Building Height Estimation from Stereo Satellite Images Using Contour Vector Registration

Yaxuan Duan [1], Wei Qin [1], Xin Huang [1], Pei Mi [1], Yang Yu [1], Pengjie Tao [1,*]

[1] School of Remote Sensing and Information Engineering, Wuhan University, Wuhan 430079, China - yxduan@whu.edu.cn, wei-qin@whu.edu.cn, lavender@whu.edu.cn, mipei2023@whu.edu.cn, yang.yu@whu.edu.cn, pjtao@whu.edu.cn

Keywords: Building height estimation; Stereo satellite images; Contour vector registration; Vertical Line Locus (VLL); High-resolution images.

Abstract

Accurate building height estimation plays a crucial role in large-scale 3D urban reconstruction. However, conventional stereo matching approaches often suffer from mismatches around building edges, leading to unreliable height retrieval in dense urban areas. To address this issue, this paper presents a novel method for building height estimation based on contour vector registration integrated with the vertical line locus technique. The proposed framework first automatically matches building contour vectors extracted from stereo high-resolution satellite images. Then, for each paired contour, a range of candidate heights is searched using a rational function model to project the reference contour from the image space to object space and then reproject it onto the conjugate image. The elevation that maximizes the overlap ratio between projected and paired contours is identified as the optimal roof elevation. Building height is subsequently derived by subtracting the ground elevation from the estimated roof elevation. Experiments conducted on SuperView-1 (SV-1) satellite stereo images over Jiuyuan District, Baotou, Inner Mongolia, China, demonstrate the effectiveness of the proposed method. The resulting building height estimates achieve a root mean square error of 0.84 m compared to manual measurements, showing strong agreement (r = 0.9993). The proposed contour-based stereo registration approach provides a robust and efficient solution for building height extraction from high-resolution satellite data, supporting precise urban 3D modeling and large-scale spatial analysis.

1. Introduction

Accurate building height estimation is essential for understanding urban morphology, supporting three-dimensional (3D) city modeling, and facilitating environmental and climate studies (Frantz et al., 2021; Weng et al., 2024; Wu et al., 2023). High-resolution stereo and multi-view satellite images have become indispensable data sources for large-scale 3D urban mapping (de Franchis et al., 2014; Du et al., 2024b; Tang et al., 2020). Building height data derived from satellite observations play a vital role in applications such as urban expansion monitoring, sustainable city design, and disaster assessment (Du et al., 2024a; Javanroodi et al., 2023). However, extracting reliable height information in dense urban areas remains challenging due to geometric distortions, radiometric inconsistencies, and the complex structures of man-made environments (Fraser & Hanley, 2005; Xu et al., 2024). The fusion of multisource remote sensing data—combining optical, SAR, and LiDAR observations—has emerged as a promising pathway to overcome these issues and improve model robustness (Lee et al., 2020; Li et al., 2021; Zhao et al., 2023). Recent studies further highlight the integration of multi-modal observations within deep learning frameworks to enhance the accuracy of large-scale urban height mapping (Li et al., 2023; Ma et al., 2024).

Traditional approaches for building height estimation are generally divided into shadow-based and stereo-based methods. Shadow-based techniques infer building height by analyzing the geometric relationship among shadow length, solar elevation angle, and image orientation (Qi et al., 2016; Xie et al., 2021). Although simple and cost-effective, these methods are highly sensitive to solar position, terrain slope, and occlusion conditions, which limits their application in urban environments (Du et al., 2024b; Zhang & Gruen, 2006). In contrast, stereo-based techniques utilize disparity between homologous pixels across multi-view images to derive elevation (Luo et al., 2016; Song et al., 2021; Wu et al., 2025). Despite their higher theoretical accuracy, conventional dense stereo matching often struggles with mismatches near building edges or regions of low texture (Memar et al., 2025; Shi et al., 2025). With the advent of deep learning, convolutional and transformer-based networks have been introduced to enhance stereo matching and directly infer building height from optical or SAR data (Chen et al., 2021; Miah et al., 2025; Prabhakar & Garg, 2023). For instance, instance-level multitask frameworks can jointly predict building boundaries and height values from single-view images (Shahin & Almotairi, 2021), while multi-level feature fusion models such as FusionHeightNet leverage multisource data to improve building height estimation at the city scale (Wang & Sertel, 2023). Meanwhile, multimodal data integration has been shown to improve the resilience of height mapping under varying urban, economic, and environmental conditions (Du et al., 2024b; Sawa et al., 2024).

