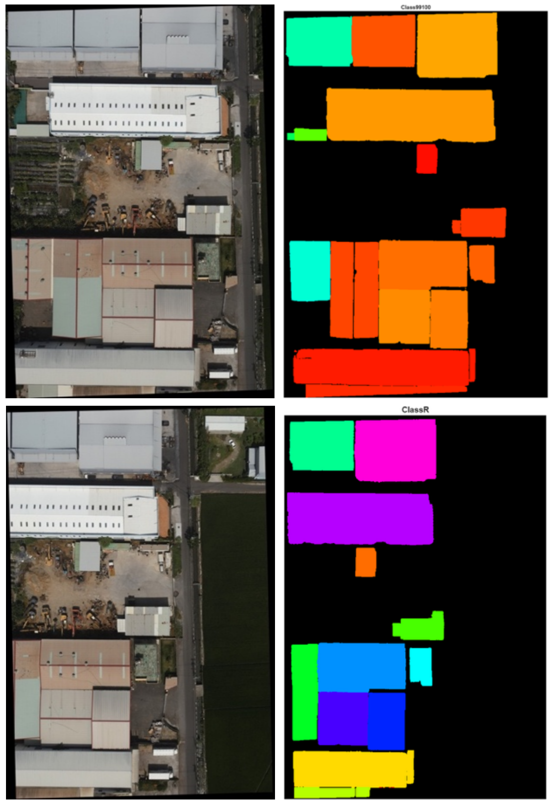


Fig. 8 Original image and classified image

Fig. 9 Attribute map

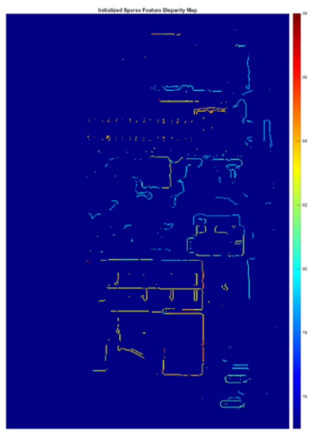


Fig. 10 Initialized sparse disparity map

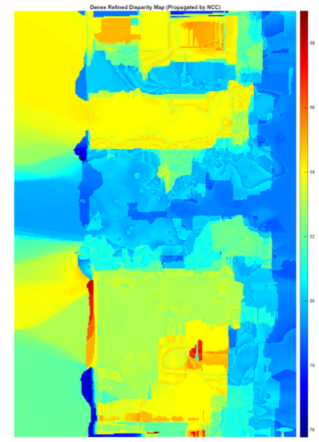


Fig. 11 Initialized disparity map

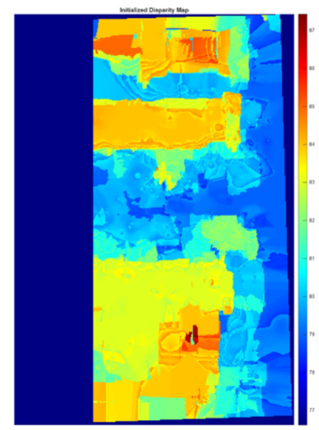


Fig. 12 Refined initialized disparity map

4.3 Point Cloud Information Generation

The final 3D reconstruction result is illustrated in Fig. 13.

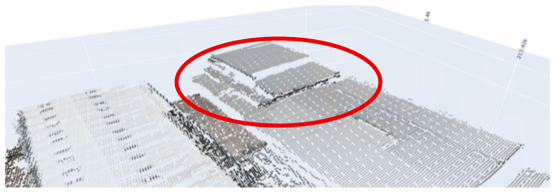
Fig. 13 Muti-view point cloud

This result serves as the foundational geometric output of the proposed algorithm.

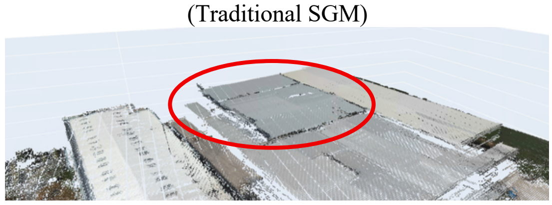
4.3.1 Geometric Completeness of Point Cloud

Geometric completeness serves as a primary indicator of reconstruction reliability, directly reflecting an algorithm's capacity to maintain surface continuity and structural integrity across challenging urban landscapes. The proposed framework shows its evidence on the rooftops, as shown in Fig. 14, where traditional SGM often suffers from structural thinning or breaks. The restriction of class-consistency and structural constraints ensures that boundaries remain sharp and the ground surface remains gap-free, successfully bridging the geometric deficiencies typical of unconstrained matching.

