

High-precision 3D Urban Reconstruction from Satellite Imagery Using Adaptive Dense Matching

Lingyan Kong,¹ Qingfeng Zeng,¹ Haiying Ren,¹ Dan Geng,² and Ruijia Deng^{3,4,5*}

¹Beijing Institute of Surveying and Mapping, Beijing 100038, China

²Comprehensive Investigation and Design Institute of Construction Co., Ltd., Beijing 100037, China

³Key Laboratory of Digital Earth Science, Aerospace Information Research Institute,
Chinese Academy of Sciences, Beijing 100094, China

⁴International Research Center of Big Data for Sustainable Development Goals, Beijing 100094, China

⁵University of Chinese Academy of Sciences, Beijing 100094, China

(Received March 31, 2026; accepted June 18, 2026)

Keywords: stereo surface reconstruction, dense matching, satellite imagery, adaptive minimum disparity, block-wise refinement, semi-global matching (SGM)

The increasing availability of high-resolution satellite stereo imagery has enabled large-scale 3D reconstruction for urban morphology analysis and monitoring change. However, conventional semi-global block matching methods are highly sensitive to the minimum disparity parameter, particularly in complex urban environments characterized by dense buildings, severe occlusions, and abrupt elevation variations, often leading to mismatches and extensive holes that degrade the quality of digital surface models and point clouds. To address this limitation, we propose a hole-driven hierarchical framework for adaptive minimum disparity estimation. The method first derives a globally adaptive minimum disparity to generate an initial disparity map, followed by the identification of invalid regions and progressive refinement through a coarse-to-fine strategy guided by local statistical characteristics. This transition from global constraint to local adaptation effectively reduces matching failures and improves disparity completeness while preserving geometric consistency. Experiments on multiple satellite stereo datasets, validated against airborne light detection and ranging point clouds, demonstrate significant improvements in the geometric completeness, structural consistency, and detail preservation of reconstructed urban surfaces. The proposed approach provides a robust and scalable solution for accurate 3D urban morphology reconstruction and supports the reliable analysis of fine-scale urban changes.

1. Introduction

With the rapid development of high-resolution satellite Earth observation technologies, 3D reconstruction from stereo imagery has become a fundamental tool for characterizing urban morphology and monitoring its dynamic evolution. Accurate 3D representations of urban structures are essential for a wide range of applications, including urban 3D modeling, disaster emergency response, 5G communication infrastructure planning, and large-scale infrastructure

*Corresponding author: e-mail: dengruijia24@mailsucas.ac.cn
<https://doi.org/10.18494/SAM6362>

monitoring,^(1–3) where detailed geometric information such as building height and spatial configuration plays a critical role. In this context, dense stereo matching serves as a key step in bridging 2D imagery and 3D urban structure reconstruction. Its objective is to estimate pixel-wise disparities for stereo image pairs, from which depth information can be recovered. However, satellite imagery typically features wide spatial coverage, pronounced terrain variations, radiometric inconsistencies, occlusions, and repetitive textures, all of which pose significant challenges to achieving accurate and efficient dense matching.⁽⁴⁾ These challenges become even more pronounced in complex urban environments, where dense building structures, severe occlusions, and abrupt elevation changes further degrade matching reliability.

Existing dense stereo matching approaches can generally be categorized into four groups in accordance with their optimization strategies: local cost aggregation methods, global optimization methods, semi-global matching (SGM), and deep-learning-based methods.⁽³⁾ Local methods, such as window-based matching using the sum of absolute differences or normalized cross-correlation measures, rely on photometric consistency within local neighborhoods. While computationally efficient, they are prone to errors in weakly textured regions, occlusions, repetitive patterns, and disparity discontinuities.⁽⁵⁾ Global optimization approaches, including graph cuts and belief propagation, formulate disparity estimation as an energy minimization problem with global smoothness constraints, thereby improving consistency across the image.^(6,7) Nevertheless, their iterative optimization procedures incur substantial computational costs, limiting their practicality for processing ultralarge satellite imagery containing hundreds of millions of pixels.^(8,9)

In recent years, deep-learning-based stereo matching has achieved remarkable progress. Networks such as a matching-cost convolutional neural network (MC-CNN)⁽¹⁰⁾ leveraged convolutional neural networks to learn robust matching costs. Subsequent architectures, including GC-Net⁽¹¹⁾ and PSMNet,⁽¹²⁾ construct cost volumes and perform 3D convolutional regularization, achieving state-of-the-art performance on benchmark datasets such as KITTI and Middlebury. Architectures such as GC-Net are reported to reduce mismatch rates in low-texture regions by 30–50%,⁽¹³⁾ significantly improving disparity completeness. Satellite-oriented adaptations of networks such as PSMNet and HSM have also been examined in comprehensive reviews.⁽¹⁴⁾ Despite their high accuracy, deep learning approaches face notable limitations in satellite remote sensing scenarios. First, the scarcity of large-scale, high-quality ground-truth disparity datasets constrains supervised training.⁽¹⁵⁾ Second, generalization across heterogeneous satellite scenes remains challenging, as models trained on specific terrains often fail to transfer reliably to regions with different geomorphological characteristics.^(16,17) Finally, the computational demands of deep networks require substantial GPU resources, complicating deployment on ultra-high-resolution satellite imagery. Consequently, traditional model-based methods continue to dominate many operational workflows where training data or computational resources are limited.⁽¹⁷⁾

SGM approximates global optimization by aggregating matching costs along multiple paths, achieving an effective balance between accuracy and computational efficiency.⁽¹⁸⁾ The method demonstrates strong performance in preserving object boundaries and fine structural details, while exhibiting robustness to radiometric differences and relatively low sensitivity to parameter

selection.⁽¹⁸⁾ In recent studies, it has been emphasized that SGM remains one of the most widely adopted stereo matching algorithms in photogrammetry and computer vision owing to its high parallelizability, real-time potential, and independence from scene-specific training data. Empirical investigations on satellite imagery further confirm its effectiveness; comparative studies have revealed that SGM-derived digital surface models (DSMs) outperform traditional window-based approaches in both completeness and geometric accuracy.^(19–23)

Despite these advantages, the semi-global block matching (SGBM) implementation commonly used in engineering frameworks introduces additional disparity range parameters, such as the minimum disparity (minDisparity), to constrain search interval and improve computational stability.^(18,24) Satellite stereo pairs often involve large baselines, significant terrain variation, and wide disparity ranges. Under such conditions, a globally fixed minimum disparity may fail to cover all terrain regions, particularly areas with large elevation differences. In previous work, it was noted that complex urban structures and low-texture surfaces can lead to unsatisfactory disparity estimation, resulting in mismatches and invalid regions.⁽⁴⁾ In practice, the improper configuration of the minDisparity parameter can generate extensive disparity holes or erroneous matches, degrading the geometric integrity of DSMs and reconstructed point clouds.

To address these challenges, in this study, we propose a block-wise adaptive minimum disparity framework driven by the hierarchical refinement of disparity holes. At the global scale, the statistical characteristics of disparity distributions are used to construct evaluation criteria, enabling the iterative estimation of an optimal minimum disparity parameter and the generation of a stable initial disparity map. Regions containing disparity holes are then identified and processed using a coarse-to-fine hierarchical partitioning strategy. Each subdivided region undergoes localized adaptive matching recalculation, allowing disparity parameters to be dynamically adjusted using local statistics. The primary contributions of this work are summarized as follows:

- an adaptive minimum disparity optimization mechanism that integrates global and local disparity statistics to improve robustness while reducing parameter sensitivity;
- a hierarchical block partition strategy that progressively refines matching in failure regions through localized disparity estimation;
- a progressive framework transitioning from global constraints to local optimization, effectively filling disparity holes and suppressing error propagation.

The proposed method is evaluated using multiple high-resolution satellite stereo datasets, and the reconstructed point clouds are quantitatively assessed against airborne light detection and ranging (LiDAR) reference data.

2. Motivation and Algorithm Framework

The rapid advancement of high-resolution satellite Earth observation technology has enabled the generation of dense point clouds from stereo imagery using accurate RPC models. As a cost-effective and efficient alternative to traditional airborne surveys, this approach supports the large-scale derivation of DSMs and digital elevation models, which serve as fundamental data

sources for 3D urban morphology analysis. These 3D representations provide essential geometric descriptors for characterizing urban structures—such as building height, spatial configuration, and surface complexity—which are critical for understanding and monitoring urban evolution. Within the complete 3D reconstruction workflow, dense stereo matching serves as the critical component that bridges 2D imagery and 3D geometric reconstruction.

Mathematically, the dense matching problem can be formulated as the search for pixel correspondences between two or more images under a predefined similarity criterion. With $I_L(x, y)$ and $I_R(x, y)$, where (x, y) represents the image coordinates, the objective is to determine a disparity function $d(x, y)$ such that, for each pixel (x_L, y_L) in I_L , the corresponding pixel in I_R is located at $(x_L - d(x_L, y_L), y_L)$.

However, several practical factors complicate this process: (1) illumination variations and image noise, (2) shadows and specular reflections causing radiometric inconsistencies, (3) perspective distortions, (4) repetitive or textureless regions, and (5) occlusions and depth discontinuities. In this study, the SGBM algorithm is experimentally evaluated for the stereo matching of satellite epipolar imagery and compared with the SGM⁽¹⁸⁾ and PatchMatch⁽²⁵⁾ algorithms. On the basis of the results of comprehensive experimental analysis, SGBM is selected as the dense matching framework because of its favorable balance between accuracy and computational efficiency. To address the limited adaptability of conventional fixed-parameter matching strategies in complex terrain scenarios, an adaptive minimum disparity estimation method is proposed. Furthermore, the generated disparity map is partitioned into multi-scale blocks, and regions with high void ratios are reprocessed using a locally reconstructed search range. This strategy significantly improves the completeness and accuracy of the final disparity results.

These factors may lead to incorrect correspondences or matching failures. Therefore, effective solutions must be developed from multiple perspectives, including image preprocessing, algorithmic optimization, and disparity map postprocessing.

As shown in Fig. 1, the reconstruction workflow consists of four main stages.

1. **Preprocessing:** Epipolar images are preprocessed to reduce the effects of shadows, weak texture, and noise, thereby improving dense matching reliability.
2. **Dense Matching:** The SGBM algorithm is employed for dense stereo matching. To address regional variability, an adaptive minimum disparity parameter is introduced.
3. **Void Region Refinement:** Block-wise statistics of invalid disparity ratios and image entropy are used to identify and locally recompute unreliable regions.
4. **Point Cloud Generation:** Conjugate points are projected using RPC parameters and the mapping matrix. Forward intersection is applied to iteratively estimate elevation and generate the final 3D point cloud.

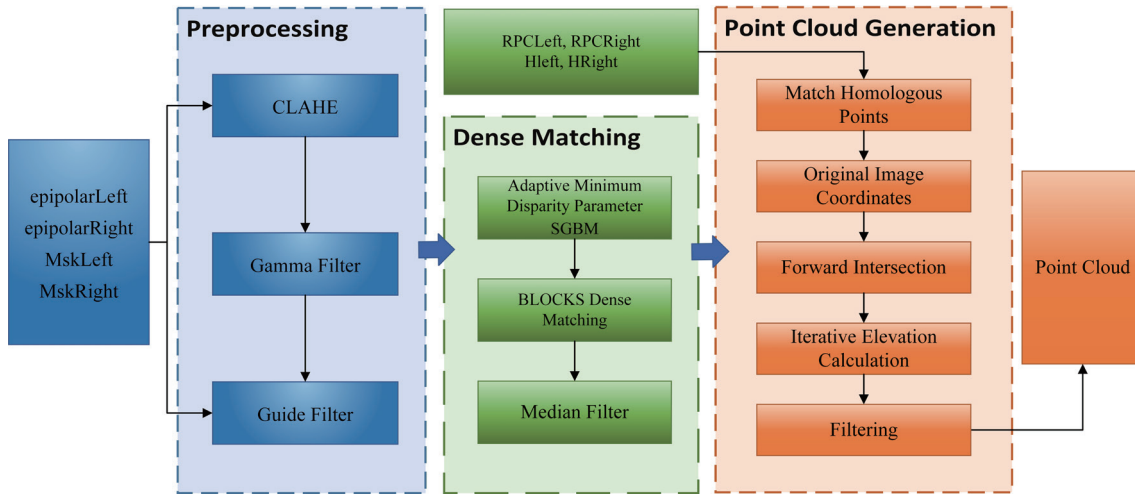


Fig. 1. (Color online) Algorithm framework.