Another key challenge lies in the accurate extraction and registration of building contours, which form the foundation for height estimation and 3D reconstruction. Vectorized building footprints extracted from high-resolution imagery provide geometric regularity and semantic structure that can substantially enhance stereo registration (Alidoost et al., 2019; Chen et al., 2020; Qin et al., 2018). Deep learning models such as Mask-Height R-CNN (Huang et al., 2016), SDBN-HCWO (Miah et al., 2025), and edge-based networks (Li et al., 2024; Zhao et al., 2022) have demonstrated superior performance in extracting precise building outlines from complex urban scenes. These vector-based approaches outperform raster correlation methods by exploiting polygon-level consistency and improving computational efficiency (Rau et al., 2024). In addition, emerging techniques employing attention-based networks (Lee

* Corresponding author

et al., 2024) and transfer learning (Zhao et al., 2023) have shown potential to improve contour registration robustness under varying spatial resolutions and imaging conditions.

In this study, we propose a novel building height estimation method that integrates contour vector registration with the vertical line locus (VLL) approach. The method combines the structured representation of vector contours with the geometric rigor of the VLL framework. Building contours extracted from stereo satellite imagery are projected into object space via the rational function model (RFM) and reprojected onto conjugate images over a range of candidate elevations. The optimal roof elevation is determined by maximizing the geometric overlap ratio between projected and paired contours, and the final building height is derived by subtracting the ground elevation from a digital elevation model (DEM). The method is validated using SuperView-1 (SV-1) satellite stereo imagery over Jiuyuan District, Baotou, Inner Mongolia, China, an area featuring diverse urban morphology and relatively flat terrain suitable for photogrammetric evaluation. Experimental results show that the proposed method achieves a root mean square error (RMSE) of 0.84 m and a correlation coefficient of 0.9993 compared with manual measurements, confirming its high accuracy and robustness for large-scale urban building height extraction (Shahin & Almotairi, 2021; Wang & Sertel, 2023; Xu et al., 2024). The proposed framework contributes to the growing field of 3D urban reconstruction by bridging pixel-level deep learning approaches and vector-level geometric modeling, offering a scalable solution for precise building height estimation from high-resolution satellite imagery.

The remainder of this paper is organized as follows. Section 2 details the study areas and experimental data. Section 3 introduces the methodology, including the mathematical formulation of contour vector registration and the VLL-based height optimization. Section 4 presents and discusses the experimental results. Section 5 concludes the paper and provides future research perspectives.

2. Study Area and Datasets

2.1 Study Area

The study area is located in Jiuyuan District, within the urban core of Baotou City, Inner Mongolia, China (≈ 40°37′ N, 109°50′ E). The terrain is relatively flat, with elevations ranging from 970 m to 1100 m above sea level, and the region contains a mixture of residential, industrial, and commercial buildings typical of northern Chinese cities. The moderate relief minimizes terrain-induced distortions in stereo projection, making the area suitable for evaluating the geometric accuracy and robustness of the proposed building-height estimation approach.

2.2 Datasets

This study utilizes high-resolution stereo images acquired by the SV-1 satellite system; a high-performance optical platform specifically designed for precise photogrammetric applications. The satellite can capture 0.5 m panchromatic and 2.0 m multispectral imagery with accurate geometric calibration. Its along-track stereo configuration provides sufficient parallax and radiometric consistency for reliable building height estimation. As shown in Figure 1, a stereo pair covering the Jiuyuan District of Baotou, Inner Mongolia, China, was acquired on 15 August 2022, with an area extent of approximately 175.57 km². The key imaging parameters are summarized in Table 1.

The stereo pair includes the left and right panchromatic images together with automatically extracted building contours that serve as geometric inputs for the subsequent processing.

Parameter	Specification
Satellite system	SuperView-1 (SV-1)
Acquisition date	15 August 2022
Spatial resolution	0.5 m (panchromatic), 2 m (multispectral)
Quantization	11 bits
Orbit	Sun-synchronous
Stereo mode	Along-track (Forward/Backward)
Area coverage	~175.57 km²

Table 1. Key specifications of the SV-1 stereo imagery utilized in this study.

In addition to the primary stereo imagery, three ancillary datasets were utilized for processing and validation:

(a) DEM Data: A pre-existing DEM with a 5 m spatial resolution was used. This data was generated from Ziyuan-3 (ZY-3) three-line array stereo imagery acquired in September 2022. The initial Digital Surface Model (DSM) derived from stereo matching was subsequently processed using automatic terrain filtering, followed by manual verification and editing, to obtain the bare-earth DEM. With an estimated vertical accuracy of approximately 3 m, it serves as a sufficient initial ground elevation (Z_ground) reference for the subsequent building height derivation.

(b) Vector Building Contours: In this study, contour extraction and their initial stereo correspondences are treated as preprocessing steps. Building footprints were obtained using a

(a) (b) (c)

Figure 1. SV-1 stereo imagery over Jiuyuan District. (a) Left image. (b) Right image. (c) Sample of extracted contours

pre-trained Mask R-CNN, and initial pairs were established via basic spatial overlap constraints. As the core contribution of this paper focuses on the subsequent VLL-based height estimation, detailed deep learning architectures and pairing metrics are omitted. These vectorized contours seamlessly served as the foundational geometric primitives for our proposed optimization framework. An example of these extracted contours, shown in green, is overlaid on the left image in Figure 1c.