Furthermore, at the ground surface, as shown in Fig. 15, the raw data exhibits significant fractures. By utilizing refined disparity priors, the proposed approach effectively fills these gaps to produce a stable, continuous geometric surface.

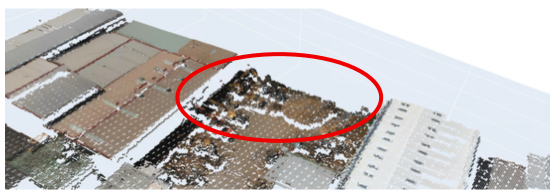


(Traditional SGM)



(Attribute-constrained SGM)

Fig. 14 Comparison of rooftop surface continuity



(Traditional SGM)

(Attribute-constrained SGM)

Fig. 15 Comparison of ground surface integrity

4.3.2 Attribute-Enriched Point Cloud Information

To further demonstrate the effectiveness of attribute propagation and the added value of multi-layer information, four specialized visualization scenarios are presented below:

Scenario 1: Semantic-Aware Point Cloud (Class Coloring)
As shown in Fig. 16, each point in the class-colored point cloud is assigned a semantic label derived from the Classified Image. By mapping the classID to a distinct colormap, we demonstrate that the reconstruction process successfully maintains semantic consistency. This visualization proves that the matching constraints effectively partitioned the scene into logical urban objects, which is a prerequisite for automated urban modeling.

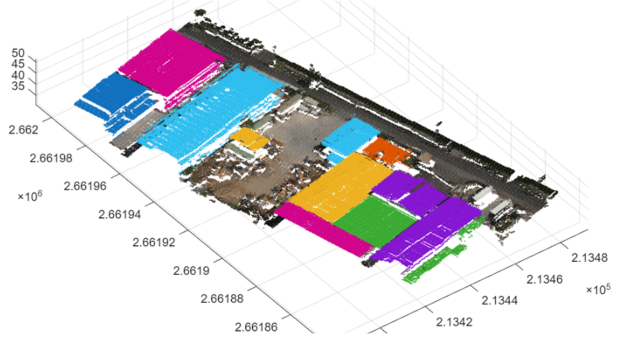


Fig. 16 class-colored point cloud

Scenario 2: Structural Skeleton and Feature Seed Distribution
The structural integrity of the point cloud, as shown in Fig. 17, is highlighted by isolating the feature seeds, such as edges (yellow), keypoints (red), and vertices(green), against the background geometry. This visualization demonstrates the successful 3D propagation of the 2D conjugate features. The dense distribution of vertices (corners) and edges along building rooflines confirms that the algorithms provided a reliable geometric skeleton that stabilized the dense matching process.

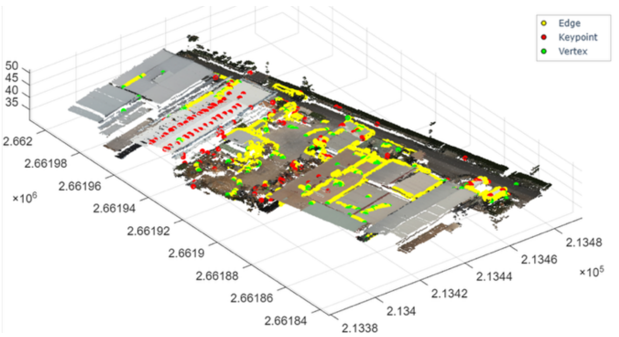


Fig. 17 Feature-highlighted point cloud

Scenario 3: Boundary Precision and Regional Interpretation
To evaluate the preservation of object limits, the point cloud, as shown in Fig. 18, is visualized by enhancing the boundary IDs on top of the classified regions. These boundaries are systematically extracted from the discrete segments within the classified image. However, a characteristic phenomenon observed in the Fig. 19 is that certain boundary points may not perfectly align with the physical roof ridges. This is primarily attributed to the 2D-to-3D projection perspective. Since the SAM-based semantic partitioning is performed in the 2D image domain, the resulting masks occasionally encapsulate portions of the building facades or eaves, depending on the camera's shooting angle, particularly in oblique imagery. Thus, when these 2D semantic boundaries are propagated into the 3D domain, they represent the apparent visual limits of the object from a specific viewpoint. While this introduces an offset relative to the idealized roof skeleton, the class consistency check ensures that these points remain semantically correct, providing a robust baseline for further structural refinement or building model reconstruction.