3. Dense Matching and 3D Reconstruction

3.1 Preprocessing of satellite epipolar imagery

Compared with traditional global histogram equalization, contrast-limited adaptive histogram equalization (CLAHE)⁽²⁶⁾ enhances local contrast while avoiding overexposure or underexposure caused by global intensity adjustments. In this study, CLAHE was applied to each epipolar image with a tile grid size of 4×4 . The clip limit was adaptively determined from the standard deviation of valid nonzero pixels as follows.

$$clipLimit = \max \left(3.0, \min \left(10.0, \frac{\sigma_I}{10} \right) \right) \quad (1)$$

Here, σ_I denotes the gray-level standard deviation of valid pixels. This setting constrains the clip limit within the range of 3.0–10.0 to balance local contrast enhancement and noise suppression.

Gamma correction was then applied to enhance details in darker regions. The gamma coefficient was set to 0.7, and the transformation was defined as

$$I_\gamma = 255 \left(\frac{I_{CLAHE}}{255} \right)^{0.7}. \quad (2)$$

To further suppress local noise while preserving edge structures critical for accurate correspondence estimation, guided filtering⁽²⁷⁾ was employed. The original epipolar image was used as the guidance image, while the gamma-corrected image was used as the filtering input. The guided filtering radius was set to 1, the regularization parameter ϵ was set to 5.1, and the

output depth was kept unchanged. Finally, radiometric normalization was performed over valid nonzero regions using the mean and standard deviation of the left and right epipolar images to reduce brightness and contrast differences between the stereo pair. Overall, the preprocessing step improves image quality and provides more stable inputs for subsequent dense matching algorithms.

3.2 Dense matching

3.2.1 Adaptive minimum disparity optimization for SGBM

To efficiently determine the optimal minimum disparity, an adaptive iterative optimization strategy is proposed for the SGBM algorithm. As shown in Fig. 2, the workflow of the proposed adaptive minimum disparity optimization strategy consists of the iterative evaluation and dynamic adjustment of the disparity search range. By integrating multi-criteria evaluation, the method efficiently determines the optimal minimum disparity while maintaining matching stability.

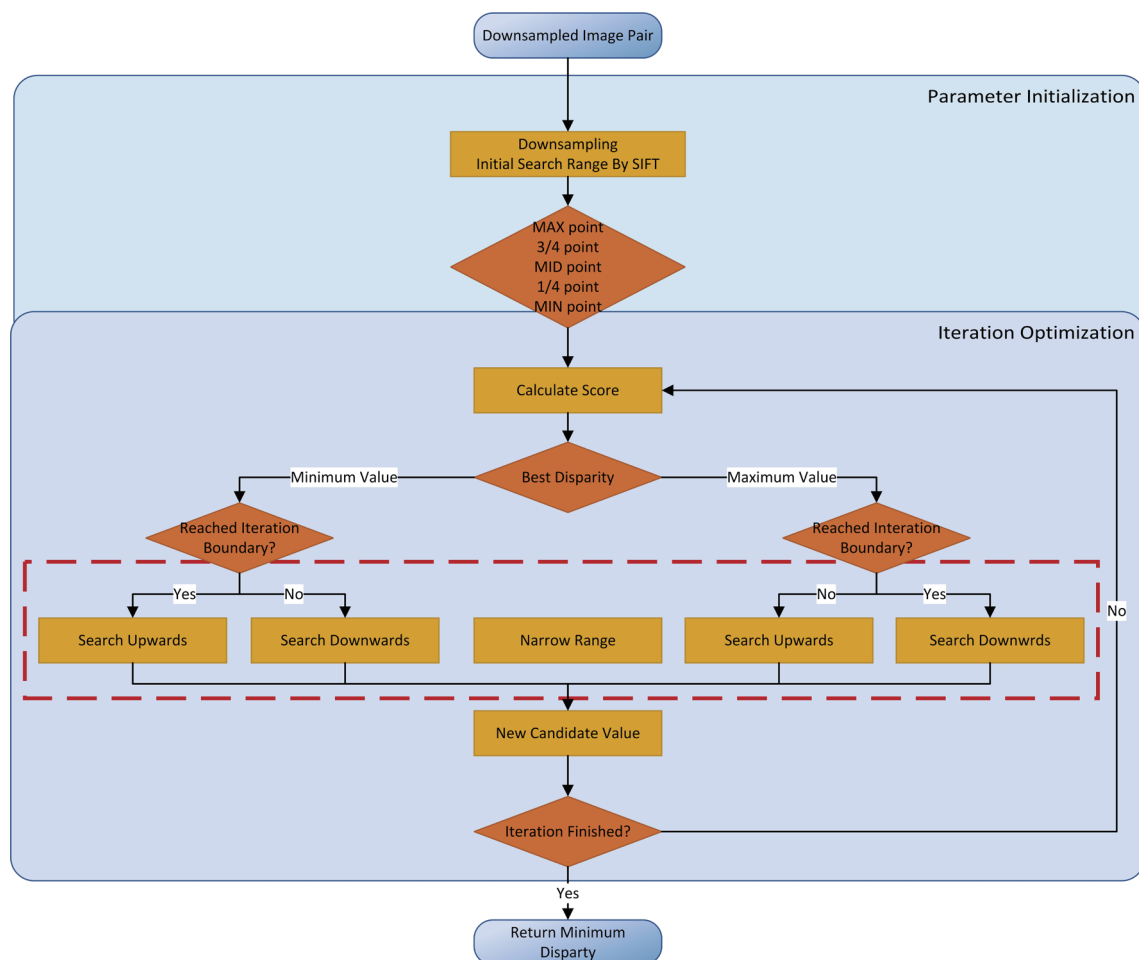


Fig. 2. (Color online) Algorithm pipeline.

3.2.1.1 Preprocessing and initialization

The stereo epipolar images are first downsampled to reduce computational cost. The initial search interval of the minimum disparity is determined from feature correspondences extracted using SIFT⁽²⁸⁾ under epipolar constraints (see Sect. 3.2.2). Within the initial interval, five candidate points are sampled, including the minimum value (*MIN*), the first quartile point, the middle point (*MID*), the third quartile point, and the maximum value (*MAX*). Preliminary disparity maps and evaluation scores are computed for each candidate.

3.2.1.2 Construction of evaluation metrics

The primary evaluation metric is defined as

$$score = a \cdot invalid_ratio + b \cdot disparity_variance, \quad (3)$$

where a and b are weighting coefficients. In this study, both coefficients were set to 0.5. The term invalid ratio reflects the proportion of unmatched or invalid pixels in the disparity map, while disparity variance describes the stability of the estimated disparity distribution. Equal weights were adopted to balance disparity completeness and disparity stability, avoiding excessive preference for either hole reduction or variance suppression.

During each iteration, beyond the computation of the primary metric, a secondary evaluation metric, $score_e$, is introduced to further assess structural quality and matching consistency.

$$score_e = c \cdot RMSE_{log} + d \cdot MAE_{log} - e \cdot SSIM \quad (4)$$

Here, $RMSE_{log}$ and MAE_{log} indicate reprojection errors under logarithmic scaling, while structural similarity index measure (SSIM) represents structural similarity. The weights c , d , and e are set to 0.4, 0.3, and 0.3, respectively. A slightly larger weight was assigned to $RMSE_{log}$ because it is more sensitive to large reprojection errors, whereas MAE_{log} and SSIM provide complementary constraints on average error magnitude and structural consistency.

3.2.1.3 Dynamic range adjustment

During each iteration, the candidate with the minimum primary $score$ is selected as the current local optimal disparity value, and its corresponding secondary metric ($score_e$) is recorded.

According to the relative position of the optimal value within the current search interval, the range is dynamically adjusted as follows.

Internal case: If the optimum lies inside the interval, the range is reduced using a bisection strategy, generating a new candidate set.

$$C^{(k+1)} = \left\{ \left\lfloor \frac{d_{min}^{(k)} + d^{*(k)}}{2} \right\rfloor, d^{*(k)}, \left\lfloor \frac{d_{max}^{(k)} + d^{*(k)}}{2} \right\rfloor \right\} \quad (5)$$

Boundary case: If the optimal value coincides with the current iteration boundary, the search step is adaptively adjusted, yielding a new candidate set:

$$C^{(k+1)} = \left\{ \left\lfloor \frac{d_{1/4}^{(k)} + d_{min}^{(k)}}{2} \right\rfloor, d^{*(k)}, \left\lfloor \frac{d_{3/4}^{(k)} + d_{max}^{(k)}}{2} \right\rfloor \right\}. \quad (6)$$

The iteration boundary is then updated in accordance with the current search direction (increasing or decreasing).

Extreme case: If the optimal disparity reaches the predefined extreme of the search range, one-sided exploration is performed along that direction, forming a new candidate set:

$$C^{(k+1)} = \begin{cases} \left\{ \left\lfloor \frac{d_{min}^{(k)} + d^{*(k)}}{2} \right\rfloor, d^{*(k)} \right\}, & \text{lower extreme,} \\ \left\{ d^{*(k)}, \left\lfloor \frac{d_{max}^{(k)} + d^{*(k)}}{2} \right\rfloor \right\}, & \text{upper extreme.} \end{cases} \quad (7)$$

The corresponding search boundary is then updated according to the search direction. In Eqs. (5)–(7), $\lfloor \cdot \rfloor$ denotes the floor operation and k the iteration index.

The candidate set and evaluation metrics are updated at each iteration. The maximum number of iterations was set to 30 to limit computational cost. In addition, an early termination criterion was used in the implementation. The iteration terminates when the updated candidate set contains only one unique minimum disparity value:

$$\left| \text{unique} \left(C^{(k+1)} \right) \right| = 1, \quad (8)$$

where $C^{(k+1)}$ denotes the updated candidate set at iteration $k + 1$. This criterion indicates that the optimization process has converged to a stable minimum disparity candidate. Therefore, the stopping condition is based on the stabilization of candidate disparity values rather than a predefined threshold of score variation.

3.2.1.4 Global optimal selection

After all iterations, all the locally optimal candidates recorded during the search process are ranked in ascending order in accordance with the primary metric score. Let the minimum

primary score be $score_{min}$. To avoid selecting an unstable candidate caused by small fluctuations in primary metric, a tolerance range is introduced. Candidate parameters satisfying

$$score \leq score_{min} + 0.15 \quad (9)$$

are retained as near-optimal solutions. The tolerance value of 0.15 was empirically selected to include candidates with comparable primary scores while excluding evidently inferior results. Finally, within this near-optimal subset, the candidate with the minimum $score_e$ is selected as the final optimized minimum disparity parameter. This two-stage selection first ensures global disparity completeness and stability, and then uses the secondary metric to select the result with better reprojection consistency and structural quality.

3.2.2 Block-wise disparity refinement with entropy constraint

Although global adaptive optimization produces stable disparity maps, it may fail in regions with abrupt terrain variations. To address this issue, a multi-scale block-wise refinement strategy is introduced.

As shown in Fig. 3, the basic principle of the block-wise disparity refinement strategy is illustrated. Image blocks with high disparity hole ratios are identified for re-estimation. Local feature correspondences are extracted using ORB⁽²⁹⁾ and matched under Hamming distance and epipolar constraints. The disparities of matched points are sorted, and a robust estimate is obtained by averaging the smallest and third smallest disparities, followed by averaging with the second smallest value to suppress mismatches. A new local search interval is defined as

$$disparity_interval = [optimal_dis - 100, optimal_dis + 100]. \quad (10)$$

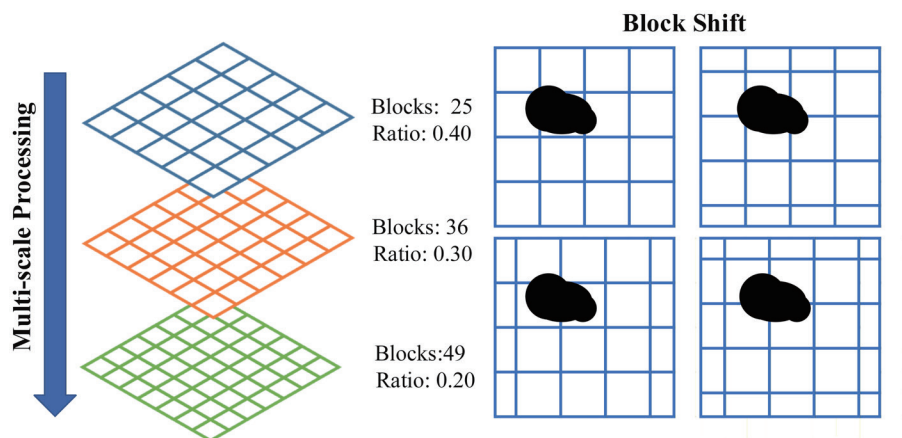


Fig. 3. (Color online) Block-wise disparity refinement.