(c) Reference Height Data: A high-accuracy validation dataset was prepared for 3,346 buildings. To obtain reliable ground truth, homologous points of building roofs and their corresponding ground bases were manually measured across the stereo images using precise photogrammetric intersection to calculate the actual building heights. These rigorous manual measurements served as the ground truth for the quantitative accuracy assessment of our results.

3. Methodology

3.1 Overview of the Proposed Framework

The proposed method estimates building height by integrating contour vector registration with VLL principle. Unlike pixel-based stereo matching, which depends on dense correlation and is prone to mismatches near edges, this framework performs matching in the object domain, using geometric consistency between building contours extracted from stereo satellite imagery.

The overall workflow consists of four main steps: (1) contour extraction and pairing from stereo images; (2) geometric projection of contours from the reference image to object space using the RFM; (3) reprojection of object-space contours onto the conjugate image at discrete elevation hypotheses using the VLL concept; and (4) determination of the optimal roof elevation by maximizing geometric overlap between projected and paired contours.

In this paper, *elevation* refers to the absolute height above mean sea level (i.e., roof or ground elevation), while *building height* denotes the relative difference between roof and ground elevations.

3.2 Geometric Foundation

All geometric transformations between the 2D image space and the 3D object space are performed using RFM, which is provided with the SV-1 satellite metadata. The RFM establishes the mathematical relationship between image coordinates (r, c) and ground coordinates (X, Y, Z) through rational polynomial functions, without explicitly modeling the sensor's physical properties.

The forward RFM, which projects a 3D ground point to 2D image coordinates, is generally expressed as:

$$\begin{aligned} r_n &= \frac{P_1(X_n, Y_n, Z_n)}{P_2(X_n, Y_n, Z_n)} \\ c_n &= \frac{P_3(X_n, Y_n, Z_n)}{P_4(X_n, Y_n, Z_n)} \end{aligned} \quad (1)$$

where (r_n, c_n) and (X_n, Y_n, Z_n) are the normalized image and ground coordinates, respectively. P_1, P_2, P_3, P_4 are third-order rational polynomial coefficients (RPCs) are supplied with the imagery. The inverse RFM, used to project an image point back to the object space at a given elevation, is formulated similarly.

This model facilitates accurate and efficient projection and reprojection of contour vertices.

3.3 VLL-Based Contour Registration and Elevation Optimization

The core of our method is an extension of the classical VLL concept. Traditionally, for a given ground point (X, Y) with an unknown height Z, the VLL defines a vertical line in the object space. As the elevation Z varies along this line, its projection onto a stereo image pair traces a corresponding path.

In this study, we advance this principle from a single point to an entire building contour. As illustrated in Figure 2, the process unfolds as follows:

- **Define Elevation Hypotheses:** For each paired contour, a search range of candidate elevations $\{Z_i\}$ is established, from a plausible minimum (Z_{min}) to a maximum (Z_{max}).
- **Projection and Reprojection:** For each elevation hypothesis Z_i, every vertex of the reference contour in the left image is projected into the 3D object space using the inverse RFM. This creates a planar 3D contour at height Z_i.
- **Matching in Conjugate Image:** This 3D contour is then reprojected onto the right image using the forward RFM, resulting in a projected 2D contour, which we denote as $C_{p,i}$.

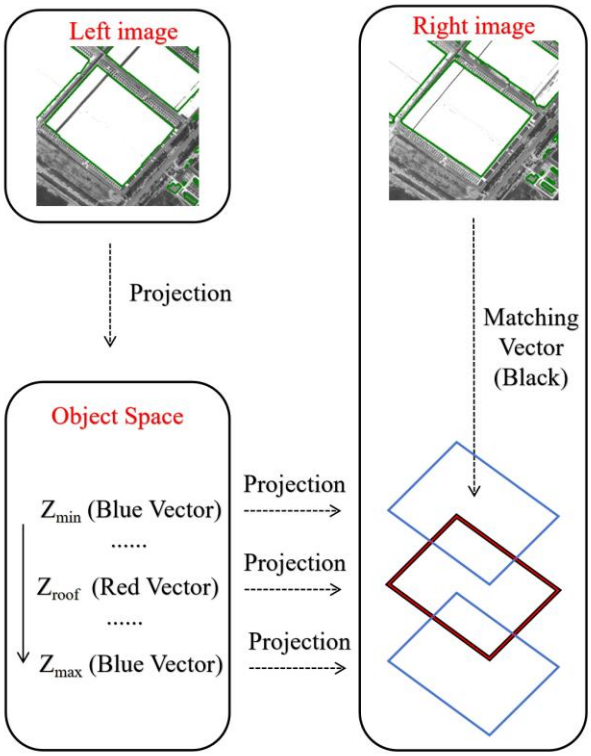

Figure 2. Schematic of the proposed VLL-based contour registration method. The building contour from the left image is projected into 3D object space under different elevation hypotheses (e.g., Z_{min}, Z_{max}). These 3D contours are then reprojected onto the right image. The optimal roof elevation (Z_{roof}) is determined when the reprojected contour (red) achieves the maximum overlap with the corresponding paired contour (black).