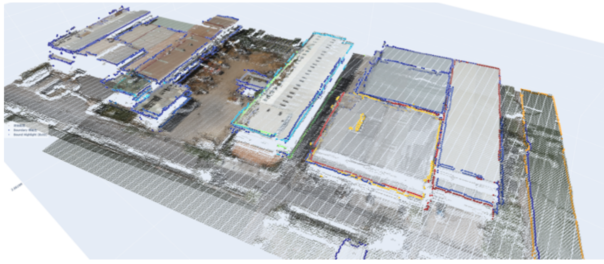


Fig. 18 Boundary-highlighted point cloud

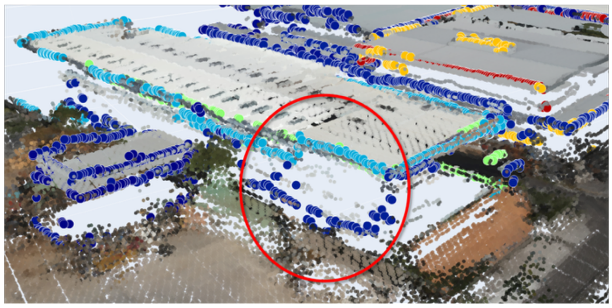
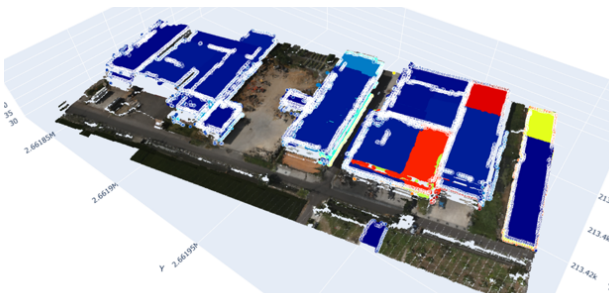


Fig. 19 Visualization of Boundary Misalignment at Building Roof Ridges

Scenario 4: Integrated Attribute-Enriched Representation
The final scenario provides a comprehensive view of the attribute-enriched point cloud information, overlaying semantic regions, enhanced boundaries, and structural feature seeds, as shown in Fig. 20. This integrated visualization represents the final product of our pipeline: a multi-layered data structure where every point carries its spatial coordinates (X, Y, Z), its semantic identity, and its structural role. As shown in Fig. 21, this integrated representation is further exemplified by the associated attribute table, where every spatial point is systematically linked to its semantic category, boundary identity, and structural role (e.g., SIFT seed or vertex). This data structure underscores the transition from a 'data-centric' point cloud to an information-driven representation, providing a queryable foundation ready for automated feature extraction and high-precision measurement.



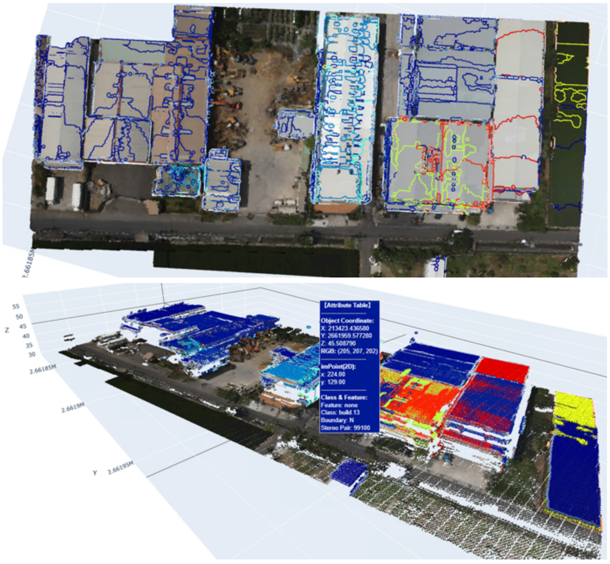


Fig. 20 Colored attribute-enriched point cloud

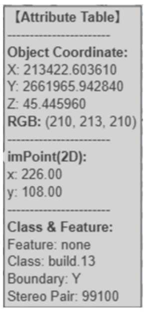
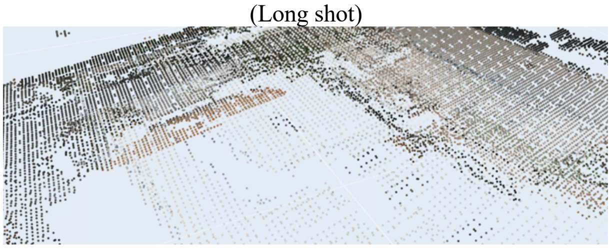