To prevent unreliable refinement in weak-texture regions (e.g., water surfaces), local image entropy is employed as a masking constraint. Blocks with entropy below a predefined threshold are excluded from re-estimation. This strategy enhances local reconstruction reliability while preserving the global consistency of the disparity map.

3.3 Forward intersection and point cloud generation

After obtaining the dense disparity map, spatial forward intersection is performed using conjugate image points and the satellite RPC model to establish the mapping from image pixel coordinates to 3D ground coordinates.⁽³⁰⁾

The RPC model of satellite imagery establishes a nonlinear relationship between the normalized image coordinates (r_n, c_n) and the normalized geographic coordinates (λ_n, ϕ_n, h_n) :⁽³¹⁾

$$\begin{cases} r_n = \frac{P_1(\lambda_n, \phi_n, h_n)}{P_2(\lambda_n, \phi_n, h_n)}, \\ c_n = \frac{P_3(\lambda_n, \phi_n, h_n)}{P_4(\lambda_n, \phi_n, h_n)}, \end{cases} \quad (11)$$

where $P_i(\cdot)$ represents a third-order polynomial with 20 terms in three variables.

Given a pair of corresponding image points (x_L, y_L) and (x_R, y_R) , the estimation of the 3D ground coordinates is formulated as a nonlinear least-squares problem that minimizes the reprojection error.⁽³²⁾

In this study, a damped Newton iterative scheme, equivalent to the Levenberg–Marquardt method, is employed for optimization:

$$\mathbf{X}_{k+1} = \mathbf{X}_k - (\mathbf{J}^T \mathbf{W} \mathbf{J} + \lambda \mathbf{I})^{-1} \mathbf{J}^T \mathbf{W} \mathbf{e}. \quad (12)$$

By applying the above iterative solution to all valid disparity pixels, the geographic coordinates corresponding to each image pixel are obtained. Combined with the spectral information from the original imagery, a georeferenced dense point cloud is finally generated.

4. Experiments and Results

4.1 Experimental setup

To evaluate the robustness and generalization capability of the proposed dense matching method, experiments were conducted across four representative regions, namely, Beijing (Tanzhe Temple), Shanghai (Jingci area), Hebei (Huailai), and Hangzhou. These study areas cover diverse environments such as mountainous, rural, and complex urban scenes; different environments introduce distinct challenges for dense matching. Mountainous regions are characterized by large elevation variations and wide disparity ranges, which increase the

difficulty of reliable correspondence estimation. In contrast, complex urban environments are dominated by dense building structures, severe occlusions, and sharp structural discontinuities, which lead to significant matching ambiguity and errors.

The acquired stereo datasets are summarized in Table 1. Only panchromatic stereo images were used in this study. Pleiades-1A provides very high resolution panchromatic imagery with an approximate spatial resolution of 0.5 m and supports stereo acquisition, making it suitable for detailed urban 3D reconstruction. GF-7 is equipped with forward- and backward-looking panchromatic cameras, with spatial resolutions of approximately 0.8 and 0.65 m, respectively, enabling large-area satellite stereo mapping. For the Beijing Tanzhe Temple dataset, Jilin-1 GF02B panchromatic imagery was used, with an approximate spatial resolution of 0.75 m. The use of these datasets allows the proposed method to be evaluated under different imaging geometries, spatial resolutions, and terrain conditions.

All experiments were conducted on a Dell Precision 3640 Tower workstation equipped with an Intel Core i9-10900K CPU at 3.70 GHz, 64 GB of RAM, and an NVIDIA Quadro RTX 4000 GPU with 8 GB of memory. The proposed SGBM-based dense matching and adaptive disparity optimization were implemented using multi-threaded CPU processing.

4.2 Evaluation metrics

Since pixel-level ground-truth disparity maps are unavailable in real satellite stereo scenarios, the optimized disparity maps were converted into 3D point clouds through RPC-based forward intersection and then compared with airborne LiDAR reference data. Two types of accuracy assessment were conducted: global spatial consistency and local elevation-profile accuracy.

For global spatial consistency, the cloud-to-cloud (C2C) absolute distance between the reconstructed point cloud and the airborne LiDAR reference point cloud was calculated. The mean value and standard deviation of the C2C distances were used to describe the overall geometric deviation and error dispersion, respectively.

For local structural accuracy, elevation profiles were extracted along representative *X*- and *Y*-direction transects. The reconstructed and reference profiles were first aligned within their common spatial range and averaged in identical spatial bins. The elevation differences of the matched bins were then used to calculate root mean square error (*RMSE*) and mean absolute error (*MAE*) as follows.

Table 1
Summary of study areas, panchromatic stereo datasets, and computational efficiency.

Test sites	Sensors	Image size (columns × rows, pixels)	Area (km ²)	Runtime (s)
Shanghai Jingci	Pleiades-1A	16089 × 15945	64.13477	7902.66
Beijing Tanzhe Temple	Jilin-1 GF02B	8867 × 7165	35.7368	3113.66
Hebei Huailai	GF-7	35818 × 42787	647.5002	37456.17
Hangzhou	Pleiades-1A	13089 × 19646	64.2866	7262.32

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (Z_i - \hat{Z}_i)^2} \quad (13)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |Z_i - \hat{Z}_i|$$

Here, n is the number of matched profile bins, Z_i is the reconstructed elevation, and $\hat{Z}_i$ is the corresponding airborne LiDAR reference elevation. In addition, the percentages of elevation differences within ± 1 and ± 2 m were calculated to further evaluate the distribution of local elevation errors.

4.3 Validation of adaptive minimum disparity strategy

The proposed global adaptive minimum disparity estimation strategy was validated using the Hangzhou stereo dataset. The minimum disparity parameter was iteratively adjusted within a predefined range, and evaluation metrics were computed for each candidate value.

Figure 4 shows a subset of intermediate disparity maps generated during the iterative search process. These disparity maps are arranged in ascending order of the minimum disparity value rather than the actual iteration sequence. Among all tested values, the configuration with a minimum disparity of -29 produced the best evaluation result and was therefore selected as the optimal value.

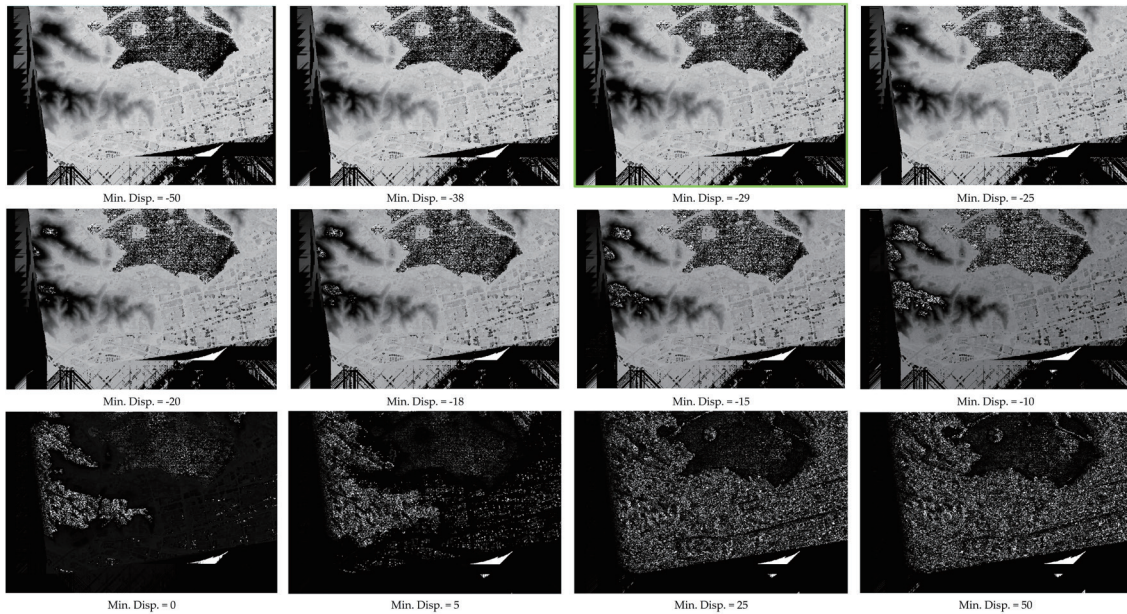


Fig. 4. Partial intermediate results of iterative minimum disparity parameter computation. The optimal minimum disparity is -29 , where disparity maps with disparity near this value exhibit enhanced geometric fidelity and reduced noise. Off-optimal settings generate significant artifacts and erroneous matches.

As shown in Fig. 4, disparity maps obtained with parameter values close to the optimal setting exhibit relatively consistent and continuous structures. In contrast, other parameter configurations lead to the significant degradation of the matching results, manifested as extensive noise artifacts or large disparity holes. This phenomenon indicates that an inappropriate disparity search range can severely affect the reliability of dense stereo matching, since the matching algorithm must search within the specified disparity interval to determine the corresponding pixels.

Figure 5 illustrates the variation of evaluation metrics with respect to the minimum disparity parameter during the iterative optimization process. The results demonstrate a clear convergence behavior as the parameter gradually approaches the optimal value.

In Fig. 5(a), the invalid pixel ratio decreases significantly when the minimum disparity approaches the optimal range. This indicates that the stereo matching algorithm can recover more valid correspondences when the disparity search interval better matches the actual disparity distribution of the scene. When the parameter deviates from the optimal value, the search range becomes inappropriate, leading to an increased number of unmatched pixels and larger disparity holes.

Figure 5(b) presents the variance of the disparity map. A smaller variance indicates a more stable and spatially consistent disparity estimation. When the minimum disparity is close to the optimal value, the variance remains relatively low, suggesting that the estimated disparities are smooth and structurally coherent. In contrast, when the parameter is very large or very small, the disparity estimation becomes unstable, resulting in increased variance owing to noise and mismatched pixels.

The error metrics shown in Figs. 5(c) and 5(d) further confirm this trend. Both $RMSE_{log}$ and MAE_{log} measure the deviation between the estimated disparity and the reference disparity, where smaller values correspond to more accurate disparity estimation. As the minimum disparity approaches the optimal value, these error metrics decrease and reach their minimum levels. This behavior is consistent with common stereo evaluation criteria, where lower $RMSE$ and MAE indicate improved matching accuracy.

Figure 5(e) shows the variation of SSIM, which indicates the structural consistency between disparity maps. SSIM is a perceptual metric designed to measure structural similarity between two images, with higher values indicating better structural preservation. As observed in the figure, SSIM remains relatively high when the minimum disparity is close to the optimal value, indicating that the structural information of the scene is well preserved. When the parameter deviates from the optimal region, SSIM decreases rapidly, reflecting the increasing distortion and structural inconsistency in the disparity maps.

Figure 5(f) illustrates the intermediate evaluation $score_e$, which combines multiple error metrics to provide a comprehensive assessment of disparity quality. The score gradually decreases as the minimum disparity approaches the optimal value, demonstrating that the optimization process effectively guides the parameter toward the best configuration.