3.4 Matching Metric and Optimization

To quantify the geometric consistency, we calculate the overlap between the projected contour $C_{p,i}$ and its corresponding original contour in the right image, C_r. We employ the Intersection over Union (IoU) as the matching metric, which is a robust measure against differences in shape and scale. The IoU for each elevation hypothesis i is calculated as:

$$IoU_i = \frac{S(C_{p,i} \cap C_r)}{S(C_r)} \quad (2)$$

where $S(\cdot)$denotes the polygon area.

The optimal roof elevation Z_{roof}, is identified by finding the elevation hypothesis that maximizes the IoU score:

$$Z_{\text{roof}} = \arg\max_{Z_i} IoU_i \quad (3)$$

After identifying Z_{roof}, the building height H is derived by subtracting the ground elevation Z_{ground} obtained from the DEM:

$$H = Z_{\text{roof}} - Z_{\text{ground}} \quad (4)$$

This vector-based optimization replaces dense pixel correlation with a purely geometric measure, enhancing robustness against illumination differences and radiometric noise. It also ensures that the final results preserve the sharp structural boundaries of buildings and maintain strong geometric consistency across stereo pairs.

To intuitively illustrate the variation pattern, Figure 3 shows the relationship between IoU_i and elevation hypothesis Z_i. Starting from Z_{min}, as Z_i increases, the projected contour gradually aligns with C_r, so IoU_i rises and approaches 1. After reaching the peak at Z_{roof} (optimal elevation), continuing to increase Z_i causes the projected contour to deviate again, making IoU_i decline. Through interpolation around this peak (e.g., cubic spline interpolation), we can precisely determine Z_{roof} for sub-unit elevation accuracy.

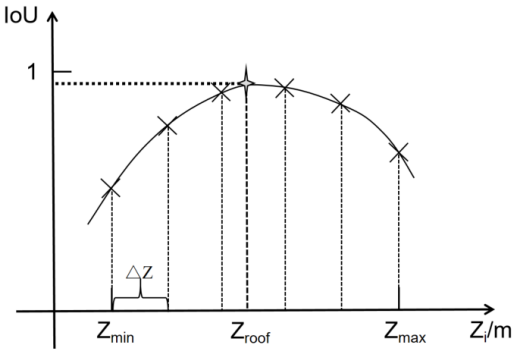

Figure 3. Schematic of the proposed VLL-based contour registration method.

The selection of elevation step size (ΔZ) directly affects the resolution and computational efficiency of the proposed contour-based height estimation. As illustrated in Figure 4, ΔZ is determined adaptively according to the ratio between the object-space elevation range ($Z_{max} - Z_{min}$) and the corresponding projection displacement between the left and right images. This adaptive strategy allows finer elevation sampling in high-relief areas while maintaining computational efficiency in relatively flat regions.

4. Results and Discussion

4.1 Building Roof Elevation Estimation

Based on the proposed VLL-based framework, roof elevations were estimated for 42,173 buildings across the study area. Table 2 summarizes the statistical distribution of these roof elevations.

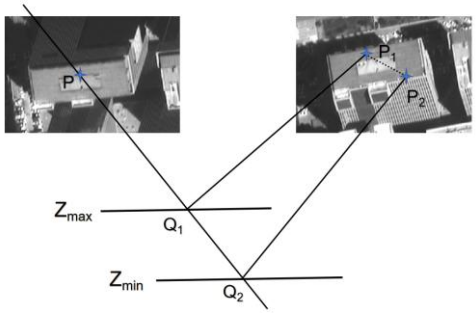

Figure 4. Determination of elevation step size (ΔZ) based on projection displacement between left and right stereo images.

The estimates predominantly range between 1000 m and 1100 m, which is consistent with the regional topography and demonstrates that our method preserves the underlying terrain gradient.

Elevation range (m)	Number of buildings	Percentage (%)
970–980	312	0.7
980–990	4089	9.7
990–1000	4619	10.9
1000–1100	33043	78.3
1100–1200	110	0.3

Table 2. Distribution of estimated roof elevations.

The roof elevations were then converted to building heights by subtracting the corresponding ground elevations from the DEM. The statistical distribution of the final building heights is summarized in Table 3. The results show that building heights range from approximately 2 m to 40 m. To better understand the urban morphology, the buildings were classified according to the Code for Design of Civil Buildings (assuming an average story height of 3 m). The classification reveals a predominance of low-rise (<9 m) and multi-story (9–18 m) buildings, which collectively account for over 92% of the total building stock. This aligns perfectly with the morphological characteristics of Jiuyuan District, where residential zones and industrial parks dominate the landscape.