Fig. 21 Attribute table of the point cloud information

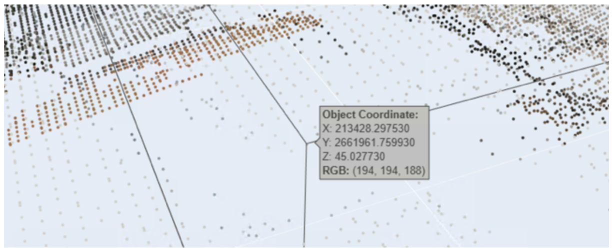
4.3.3 Comparative Analysis: From Raw Point Cloud Data to Attribute-Enriched Information

This section presents a comparative analysis between the products of the traditional SGM and the proposed attribute-constrained-and-enriched point cloud information generation framework. The experimental results demonstrate a fundamental shift from generating raw point cloud data to producing structured point cloud information. While traditional SGM provides basic geometric coordinates *(X, Y, Z)* and color information *(RGB)*, as shown in Fig. 22, it often falls short in complex urban scenarios where edge blurring and matching noise hinder precise feature identification.

(Long shot)

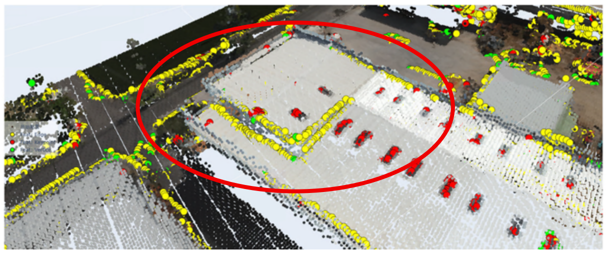


(Close-up)

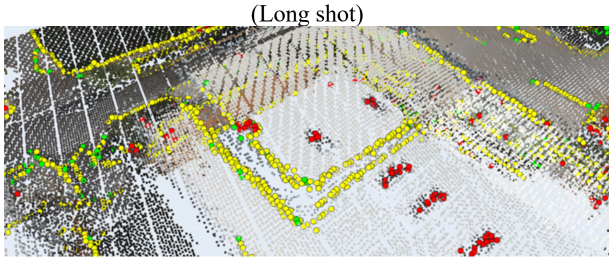


(Point cloud data)
Fig. 22 Raw point cloud data

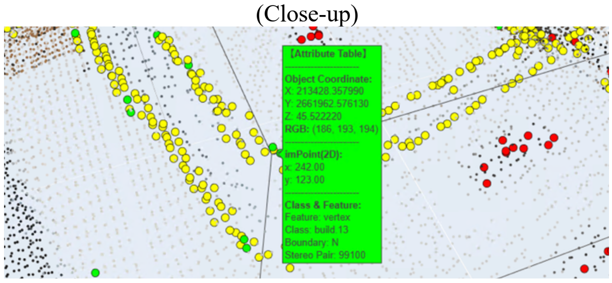
The primary advantage of the proposed "attribute-enrichment" framework lies in transforming these simple semantic-geometric collections into high-fidelity information assets. In scenarios where traditional point clouds present ambiguous point clusters, the integration of structural attributes allows for the rapid identification of specific geometric characteristics. For instance, as illustrated in Fig. 23, the structural vertices can be instantly distinguished from surrounding interference information. This capability enables operators to quickly evaluate and select the precise points required for engineering or modeling tasks, significantly reducing manual interpretation errors.



(Long shot)



(Close-up)



(Point cloud information colored by features)
Fig. 23 Attribute-enriched point cloud information showing green vertex markers

It is able to capitalize on redundant observations generated across multiple stereo configurations. A single spatial location is frequently captured by various stereo pairs, as shown in Fig. 24, resulting in multiple candidate points for a single feature. The system utilizes this redundancy as a robust source for cross-validation. By leveraging the comprehensive Attribute Table, which includes object coordinates, image coordinates, and stereo pair sources, the system facilitates the reprojection verification process. As shown in Fig. 25, this allows for a direct consistency check between the 3D point position and its original 2D image location. Consequently, mismatched points caused by repetitive façade patterns or occlusions can be systematically excluded, ensuring that the final output is not just a collection of points, but a reliable, verifiable, and information-enriched 3D point cloud.