Finally, Fig. 5(g) presents the overall evaluation score obtained during the iterative optimization process. The curve clearly shows a convergence trend: the score decreases rapidly

Disparity Metrics Analysis

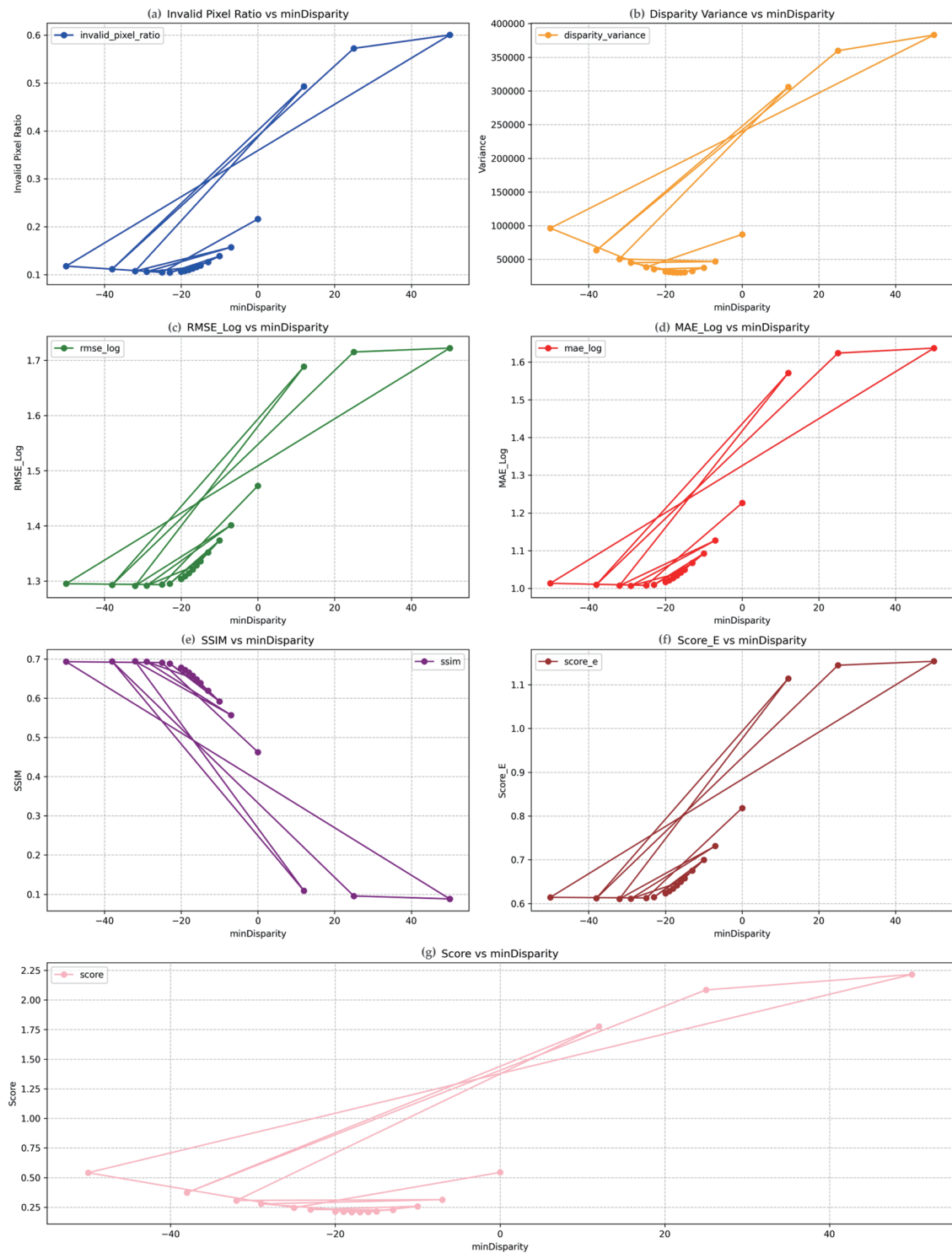


Fig. 5. (Color online) Variation of evaluation metrics with respect to the minimum disparity parameter during adaptive optimization. (a) Invalid pixel ratio; (b) disparity variance; (c) $\text{RMSE}_{\log}$ (labeled $\text{RMSE}_{\log}$ in the plot); (d) $\text{MAE}_{\log}$ (labeled $\text{MAE}_{\log}$); (e) SSIM ; (f) score_e (labeled Score_E); and (g) overall score.

during the early iterations as the search interval approaches the correct disparity range, and then stabilizes near the optimal value.

Results show that when the minimum disparity approaches the optimal value, error metrics reach minimum levels while structural similarity remains high. As the parameter deviates from this value, $RMSE_{log}$ and MAE_{log} increase and SSIM decreases, indicating degraded matching quality. Disparity maps near the optimal parameter exhibit fewer noise artifacts and better structural consistency. These results confirm that the adaptive strategy can reliably determine the optimal search interval and improve dense matching stability. By automatically determining the optimal minimum disparity parameter, the method improves the robustness and reliability of the dense matching results while reducing the need for manual parameter tuning.

4.4 Comparison of hole-filling strategies

To address the issue of frequent disparity holes in mountainous regions and mountain–urban transitional areas, comparative experiments on disparity hole-filling algorithms were designed. The experimental areas include pure mountainous terrain, mountain–urban transition zones, and high-density urban regions. We analyze the recovery capability in areas with high hole rates, boundary continuity, and the suppression of mismatches in weak-texture regions, in order to verify the effectiveness of the proposed multi-scale block-wise optimization strategy in complex scenarios.

In the Tanzhe Temple mountainous area, the proposed method effectively suppresses erroneous matches and reduces voids caused by abrupt elevation changes. As shown in Figs. 6(a) and 6(b), the comparison of red-boxed areas between the original and refined disparity maps clarifies that the method improves matching reliability and significantly decreases hole formation in steep terrain regions.

In Hangzhou's water-body regions, the entropy-based threshold constraint—introduced in Sect 3.2.2—is applied to identify homogeneous or weak-texture areas (e.g., water surfaces), thereby preventing the propagation of incorrect disparities. As illustrated in the red-boxed region of West Lake in Figs. 6(c) and 6(d), the refined disparity map exhibits markedly reduced noise (i.e., spurious disparity values) compared with the unrefined result, confirming the effectiveness of the entropy-guided constraint in preserving structural integrity under low-texture conditions.

As illustrated in Fig. 7, the proposed disparity refinement method significantly improves the quality of disparity maps across six representative regions with various terrain and settlement characteristics. Compared with the original results, the refined disparity maps exhibit substantial reductions in the number of void regions and noise artifacts, leading to enhanced disparity completeness and consistency. In mountainous areas, the refined results show smoother and more continuous terrain surfaces, effectively handling large elevation variations. In hilly regions with settlements, both terrain transitions and building boundaries are more clearly delineated. In dense urban areas, the method effectively suppresses mismatches caused by repetitive textures and shadowed regions, resulting in the improved structural consistency and more reliable representation of urban features.

Overall, the proposed method demonstrates strong robustness under both terrain-driven and structure-driven challenges, achieving improved disparity completeness while preserving

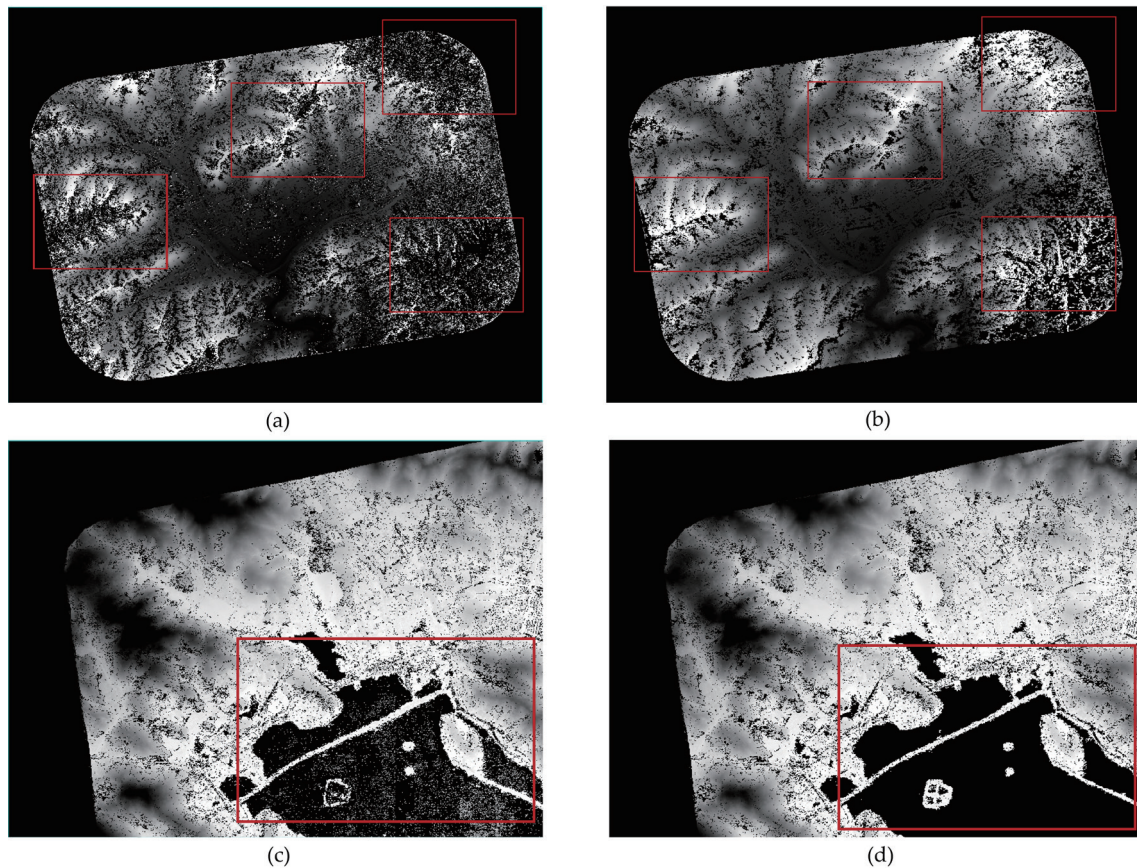


Fig. 6. Comparative analysis results of the entropy-guided hole-filling strategy. (a) Original disparity map of Tanzhe Temple (Beijing). (b) Refined disparity map after hole filling, where the red boxes highlight reduced artifacts and improved geometric continuity. (c) Original disparity map of West Lake (Hangzhou). (d) Processed result demonstrating enhanced structural integrity through entropy-based thresholding. The areas in the red boxes in (b) and (d) visually validate the effectiveness of the proposed adaptive refinement in suppressing voids while preserving topographic fidelity.

boundary continuity. This leads to more accurate and reliable 3D reconstruction results, which are essential for urban morphology analysis.

4.5 Point cloud reconstruction results

Using the optimized disparity maps and RPC model, 3D point clouds were generated for Hangzhou, Shanghai (Jingci), Beijing (Tanzhe Temple), and Hebei (Huailai). Table 1 shows the experimental datasets and computational efficiency. The image size refers to the epipolar panchromatic stereo image dimensions used for dense matching, and the reported runtime corresponds to the complete dense matching and disparity refinement process for each study area.

Figure 8 illustrates the reconstructed point clouds for these representative areas. The results show consistent geometric structures across diverse terrain types. Urban areas show well-

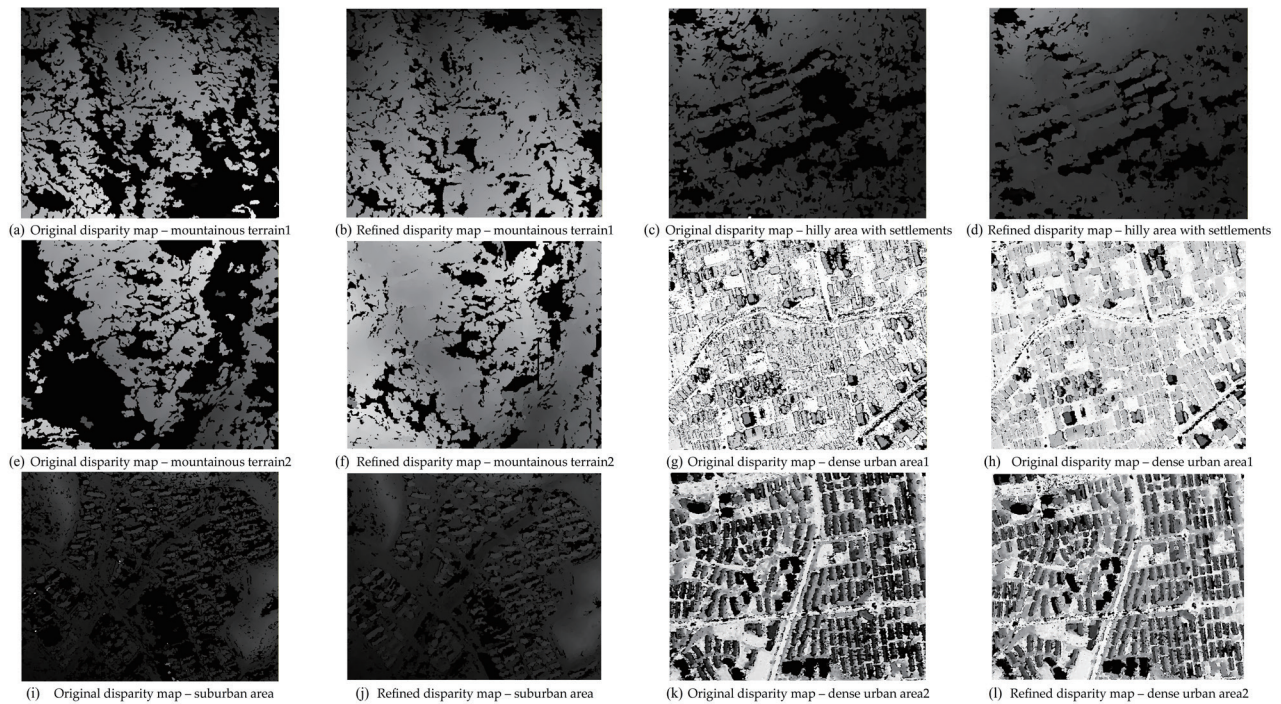


Fig. 7. Comparative analysis results of disparity map refinement across six distinct regions: (a, b) mountainous terrain 1, (c, d) hilly area with settlements 1, (e, f) mountainous terrain 2, (g, h) hilly area with settlements 2, (i, j) dense urban area 1, and (k, l) dense urban area 2. Original and refined disparity maps demonstrate the marked suppression of voids and noises across all regions, with improved geometric continuity in terrain transitions and urban structures. The results validate the method's adaptability to various topographic complexities and settlement densities.

preserved building outlines and structural integrity, whereas mountainous regions maintain a continuous terrain morphology. Although limited noise appears in shadowed or repetitive-texture areas, the overall terrain continuity remains stable.