Building Type	Height Range (m)	Number of Buildings
Low-Rise (1-3 stories)	<9	31081
Multi-Story (4-6 stories)	9-18	8029
Mid-Rise (7-9 stories)	18-27	1263
High-Rise (≥10 stories)	>27	1800

Table 3. Classification of building heights in the study area.

The spatial distribution of the calculated building heights is visualized in Figure 5. This figure presents the original satellite imagery (5a, 5d), the extracted building contours (5b, 5e), and the final color-coded height map (5c, 5f). The results demonstrate that our vector-based method successfully captures

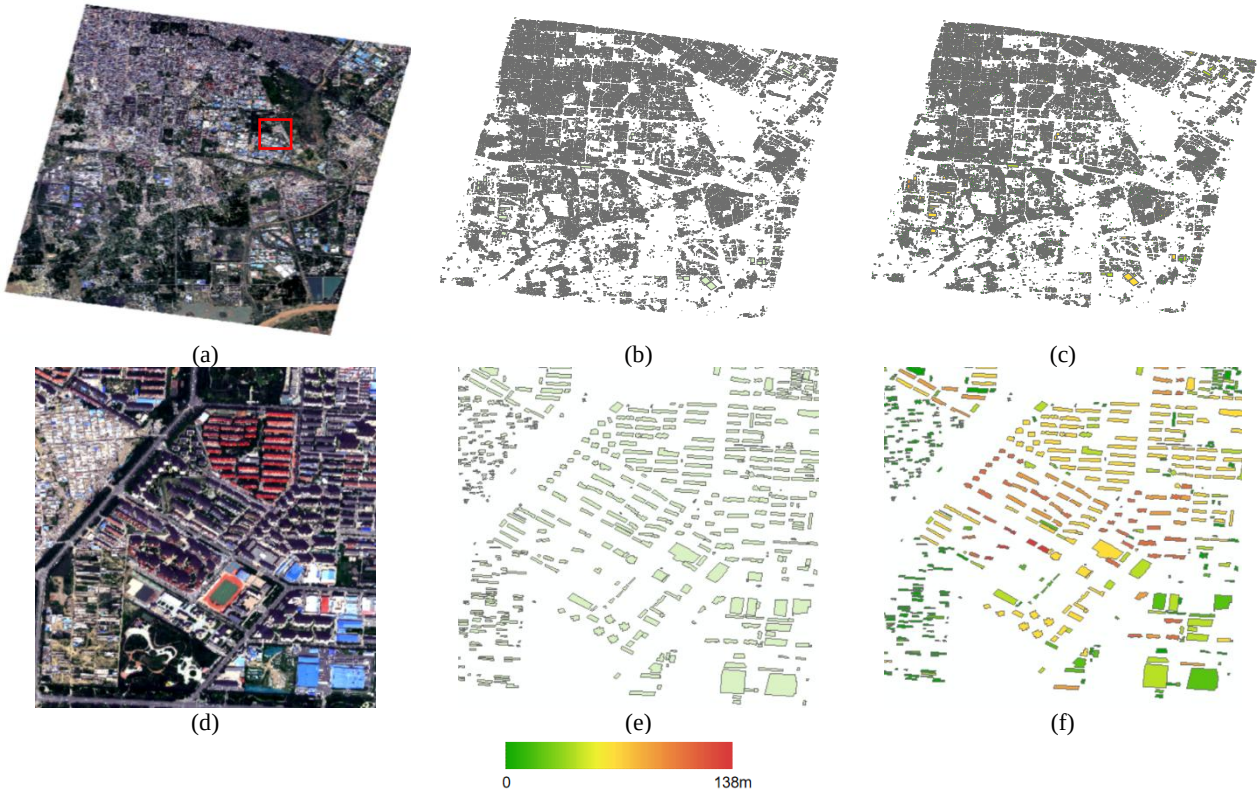

Figure 5. Building height estimation results for the study area. (a) Original image. (b) Extracted building contours. (c) Calculated building heights. (d-f) show corresponding close-up views of a subregion.

subtle height differences between neighbouring buildings, a task often challenging for conventional pixel-based stereo matching due to edge-related artifacts.

4.2 3D Reconstruction and Visualization

To further evaluate the effectiveness of the proposed approach, three-dimensional building models were reconstructed using the estimated building heights and the corresponding footprint polygons. As shown in Figure 6, the reconstructed models exhibit realistic geometry and proportional height relations across different urban functional zones.

Three typical subregions were selected for detailed visualization: a residential area with uniform low-rise buildings (Figure 6a), an and a commercial cluster with several high-rise buildings (Figure 6c). In each case, the reconstructed building morphology aligns closely with the true urban form observed in the satellite images, validating the geometric accuracy of our method. The contour-based modelling approach also offers superior interpretability compared to raster-based DSM generation, as the derived 3D models maintain explicit topological boundaries and preserve semantic relationships, making them readily integrable into GIS and 3D urban databases (Chen et al., 2021; Javanroodi et al., 2023; Wang & Sertel, 2023).