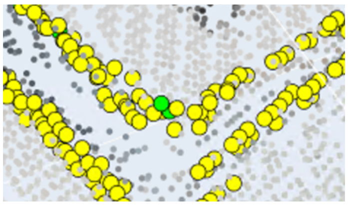


Fig. 24 Multiple candidate points from different image pairs

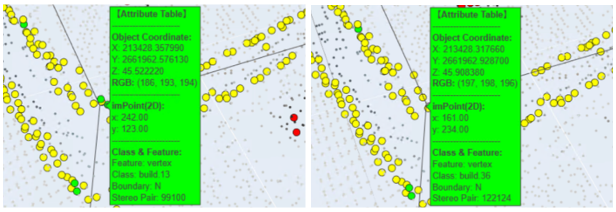


Fig. 25 Illustration of the attribute table application

5. Conclusion

This study has developed an attribute-enriched dense matching framework specifically designed to overcome reconstruction challenges in complex industrial and urban environments. By transforming raw point cloud data into structured point cloud information, the proposed method effectively mitigates mismatches and geometric blurring commonly found at large disparity discontinuities.

Experimental results demonstrate the robustness of the framework across multiple stereo configurations. A key contribution of this research is the fundamental transition from simple *(X, Y, Z, RGB)* data to an attribute-enriched output. Through the integration of attribute-constrained SGM and attribute map, the system provides stable anchors that allow operators to rapidly identify and verify precise feature points even in high-noise or low-texture scenarios. This capability significantly enhances the interpretability of 3D models compared to traditional unconstrained matching methods.

Furthermore, the framework capitalizes on the redundant observations generated by multiple stereo pairs at the same spatial location. By utilizing the comprehensive attribute table, containing object coordinates, image coordinates, and stereo source data, the system facilitates a rigorous reprojection verification process. This mechanism transforms multiple candidate points from an ambiguity into a robust source for cross-validation, ensuring that the final 3D model is both geometrically coherent and semantically reliable. While minor boundary offsets due to perspective projection remain a challenge, the framework provides a high-fidelity foundation for automated urban mapping, modeling, and industrial asset management. Future work will focus on optimizing 3D boundary refinement to further mitigate perspective-induced shifts and exploring the integration of multi-view oblique imagery for complete façade reconstruction.

References

Akinlar, C., Topal, C., 2011. EDLines: A real-time line segment detector with a false detection control. *Pattern Recognition Letters*, 32(13), pp. 1633-1642.

Chuang, T.Y., Ting, H.W., Jaw, J.J., 2018. Dense stereo matching with edge-constrained penalty tuning. *IEEE Geoscience and Remote Sensing Letters*, 15(5), pp. 664-668.

Douglas, D.H., Peucker, T.K., 1973. Algorithms for the reduction of the number of points required to represent a digitized line or its caricature. *Cartographica: The International Journal for Geographic Information and Geovisualization*, 10(2), pp. 112-122.

Hirschmuller, H., 2005. Accurate and efficient stereo processing by semi-global matching and mutual information. *In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, San Diego, CA, USA, Vol. 2, pp. 807-814.

Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A.C., Lo, W.Y., Dollár, P., 2023. Segment anything. *In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 4015-4026.

Lowe, D.G., 2004. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2), pp. 91-110.

Yang, Y.C., Jaw, J.J., 2024. Cross-based matching constrained by the classes of pixels. *In: Proceedings of the 45th Asian Conference on Remote Sensing (ACRS 2024)*, Colombo, Sri Lanka, Paper AB0218.

Yang, Y.C., Jaw, J.J., 2025. Fusing semantic segmentation and topological constraints for region matching of stereo pair. *In: Proceedings of the 43rd Conference on Surveying and Geomatics*, Yilan, Taiwan. (in Chinese)

Yang, Y.C., Chen, Y.Y., Jaw, J.J., 2025. From data to information: evaluating attribute-enriched point clouds for accurate urban corner and edge measurement (abstract only). *In: Proceedings of the 46th Asian Conference on Remote Sensing (ACRS 2025)*, Makassar, Indonesia.

Acknowledgements

This research was supported by the National Science and Technology Council (NSTC 113-2121-M-002-006) and the Chinese Society of Photogrammetry and Remote Sensing. The authors gratefully acknowledge their financial support and contribution to this work.