From a sensor-specific perspective, the datasets also exhibit different reconstruction characteristics. The Pleiades-1A datasets, with higher panchromatic spatial resolution, provide richer structural details for the Shanghai Jingci and Hangzhou areas, which is beneficial for preserving building outlines and urban morphology. The GF-7 dataset covers a much larger Huailai region with agricultural land, woodland, rural settlements, and mountainous terrain; despite the larger image size and more heterogeneous land cover, the reconstructed profiles remain stable. The Jilin-1 dataset is further used to evaluate the proposed method in mountainous terrain with large elevation variations and shadowed areas. These observations indicate that the proposed adaptive disparity optimization strategy is applicable to different high-resolution panchromatic stereo satellite datasets, while the final reconstruction quality remains affected by spatial resolution, stereo imaging geometry, radiometric consistency, terrain complexity, and RPC accuracy. These results demonstrate that the proposed method achieves reliable and adaptable 3D reconstruction performance in complex urban and hilly environments.

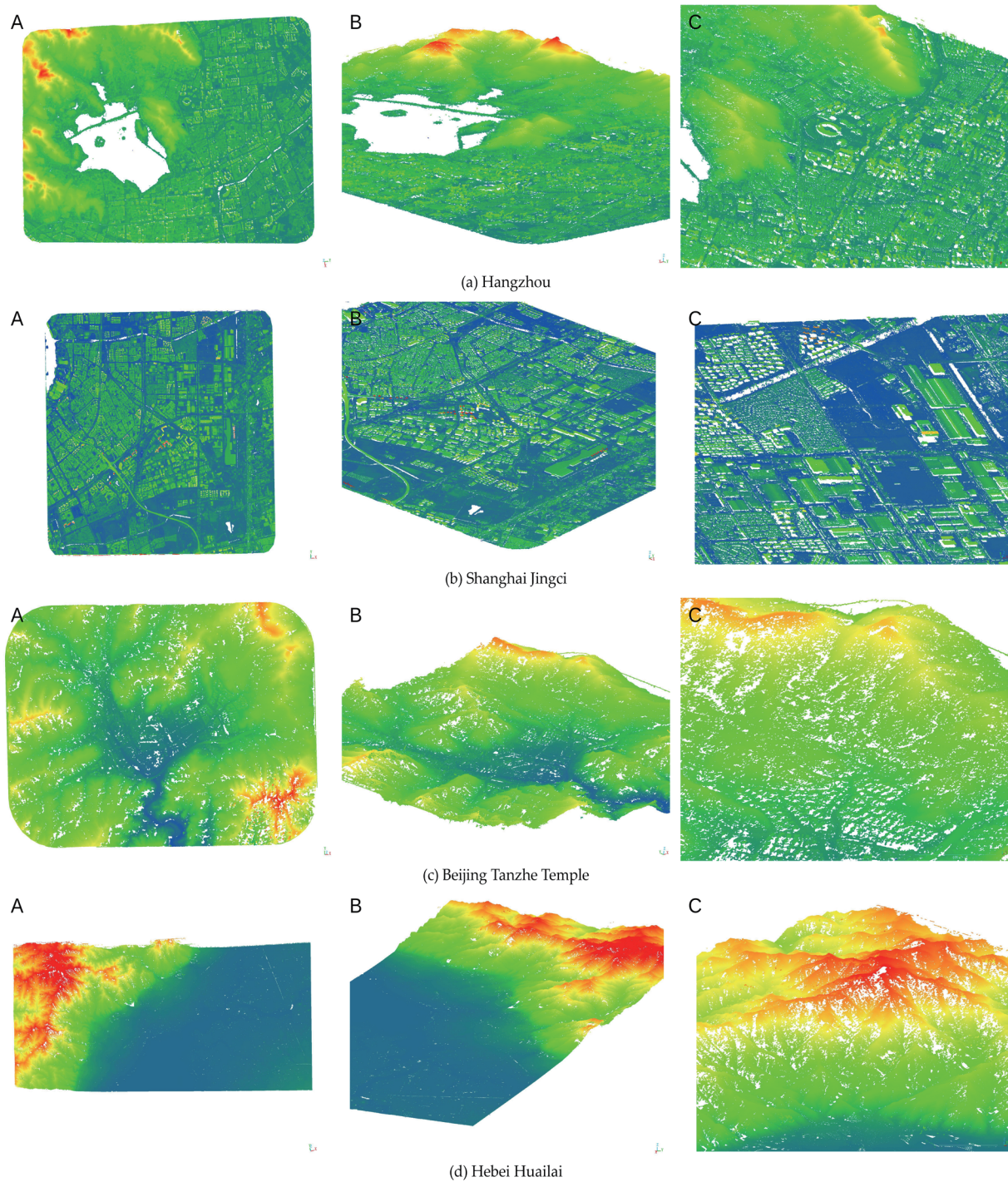


Fig. 8. (Color online) Disparity-based point cloud reconstruction across four geographic regions: (a) Hangzhou, (b) Shanghai Jingci, (c) Beijing Tanzhe Temple, and (d) Hebei Huailai. Each row corresponds to a unique region, while columns 1–3 (subfigure labels A, B, and C) represent multi-viewpoint perspectives derived from stereo disparity maps using RPC models. The consistent preservation of topographic fidelity and geometric integrity across all viewing angles and terrain types validates the robustness of the disparity-to-point-cloud pipeline under diverse acquisition geometries.

5. Discussion

To further evaluate the geometric accuracy and reliability of the reconstructed point clouds, quantitative validation was conducted using airborne LiDAR data as reference. Two representative test sites were selected: the Shanghai Jingci region characterized by dense urban structures and a complex building morphology, and the Hebei Huailai region dominated by agricultural plains, mountainous terrain, and vegetated areas.

The following discussion is focused on two aspects. First, global spatial consistency is analyzed using the C2C absolute distance between the reconstructed point cloud and the airborne LiDAR reference data. Second, local structural accuracy is examined using the elevation profiles extracted along the *X*- and *Y*-direction transects, where *RMSE*, *MAE*, and the percentages of points within ± 1 and ± 2 m are used to quantify elevation differences.

5.1 Global absolute distance error analysis

The statistical results indicate that the reconstructed point clouds maintain good geometric consistency in both study areas.

For the Jingci area shown in Fig. 9, the mean errors in the two experiments are 1.241 and 1.119 m, with standard deviations of 1.944 and 1.762 m, respectively. The error distribution is right-skewed, with the main peak concentrated at around 1 m, indicating that most points exhibit relatively small spatial deviations from the LiDAR reference.

For the Huailai1 and Huailai2 areas shown in Fig. 10, the mean errors are 1.384 and 1.176 m, respectively, with corresponding standard deviations of 0.497 and 0.407 m. These results indicate compact error distributions and stable global spatial consistency across both test areas.

Overall, the average absolute error of the reconstructed point clouds remains stable at approximately 1.2 m across both urban and natural terrain environments, demonstrating consistent geometric accuracy.

5.2 Elevation profile accuracy analysis

To further evaluate local structural reconstruction, representative elevation profiles were extracted along the *X*- and *Y*-directions, and the elevation differences between the reconstructed and reference data were analyzed.

The quantitative accuracy assessment results of the elevation profiles are shown in Table 2, while the corresponding profiles are illustrated in Fig. 11. In the Jingci area, the first experiment yields an average *RMSE* of 5.071 m and an *MAE* of 2.803 m, with 45.3% of the points within ± 1 m and 67.5% within ± 2 m. The second experiment produces an *RMSE* of 5.645 m and an *MAE* of 2.743 m, with 48.8 and 66.9% of points within ± 1 and ± 2 m, respectively. The *MAE* values of both experiments remain close (approximately 2.7–2.8 m), indicating stable elevation accuracy. Profile comparisons show that the reconstructed elevation trends generally agree with the LiDAR data. Local discrepancies are mainly caused by temporal differences between datasets, such as building changes and vegetation growth.

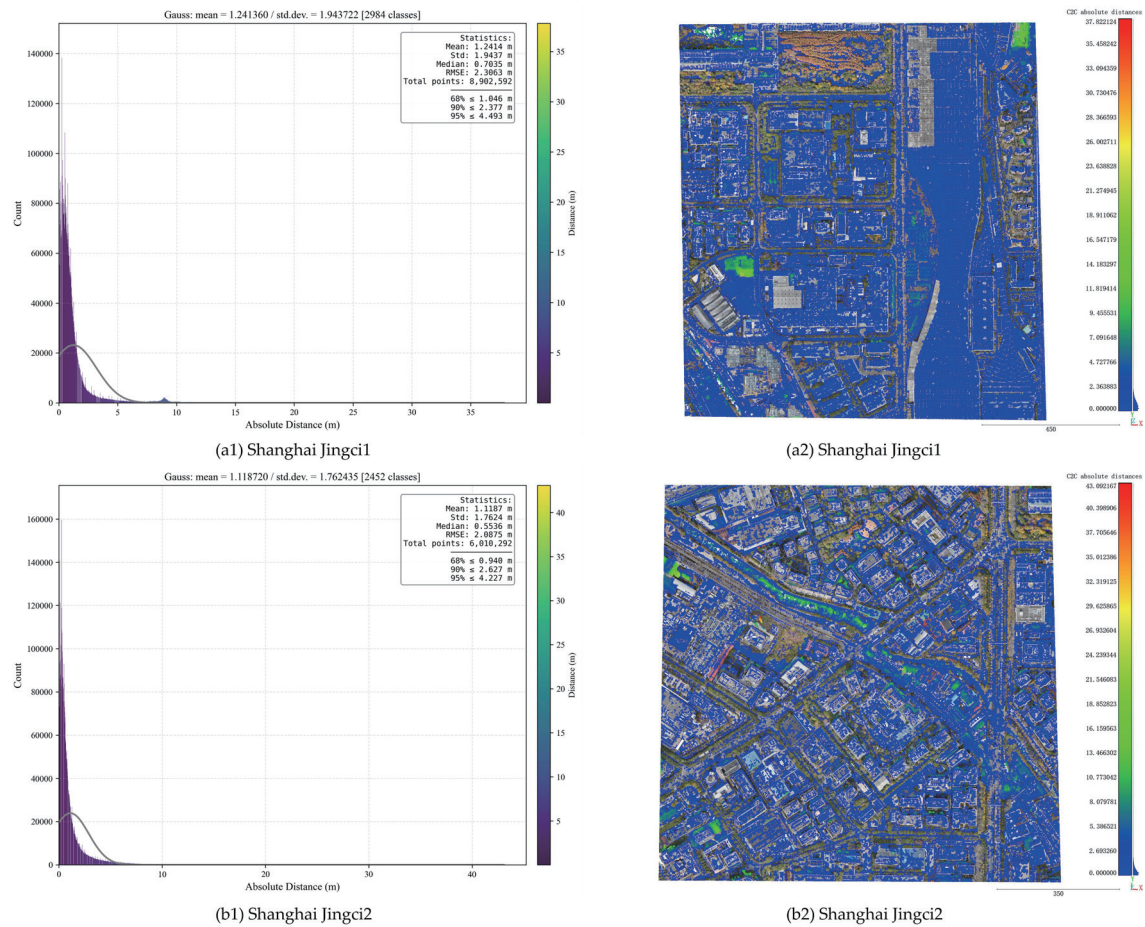


Fig. 9. (Color online) Global spatial consistency evaluation results of derived point clouds and airborne LiDAR reference data in Shanghai Jingci area. (a1, b1) Histograms of 3D Euclidean distances with Gaussian distribution fit (mean = 1.24 m, std.dev. = 1.94 m for Jingci1; mean = 1.12 m, std.dev. = 1.76 m for Jingci2). (a2, b2) Corresponding spatial distributions of absolute geometric deviations. The low mean errors and uniform dispersion patterns across both Shanghai Jingci regions demonstrate the method's high geometric fidelity and robustness in preserving spatial integrity under complex urban topographies.