4.3 Accuracy Assessment and Error Analysis

Quantitative validation was performed by comparing the estimated building roof elevations with manually measured reference values for 3,346 buildings. The validation yielded a Root Mean Square Error (RMSE) of 0.84 m and a Pearson correlation coefficient (r) of 0.9993, indicating an excellent agreement between the estimated and reference heights.

The error distribution, visualized in the histogram in Figure 7, closely follows a normal distribution centered near zero. Over 95% of the samples exhibit absolute errors below 1.5 m, demonstrating the high precision and reliability of our method. The primary sources of error can be attributed to two main factors: (1) minor imperfections in the initial contour extraction due to occlusions or shadow effects, and (2) local inaccuracies within the 5 m resolution DEM, particularly in densely built-up areas.

(a) (b) (c)

Figure 6. 3D visualization of typical urban subregions. (a) Residential zone. (b) Industrial zone. (c) Commercial zone.

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-XI-2-2026-251-2026 | © Author(s) 2026. CC BY 4.0 License.

Nevertheless, the geometric consistency enforced by the vector registration framework significantly reduces the random matching noise commonly encountered in pixel-based stereo methods. Comparatively, our achieved RMSE of 0.84 m is highly competitive with state-of-the-art deep learning-based

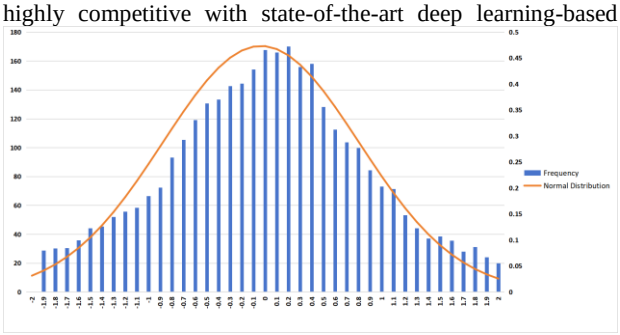

Figure 7. Histogram of elevation errors. The distribution of differences between the estimated and manually measured roof elevations shows a normal-like distribution centered at zero, with the orange line representing the fitted normal distribution curve.

disparity estimation networks, while offering superior geometric integrity at building boundaries. The proposed method thus provides a favorable balance between computational efficiency and geometric accuracy, making it well-suited for operational large-scale building height mapping and 3D urban modeling. While a quantitative benchmarking against dense stereo matching or recent deep learning-based height estimation baselines (e.g., FusionHeightNet) is not presented here, this study successfully demonstrates the geometric robustness of the vector-registration framework. Comprehensive baseline comparisons will be investigated in future work.

4.4 Parameter Sensitivity and Error Sources

To further evaluate the robustness of the proposed method, a sensitivity analysis was conducted with respect to the elevation step (ΔZ) and the DEM resolution. Reducing ΔZ from 1 m to 0.5 m slightly improved the RMSE by 0.06 m, at the cost of approximately 40% additional computation time, indicating a trade-off between accuracy and efficiency. Moreover, replacing the 5 m DEM with a 10 m version increased the RMSE by 0.18 m, confirming that ground elevation precision significantly influences final building height accuracy. Visual inspection revealed that residual errors mainly occur at building edges affected by occlusions or shadow artifacts. These results highlight that the proposed vector-based matching remains robust under moderate parameter variations.

5. Conclusion

This study presented a novel contour vector registration–based method for building height estimation from stereo satellite imagery, integrating the geometric rigor of the VLL with the flexibility of the RFM. By performing elevation optimization through contour-level overlap analysis rather than dense pixel correlation, the proposed framework achieves robust, accurate, and interpretable height retrieval across complex urban areas. Experiments using SV-1 stereo imagery over Jiuyuan District, Baotou, demonstrated sub-meter accuracy with an RMSE of 0.84 m and a correlation coefficient of 0.9993 when compared with reference measurements. The results confirm that the contour-driven approach effectively mitigates edge mismatches and enhances geometric consistency, offering an efficient alternative for large-scale urban 3D modeling. Future research will focus on improving contour extraction using advanced deep learning techniques, extending the VLL optimization to more complex terrains, and integrating multisource data such as LiDAR and SAR to further enhance precision and adaptability in operational urban mapping. The approach also has the potential to support digital twin city applications and automated large-scale 3D urban modeling.

Acknowledgments

The authors would like to thank China Siwei Surveying and Mapping Technology Co. Ltd. for providing the high-resolution stereo images from SuperView-1 satellite, which has provided invaluable support for the experimental research in this paper.

References

Alidoost, F., Arefi, H., Tombari, F., 2019. Building outline extraction from aerial images using convolutional neural networks. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XLII-4-W18*, 57–61. https://doi.org/10.5194/isprs-archives-XLII-4-W18-57-2019

Chen, Q., Wang, L., Waslander, S.L., Liu, X., 2020. An end-to-end shape modeling framework for vectorized building outline generation from aerial images. *ISPRS Journal of Photogrammetry and Remote Sensing* 170, 114–126.

Chen, S., Mou, L., Li, Q., Sun, Y., Zhu, X.X., 2021. Mask-Height R-CNN: An End-to-End Network for 3D Building Reconstruction from Monocular Remote Sensing Imagery, in: *2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS*. Presented at the 2021 *IEEE International Geoscience and Remote Sensing Symposium IGARSS*, pp. 1202–1205.

de Franchis, C., Meinhardt-Llopis, E., Michel, J., Morel, J.-M., Facciolo, G., 2014. An automatic and modular stereo pipeline for pushbroom images, in: *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*. Zürich, Switzerland.

Du, H., Lin, X., Jiang, J., Lu, Y., Du, H., Zhang, F., Yu, F., Feng, T., Wu, X., Peng, G., Deng, S., He, S., Bai, X., 2024a. A single-building damage detection model based on multi-feature fusion: A case study in Yangbi. *iScience* 27.

Du, S., Liu, H., Xing, J., Du, S., 2024b. Fusing multimodal data of nature-economy-society for large-scale urban building height estimation. *International Journal of Applied Earth Observation and Geoinformation* 129, 103809.

Frantz, D., Schug, F., Okujeni, A., Navacchi, C., Wagner, W., van der Linden, S., Hostert, P., 2021. National-scale mapping of building height using Sentinel-1 and Sentinel-2 time series. *Remote Sensing of Environment* 252, 112128.

Fraser, C.S., Hanley, H.B., 2005. Bias-compensated RPCs for Sensor Orientation of High-resolution Satellite Imagery. *Photogrammetric Engineering & Remote Sensing* 71, 909–915.

Huang, Z., Cheng, G., Wang, H., Li, H., Shi, L., Pan, C., 2016. Building extraction from multi-source remote sensing images via deep deconvolution neural networks. Presented at the 2016 *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, pp. 1835–1838.

Javanroodi, K., Perera, A.T.D., Hong, T., Nik, V.M., 2023. Designing climate resilient energy systems in complex urban areas considering urban morphology: A technical review. *Advances in Applied Energy* 12, 100155.

Lee, G.Y., Dam, T., Ferdaus, Md.M., Poenar, D.P., Duong, V.N., 2024. Unlocking the capabilities of explainable few-shot learning in remote sensing. *Artif Intell Rev* 57, 169. https://doi.org/10.1007/s10462-024-10803-5

Lee, P.Q., Xu, L., Clausi, D.A., 2020. Sentinel-1 additive noise removal from cross-polarization extra-wide TOPSAR with dynamic least-squares. *Remote Sensing of Environment* 248, 111982.

Li, C., Guan, T., Yang, M., Zhang, C., 2021. Combining data-and-model-driven 3D modelling (CDMD3DM) for small indoor scenes using RGB-D data. *ISPRS Journal of Photogrammetry and Remote Sensing* 180, 1–13.

Li, Q., Mou, L., Hua, Y., Shi, Y., Chen, S., Sun, Y., Zhu, X.X., 2023. 3DCentripetalNet: Building height retrieval from monocular remote sensing imagery. *International Journal of Applied Earth Observation and Geoinformation* 120, 103311.

Li, Z., Weng, Q., Zhou, Y., Dou, P., Ding, X., 2024. Learning spectral-indices-fused deep models for time-series land use and land cover mapping in cloud-prone areas: The case of Pearl River Delta. *Remote Sensing of Environment* 308, 114190.

Luo, W., Schwing, A.G., Urtasun, R., 2016. Efficient Deep Learning for Stereo Matching. Presented at the Proceedings of the *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 5695–5703.

Ma, C., Zhang, Y., Guo, J., Zhou, G., Geng, X., 2024. FusionHeightNet: A Multi-Level Cross-Fusion Method from Multi-Source Remote Sensing Images for Urban Building Height Estimation. *Remote Sensing* 16, 958. https://doi.org/10.3390/rs16060958

Memar, B., Russo, L., Ullo, S.L., Gamba, P., 2025. An Object-Based Deep Learning Approach for Building Height Estimation from Single SAR Images. *Remote Sensing* 17, 2922. https://doi.org/10.3390/rs17172922

Miah, M.H., Jin, S., Jubaid, M.U., Hossin, M.A., 2025. High precision building detection from satellite imagery with a novel SDBN-HCWO method. *Remote Sensing Applications: Society and Environment* 39, 101698.