The discrepancy between the global C2C distance statistics and the elevation-profile *RMSE* can be mainly explained by the different evaluation scales and error sensitivities of the two metrics. The global C2C statistics in Fig. 9 are computed over the entire point cloud and are dominated by a large number of stable surfaces, such as roads, roofs, and open terrain. Therefore, the mean C2C errors remain relatively low, with values of 1.241 and 1.119 m for the two Jingci experiments. In contrast, elevation-profile *RMSE* is calculated along narrow transects and is more sensitive to local height discontinuities and changed objects. Because *RMSE* squares the elevation residuals, a small number of local outliers can significantly increase the final value.

To further identify the sources of these local errors, Fig. 12 shows the satellite stereo images and the corresponding airborne LiDAR reference data for high-*RMSE* profile segments in the

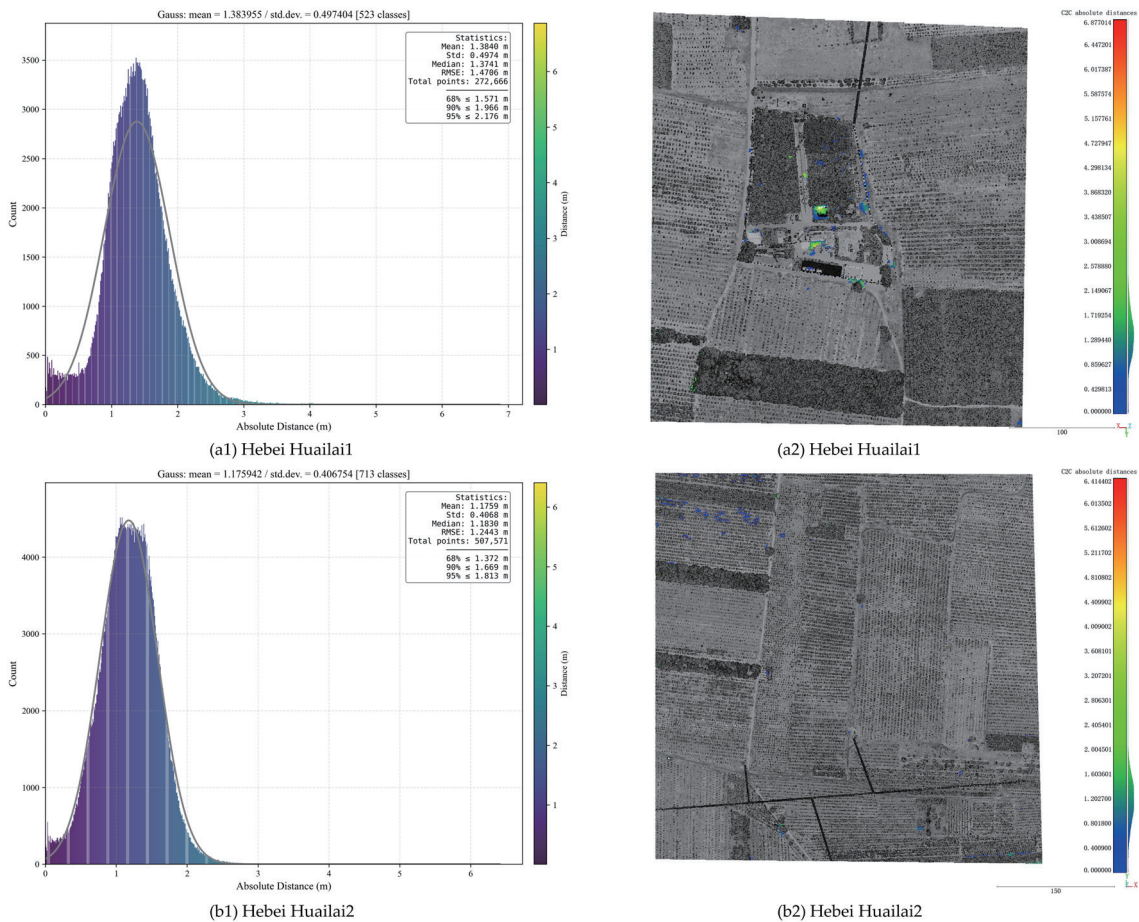


Fig. 10. (Color online) Global spatial consistency evaluation results of derived point clouds and airborne LiDAR reference data in Hebei Huailai region. (a1, b1) Histograms of 3D Euclidean distances with Gaussian distribution fit (mean = 1.18 m, std.dev. = 0.50 m for Huailai1; mean = 1.18 m, std.dev. = 0.41 m for Huailai2). (a2, b2) Corresponding spatial distributions of absolute geometric deviations. The low mean errors and tight error dispersion patterns across both subregions (Huailai1 and Huailai2) demonstrate the method's high geometric fidelity and robustness in preserving spatial integrity for agricultural terrains with complex field structures.

Table 2
Quantitative statistics of elevation profile results for reconstructed point clouds and LiDAR reference in the Jingci test sites.

Test sites	Profile	RMSE (m)	MAE (m)	±1 m (%)	±2 m (%)	Points
Shanghai Jingci1	X1	3.918	2.305	48.5	72.6	4755
	X2	3.235	1.857	52.8	74.3	5679
	Y1	7.983	3.406	42.7	68.5	4089
	Y2	5.146	2.645	37.3	54.7	4188
	Average	5.071	2.803	45.3	67.5	4677
Shanghai Jingci2	X1	5.226	2.754	42.3	62.6	3226
	X2	3.280	1.834	57.5	75.6	5890
	Y1	7.086	2.912	51.2	68.8	2930
	Y2	6.987	3.470	43.9	60.8	2826
	Average	5.645	2.743	48.8	66.9	3718

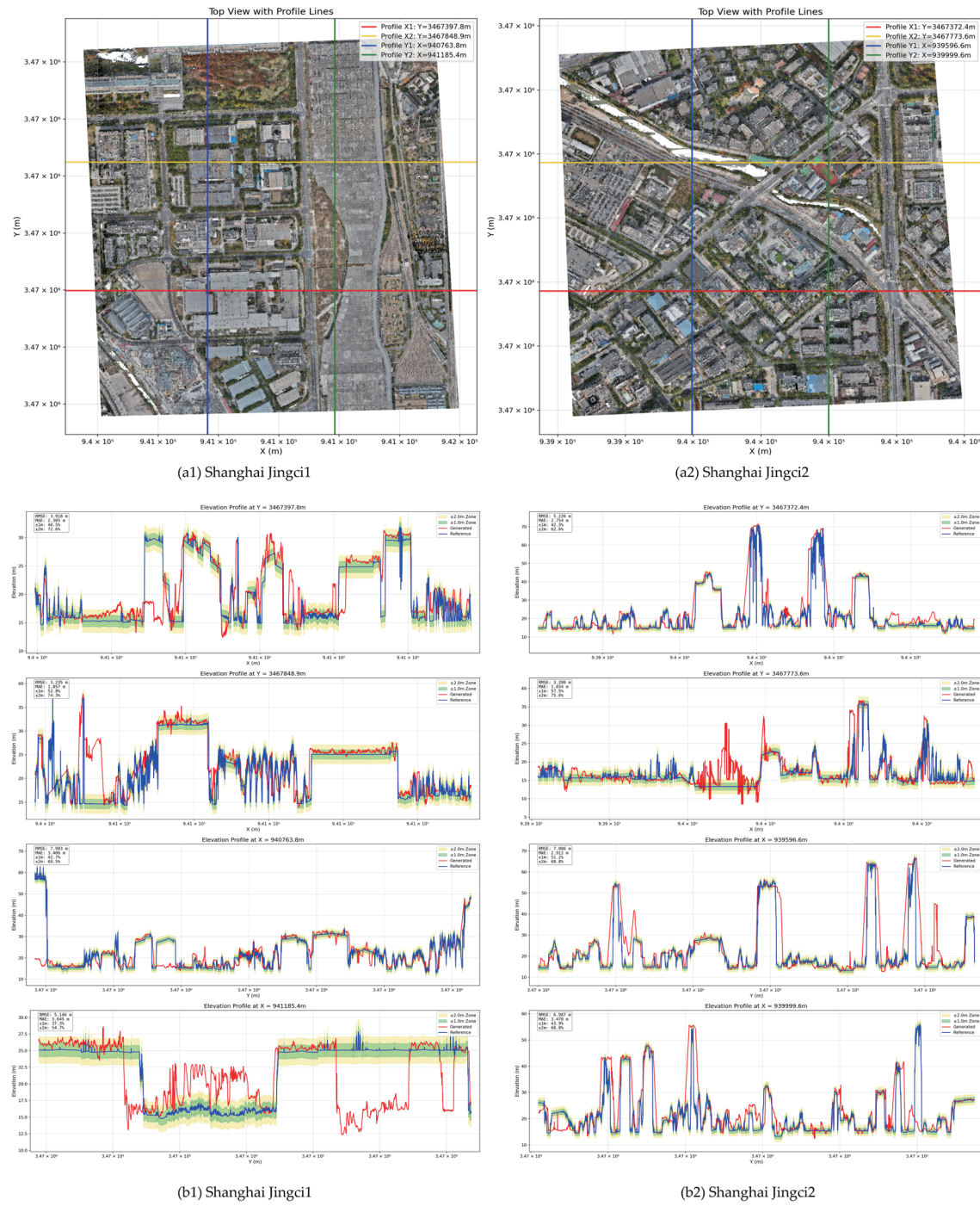


Fig. 11. (Color online) Quantitative evaluation results of local structural precision through cross-sectional elevation profile comparison. (a1, a2) Top views of Shanghai Jingci subregions with profile lines (X1, X2, Y1, Y2) indicating transect locations. (b1, b2) Elevation profiles along X- and Y-axes showing derived point clouds (red) and airborne LiDAR reference data (blue), with MAE and RMSE values displayed in the insets. The consistent alignment of elevation curves across multiple transects validates the method's capability in preserving fine-scale structural details and terrain features in complex urban environments.