Prabhakar, D., Garg, P.K., 2023. Building edge detection from very high-resolution remote sensing imagery using deep learning. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XLVIII-M-3–2023*, 189–196. https://doi.org/10.5194/isprs-archives-XLVIII-M-3-2023-189-2023

Qi, F., Zhai, J.Z., Dang, G., 2016. Building height estimation using Google Earth. *Energy and Buildings* 118, 123–132. https://doi.org/10.1016/j.enbuild.2016.02.044

Qin, X., He, S., Yang, X., Dehghan, M., Qin, Q., Martin, J., 2018. Accurate Outline Extraction of Individual Building From Very High-Resolution Optical Images. *IEEE Geoscience and Remote Sensing Letters* 15, 1775–1779.

Rau, M.I., Julzarika, A., Yoshikawa, N., Nagano, T., Kimura, M., Setiawan, B.I., Ha, L.T., 2024. Application of topographic elevation data generated by remote sensing approaches to flood inundation analysis model. *Paddy Water Environ* 22, 285–299. https://doi.org/10.1007/s10333-023-00967-1

Sawa, K., Yalcin, I., Kocaman, S., 2024. Building Detection from SkySat Images with Transfer Learning: a Case Study over Ankara. *PFG* 92, 163–175. https://doi.org/10.1007/s41064-024-00279-x

Shahin, A.I., Almotairi, S., 2021. DCRN: An Optimized Deep Convolutional Regression Network for Building Orientation Angle Estimation in High-Resolution Satellite Images. *Electronics* 10, 2970.

Shi, W., Meng, Q., Zhang, L., Zhao, M., Su, C., Guo, G., Li, T., Atkinson, P.M., 2025. Instance-Level Multitask Learning for 3-D Building Extraction From Monocular Off-Nadir Satellite Sensor Imagery. *IEEE Transactions on Geoscience and Remote Sensing* 63, 1–24. https://doi.org/10.1109/TGRS.2025.3591180

Song, X., Yang, G., Zhu, X., Zhou, H., Wang, Z., Shi, J., 2021. AdaStereo: A Simple and Efficient Approach for Adaptive Stereo Matching. Presented at the Proceedings of the *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 10328–10337.

Tang, X., Gao, X., Cao, H., Mo, F., Wang, Z., Xu, W., Zhu, G., Yue, Q., Hu, F., Zhu, H., Lu, J., 2020. The China ZY3-03 Mission: Surveying and Mapping Technology for High-Resolution Remote Sensing Satellites. *IEEE Geoscience and Remote Sensing Magazine* 8, 8–17.

Wang, P., Sertel, E., 2023. Multi-frame super-resolution of remote sensing images using attention-based GAN models. *Knowledge-Based Systems* 266, 110387.

Weng, Q., Li, Z., Cao, Y., Lu, X., Gamba, P., Zhu, X., Xu, Y., Zhang, F., Qin, R., Yang, M.Y., Ma, P., Huang, W., Yin, T., Zheng, Q., Zhou, Y., Asner, G., 2024. How will ai transform urban observing, sensing, imaging, and mapping? *npj Urban Sustain* 4, 50. https://doi.org/10.1038/s42949-024-00188-3

Wu, F., Gong, L., Zhang, B., Wang, C., Zhang, H., Wang, Y., 2025. A Review of Urban Building Extraction From Synthetic Aperture Radar Imagery Based on Deep Learning. *IEEE Access* 13, 137896–137918.

Wu, W.-B., Ma, J., Banzhaf, E., Meadows, M.E., Yu, Z.-W., Guo, F.-X., Sengupta, D., Cai, X.-X., Zhao, B., 2023. A first Chinese building height estimate at 10 m resolution (CNBH-10 m) using multi-source earth observations and machine learning. *Remote Sensing of Environment* 291, 113578.

Xie, Y., Feng, D., Xiong, S., Zhu, J., Liu, Y., 2021. Multi-Scene Building Height Estimation Method Based on Shadow in High Resolution Imagery. *Remote Sensing* 13, 2862. https://doi.org/10.3390/rs13152862

Xu, W., Feng, Z., Wan, Q., Xie, Y., Feng, D., Zhu, J., Liu, Y., 2024. Building Height Extraction From High-Resolution Single-View Remote Sensing Images Using Shadow and Side Information. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 17, 6514–6528.

Zhang, L., Gruen, A., 2006. Multi-image matching for DSM generation from IKONOS imagery. *ISPRS Journal of Photogrammetry and Remote Sensing, Extraction of Topographic Information from High-Resolution Satellite Imagery* 60, 195–211.

Zhao, H., Xu, G., Shi, Y., Li, J., Zhang, Y., 2022. The characteristics of dynamic and non-uniform thermal radiation experienced by pedestrians in a street canyon. *Building and Environment* 222, 109361.

Zhao, Q., Zhou, L., Lv, G., 2023. A 3D modeling method for buildings based on LiDAR point cloud and DLG. *Computers, Environment and Urban Systems* 102, 101974.