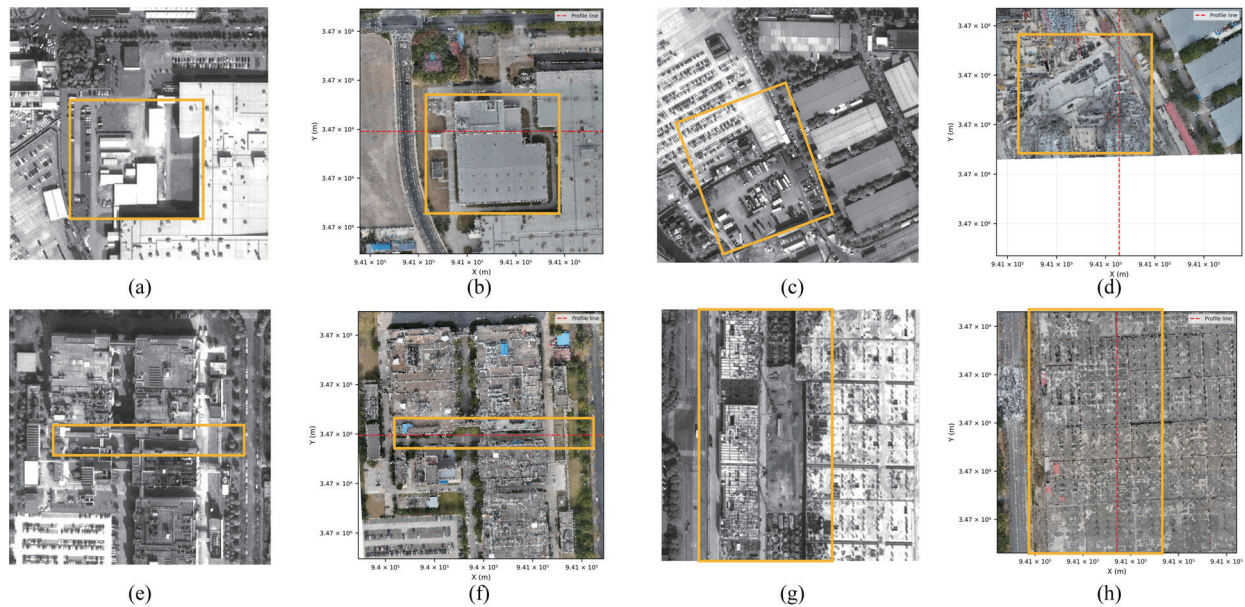


Fig. 12. (Color online) High-*RMSE* profile segments in the Shanghai Jingci area used for local comparison. Satellite stereo images and airborne LiDAR reference data in paired panels are compared. Yellow boxes mark local inconsistencies between the airborne LiDAR reference data and the satellite stereo imagery.

Jingci area. As shown in Figs. 12(a)–12(d) and 12(g) and 12(h), several regions within the yellow boxes exhibit clear land-cover or building changes between the satellite image acquisition and the airborne LiDAR survey. Some buildings present in the LiDAR reference data are absent or structurally different in the stereo images, resulting in inconsistent elevation profiles. In addition, Figs. 12(e) and 12(f) show that vegetation growth is another important source of local discrepancy. Tall vegetation or trees are not evident in the satellite stereo images but appear clearly in the airborne LiDAR point cloud, producing local elevation differences along the selected transects.

Furthermore, a specific anomaly is observed in a segment of profile X2 [Fig. 11(b2)], where the generated point cloud exhibits irregular elevation spikes significantly higher than those of the LiDAR reference. Cross-referencing reveals that this segment corresponds to a small water body. The inherent weak texture and specular reflection of water surfaces inevitably cause local matching failures in the optical stereo images, leading to these anomalous elevation artifacts. This is also consistent with the spatial error distribution in Fig. 9(b2), where many high-error points protruding from the reconstructed dense matching result correspond to trees, vegetated areas, or water bodies. Therefore, the relatively high elevation-profile *RMSE* in the Jingci area does not indicate a systematic degradation of the reconstructed point cloud, but rather reflects localized temporal changes, vegetation-related differences, and inherent optical matching limitations over untextured surfaces between the two datasets.

Similar observations are found in the Shanghai Jingci2 area, as shown in Fig. 13. The paired comparison indicates that the high-*RMSE* profile segments are mainly associated with local inconsistencies between the satellite stereo imagery and the airborne LiDAR reference data. In Figs. 13(a) and 13(b), and 13(e) and 13(f), vegetation between buildings shows notable differences in canopy height and spatial extent. In contrast, the yellow-box regions in Figs. 13(c) and 13(d), and 13(g) and 13(h) correspond mainly to building changes, where constructed structures are evident in the LiDAR data but differ from the corresponding satellite-image observations. These results further support that the higher profile *RMSE* in Jingci is primarily caused by localized temporal changes rather than systematic reconstruction errors.

The quantitative evaluation results of the cross-sectional elevation profiles in the Huailai test sites are presented in Fig. 14 and Table 3. Compared with the Jingci urban area, the Huailai results show higher stability and consistency. The first experiment achieves an average *RMSE* of 2.274 m and an *MAE* of 1.843 m, with 74.0% of the profile points within ± 2 m. The second experiment further improves the accuracy, yielding an *RMSE* of 1.451 m and an *MAE* of 1.339 m, with 92.8% of the points within ± 2 m.

The remaining local errors are mainly related to temporal and seasonal differences between the datasets. As shown in Fig. 15, the Huailai area is dominated by agricultural land and woodland, and the yellow boxes mark regions with relatively large profile discrepancies. The airborne LiDAR data were acquired on July 8, 2021, whereas the GF-7 panchromatic stereo images were acquired on February 7, 2021 and March 3, 2024. Differences in vegetation growth, crop coverage, and canopy height between these acquisition dates can therefore introduce local elevation deviations. In addition, the tower crane observed in the LiDAR data in Figs. 15(a) and

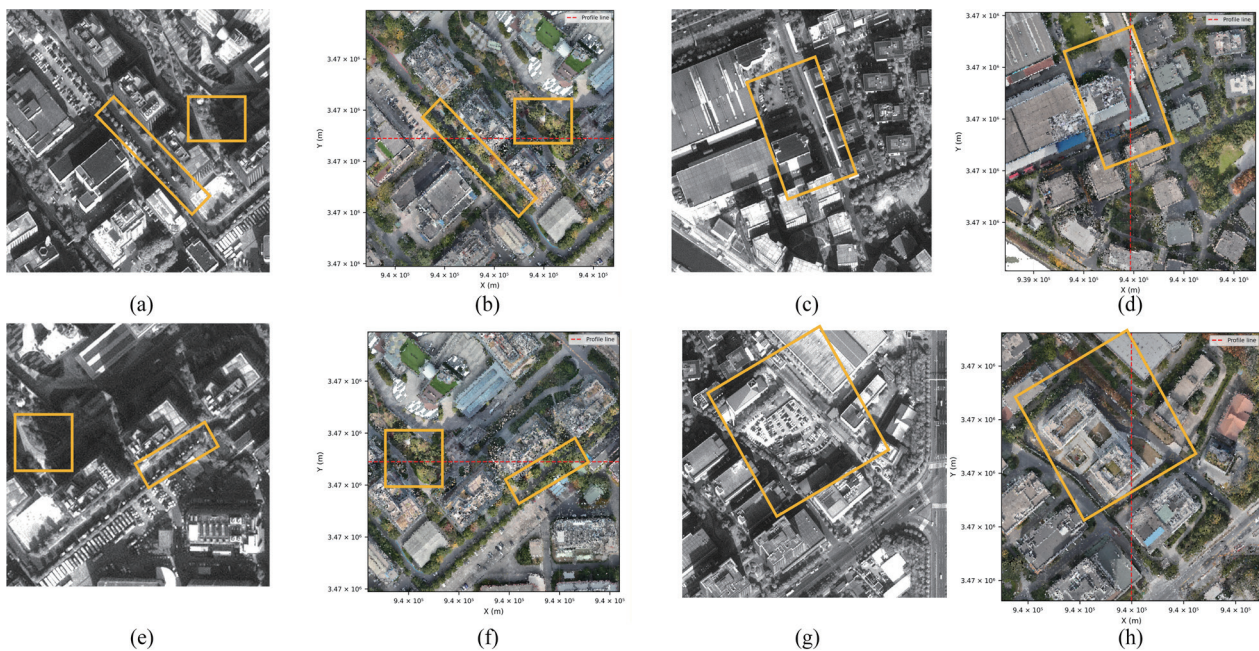


Fig. 13. (Color online) High-*RMSE* profile segments in the Shanghai Jingci2 area used for local comparison.

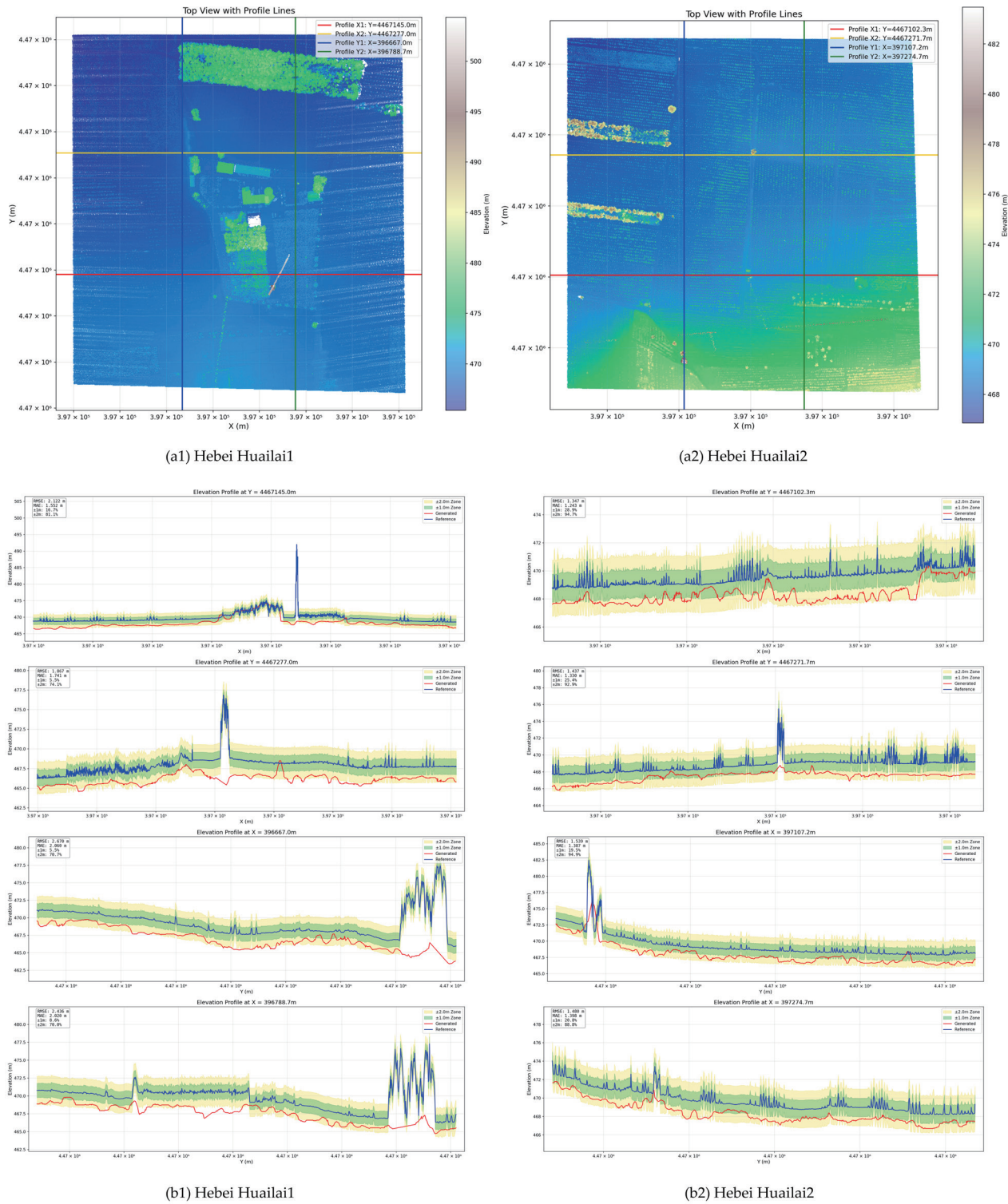


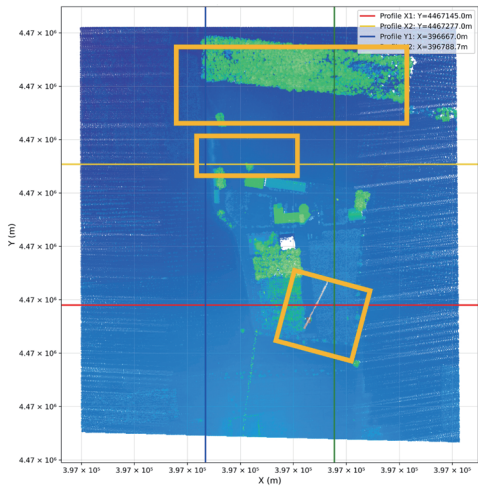
Fig. 14. (Color online) Quantitative evaluation results of local structural precision through cross-sectional elevation profile comparison. (a1, a2) Top views of Hebei Huailai subregions with profile lines (X1, X2, Y1, Y2) indicating transect locations. (b1, b2) Elevation profiles along X- and Y-axes for derived point clouds (red) and airborne LiDAR reference data (blue), with MAE and RMSE values displayed in the insets.

Table 3
Quantitative statistics of elevation profile for reconstructed point clouds and LiDAR reference in the Huailai test sites.

Test sites	Profile	<i>RMSE</i> (m)	<i>MAE</i> (m)	±1 m (%)	±2 m (%)	Points
Hebei Huailai1	X1	2.122	1.552	16.7	81.1	747
	X2	1.867	1.741	5.5	74.1	714
	Y1	2.670	2.060	5.5	70.7	587
	Y2	2.436	2.020	8.6	70.0	590
	Average	2.274	1.843	9.1	74.0	660
Hebei Huailai2	X1	1.347	1.243	28.9	94.7	1017
	X2	1.437	1.330	25.4	92.9	1007
	Y1	1.539	1.387	19.5	94.9	799
	Y2	1.480	1.398	20.8	88.8	807
	Average	1.451	1.339	23.7	92.8	908



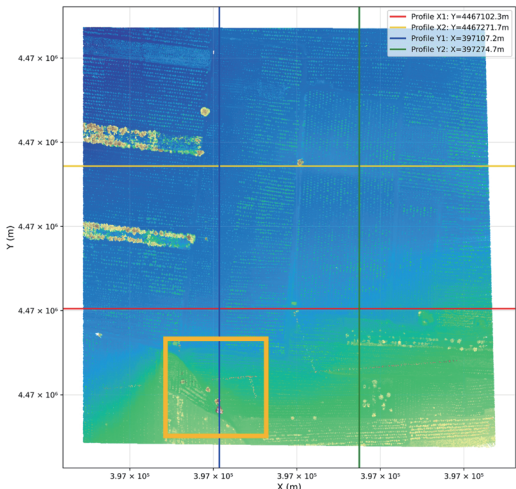
(a)



(b)



(c)



(d)

Fig. 15. (Color online) High-*RMSE* profile segments in the Huailai area used for local comparison.

15(b) represents a temporary object, which also contributes to local residuals. The slight overall offset may also be partly associated with residual georeferencing uncertainty in the airborne LiDAR reference data. Since the Huailai LiDAR point cloud was generated using a real-time positioning solution, errors in ground base-station coordinates during point-cloud coordinate computation, together with residual GNSS/INS trajectory errors, may propagate into the final georeferenced point cloud. These factors explain the local high-error segments without indicating a systematic failure of the proposed method.

Overall, the reconstructed elevation profiles exhibit stable trends and concentrated error distributions, confirming the reliability of the proposed method for continuous terrain reconstruction.

6. Conclusions and Future Work

In this study, we presented a dense stereo matching framework for high-resolution optical satellite imagery, incorporating an adaptive minimum disparity estimation strategy and a multi-scale block-wise optimization scheme. The proposed method effectively addresses the sensitivity of conventional approaches to minimum disparity settings, particularly in complex terrain and urban environments.

Experimental results across urban, mountainous, and agricultural regions demonstrate that the proposed method achieves stable reconstruction performance, with overall geometric errors maintained at approximately the meter level. The reconstructed elevation profiles show consistent trends with airborne LiDAR data, indicating reliable terrain and structural recovery.

Satellite-based 3D reconstruction offers significant advantages in cost efficiency and large-area coverage compared with airborne or terrestrial acquisition methods. However, reconstruction accuracy remains affected by several factors, including image quality, imaging geometry, and RPC model accuracy. High-quality commercial satellite imagery provides better disparity continuity and noise suppression, leading to more stable reconstruction results. In addition, inaccurate or unavailable RPC parameters may introduce systematic geometric errors if not properly refined using ground control points or block adjustment.

Occlusions and weak-texture regions in dense urban areas and mountain shadows may still produce mismatches or disparity gaps, indicating the limitations of purely image-based matching approaches.

Future work will be focused on integrating multi-source auxiliary constraints to improve matching robustness in occluded and weak-texture regions, developing more adaptive cost aggregation and disparity optimization strategies, and exploring hybrid frameworks that combine deep learning priors with geometric constraints to further enhance reconstruction accuracy and robustness.

Acknowledgments

This research did not receive any specific grant from any funding agency in the public, commercial, or nonprofit sectors. All authors read and approved the final manuscript. The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

References

- 1 D. Poli and T. Toutin: Photogramm. Rec. **27** (2012) 58. <https://doi.org/10.1111/j.1477-9730.2011.00665.x>
- 2 R. Qin: ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci. **III-1** (2016) 77. <https://doi.org/10.5194/isprs-annals-III-1-77-2016>
- 3 B. Zhu, X. Tan, and H. Li: Sci. Rep. **14** (2024) 9309. <https://doi.org/10.1038/s41598-024-59980-x>
- 4 X. Wang, L. Jiang, F. Wang, H. You, and Y. Xiang: Remote Sens. **16** (2024) 487. <https://doi.org/10.3390/rs16030487>
- 5 R. A. Hamzah and H. Ibrahim: J. Sens. **2016** (2016) 8742920. <https://doi.org/10.1155/2016/8742920>
- 6 Y. Boykov, O. Veksler, and R. Zabih: IEEE Trans. Pattern Anal. Mach. Intell. **23** (2001) 1222. <https://doi.org/10.1109/34.969114>
- 7 J. Sun, N. N. Zheng, and H. Y. Shum: IEEE Trans. Pattern Anal. Mach. Intell. **25** (2003) 787. <https://doi.org/10.1109/TPAMI.2003.1206509>
- 8 Y. Li, S. Zheng, X. Wang, and H. Ma: Comput. Geosci. **97** (2016) 58. <https://doi.org/10.1016/j.cageo.2016.09.002>
- 9 A. Geiger, M. Roser, and R. Urtasun: Proc. 10th Asian Conf. Computer Vision (Springer, Berlin, Heidelberg, 2011) 25-38. https://doi.org/10.1007/978-3-642-19315-6_3
- 10 J. Žbontar and Y. LeCun: J. Mach. Learn. Res. **17** (2016) 1. <https://www.jmlr.org/papers/v17/15-535.html>
- 11 A. Kendall, H. Martirosyan, S. Dasgupta, P. Henry, R. Kennedy, A. Bachrach, and A. Bry: Proc. IEEE Int. Conf. Comput. Vis. (IEEE, 2017) 66-75. <https://doi.org/10.1109/ICCV.2017.17>
- 12 J. R. Chang and Y. S. Chen: Proc. IEEE/CVF Conf. Computer Vision Pattern Recognition (IEEE, 2018) 5410-5418. <https://doi.org/10.1109/CVPR.2018.00567>
- 13 Z. Zhang, W. Chang, F. Mo, G. Mao, G. Chen, X. Niu, and J. Song: Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci. **XLVIII-4/W14-2025** (2025) 429. <https://doi.org/10.5194/isprs-archives-XLVIII-4-W14-2025-429-2025>
- 14 R. Mari, T. Ehret, and G. Facciolo: Image Process. On Line **12** (2022) 501. <https://doi.org/10.5201/ipol.2022.435>
- 15 X. He, S. Jiang, S. He, Q. Li, W. Jiang, and L. Wang: Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci. **XLVIII-1/W2-2023** (2023) 1635. <https://doi.org/10.5194/isprs-archives-XLVIII-1-W2-2023-1635-2023>
- 16 H. Albanwan and R. Qin: arXiv:2205.14051 (2022). <https://arxiv.org/abs/2205.14051>
- 17 O. Krejcar and H. Namazi: Environ. Earth Sci. **85** (2026) 78. <https://doi.org/10.1007/s12665-025-12798-w>
- 18 H. Hirschmüller: IEEE Trans. Pattern Anal. Mach. Intell. **30** (2008) 328. <https://doi.org/10.1109/TPAMI.2007.1166>
- 19 M. Fuentes Reyes, P. d'Angelo, and F. Fraundorfer: Remote Sens. **17** (2025) 1. <https://doi.org/10.3390/rs17010001>
- 20 Y. Han, R. Qin, and X. Huang: Photogramm. Rec. **35** (2020) 58. <https://doi.org/10.1111/phor.12310>
- 21 M. A. Aguilar, A. Nemmaoui, F. J. Aguilar, and R. Qin: GISci. Remote Sens. **56** (2019) 109. <https://doi.org/10.1080/15481603.2018.1494408>
- 22 A. Nemmaoui, F. J. Aguilar, M. A. Aguilar, and R. Qin: Comput. Electron. Agric. **164** (2019) 104903. <https://doi.org/10.1016/j.compag.2019.104903>
- 23 M. A. Aguilar, R. Jiménez-Lao, and F. J. Aguilar: ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci. **V-2-2022** (2022) 31. <https://doi.org/10.5194/isprs-annals-V-2-2022-31-2022>
- 24 H. Hirschmüller: Photogramm. Week **11** (2011) 173. <https://elib.dlr.de/73119/>
- 25 M. Bleyer, C. Rhemann, and C. Rother: Proc. British Machine Vision Conf. (BMVA Press, Dundee, 2011) 14.1-14.11. <https://doi.org/10.5244/C.25.14>

- 26 K. J. Zuiderveld: Graphics Gems IV, P. S. Heckbert, Ed. (Academic Press Professional, San Diego, 1994) pp. 474–485. <https://doi.org/10.1016/B978-0-12-336156-1.50061-6>
- 27 K. He, J. Sun, and X. Tang: IEEE Trans. Pattern Anal. Mach. Intell. **35** (2013) 1397. <https://doi.org/10.1109/TPAMI.2012.213>
- 28 D. G. Lowe: Proc. 7th IEEE Int. Conf. Computer Vision (IEEE, 1999) 1150–1157. <https://doi.org/10.1109/ICCV.1999.790410>
- 29 E. Rublee, V. Rabaud, K. Konolige, and G. Bradski: Proc. 2011 IEEE Int. Conf. Computer Vision (IEEE, 2011) 2564–2571. <https://doi.org/10.1109/ICCV.2011.6126544>
- 30 C. V. Tao and Y. Hu: Photogramm. Eng. Remote Sens. **67** (2001) 1347. https://www.asprs.org/wp-content/uploads/pers/2001journal/december/2001_dec_1347-1357.pdf
- 31 Y. Hu and C. V. Tao: Photogramm. Eng. Remote Sens. **68** (2002) 715. https://www.asprs.org/wp-content/uploads/pers/2002journal/july/2002_jul_715.pdf
- 32 J. Grodecki and G. Dial: Proc. Joint Workshop ISPRS Working Groups I/2, I/5, and IV/7 on High Resolution Mapping from Space (University of Hannover, Hannover, 2001) 77–86.

About the Authors



Lingyan Kong is a professor-level senior engineer and vice president of the Beijing Institute of Surveying and Mapping, China. He has long been engaged in geomatics and GIS research and applications, leading numerous major projects and receiving over ten provincial- and ministerial-level awards. His work has significantly contributed to urban planning and the surveying industry. (kongly@163.com)



Qingfeng Zeng is a professor-level senior engineer with an M.S. degree. He has long focused on GIS development and research, specializing in geospatial data processing, industry-oriented GIS systems, and engineering applications. His work emphasizes technological innovation and the practical transformation of geospatial solutions. (zengqingfeng11@163.com)



Haiying Ren is a professor-level senior engineer with an M.S. degree. She has long been engaged in geomatics and geographic information applications, receiving more than 60 national, municipal, and institutional awards. She has published over 40 scientific papers and led more than ten major projects in real estate management, urban governance, and low-altitude monitoring, contributing significantly to urban governance and the low-altitude economy. (14512829@qq.com)



Dan Geng is a professor-level senior engineer with a master's degree. She is currently a deputy chief engineer at the Comprehensive Investigation and Design Institute of Construction Co., Ltd. She has long been engaged in surveying, mapping, and geospatial information technologies. She has received multiple national and provincial-level awards and has made significant contributions to geospatial industry development and technological innovation. (gengd.bj@qq.com)



Ruijia Deng received his B.S. degree in surveying and mapping engineering from the China University of Mining and Technology, China, in 2024. He is currently pursuing his M.S. degree in cartography and geographic information systems at the University of Chinese Academy of Sciences, Beijing, China, and also conducts research at the Key Laboratory of Digital Earth Science, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing, China, and the International Research Center of Big Data for Sustainable Development Goals, Beijing, China. His research interests are in LiDAR remote sensing. (dengruijia24@mails.ucas.ac.cn)