remote sensing

MDPI

Article

Hierarchical Edge-Preserving Dense Matching by Exploiting Reliably Matched Line Segments

Yi Yue [1,2,3], Tong Fang [2], Wen Li [2], Min Chen [1,2,*], Bo Xu [2], Xuming Ge [2], Han Hu [2] and Zhanhao Zhang [4]

[1] The Key Laboratory of Urban Land Resources Monitoring and Simulation, Ministry of Natural Resources, Shenzhen 518040, China; yueyi1998@foxmail.com
[2] Faculty of Geosciences and Environmental Engineering, Southwest Jiaotong University, Chengdu 611756, China; ft@my.swjtu.edu.cn (T.F.); keylw@my.swjtu.edu.cn (W.L.); xubo@swjtu.edu.cn (B.X.); xuming.ge@swjtu.edu.cn (X.G.); han.hu@swjtu.edu.cn (H.H.)
[3] Hunan Institute of Surveying and Mapping Science and Technology, Changsha 410007, China
[4] China Railway Eryuan Engineering Group Co., Ltd., Chengdu 610038, China; 15146378899@163.com
* Correspondence: minchen@home.swjtu.edu.cn

Abstract: Image dense matching plays a crucial role in the reconstruction of three-dimensional models of buildings. However, large variations in target heights and serious occlusion lead to obvious mismatches in areas with discontinuous depths, such as building edges. To solve this problem, the present study mines the geometric and semantic information of line segments to produce a constraint for the dense matching process. First, a disparity consistency-based line segment matching method is proposed. This method correctly matches line segments on building structures in discontinuous areas based on the assumption that, within the corresponding local areas formed by two corresponding line pairs, the disparity obtained by the coarse-level matching of the hierarchical dense matching is similar to that derived from the local homography estimated from the corresponding line pairs. Second, an adaptive guide parameter is designed to constrain the cost propagation between pixels in the neighborhood of line segments. This improves the rationality of cost aggregation paths in discontinuous areas, thereby enhancing the matching accuracy near building edges. Experimental results using satellite and aerial images show that the proposed method efficiently obtains reliable line segment matches at building edges with a matching precision exceeding 97%. Under the constraint of the matched line segments, the proposed dense matching method generates building edges that are visually clearer, and achieves higher accuracy around edges, than without the line segment constraint.

Keywords: dense matching; edge preserving; line segment matching; rough disparity; derived disparity; disparity consistency

check for updates

Citation: Yue, Y.; Fang, T.; Li, W.; Chen, M.; Xu, B.; Ge, X.; Hu, H.; Zhang, Z. Hierarchical Edge-Preserving Dense Matching by Exploiting Reliably Matched Line Segments. *Remote Sens.* **2023**, *15*, 4311. https://doi.org/10.3390/rs15174311

Academic Editor: Geert Verhoeven

Received: 25 July 2023
Revised: 18 August 2023
Accepted: 23 August 2023
Published: 1 September 2023

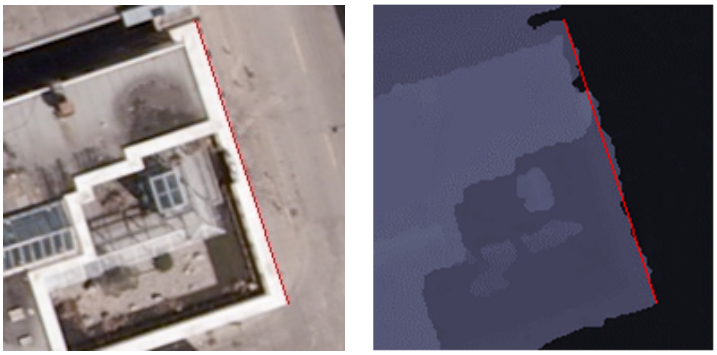
Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

The automatic reconstruction of three-dimensional (3D) models of buildings in urban areas is a fundamental issue in the fields of photogrammetry and computer vision [1–4]. Image dense matching-based 3D reconstruction has become one of the most widely used technologies in the 3D reconstruction of building models due to its economy, flexibility, and high-density point clouds [5,6].

Many dense matching methods have been proposed, encompassing both hand-crafted feature-based approaches and deep learning-based methods [7]. Given that the generalization ability of deep learning-based methods restricts their widespread application, this study focuses on hand-crafted feature-based methods. The existing hand-crafted algorithms can be divided into global methods and local methods [8]. The global matching method often entails high computational costs and is limited in its applicability, whereas the local matching method exhibits limitations in terms of model versatility and accuracy. Semi-global matching (SGM) amalgamates the strengths of global and local algorithms [9,10]

by computing the disparities for each pixel within a certain range along several one-dimensional paths and aggregating the costs. The cost values calculated from multiple directional paths are then summed, and the disparity with the minimum total energy for each pixel is retained using the winner-takes-all strategy. SGM achieves results similar to global methods while maintaining high computational efficiency. The SGM method, along with its improved variants, has extensive applications in the field of building reconstruction [11–14]. However, they rely on fixed support windows, which assume consistent disparities for all points within the window. This assumption neglects discontinuities in disparities within the window, leading to matching ambiguity. Geometric errors such as outward expansion and distortion occur at object edges, as illustrated in Figure 1.



(**a**) (**b**)

Figure 1. Example of matching errors at building edges. The red lines indicate building edges. (**a**) Original image and (**b**) disparity map generated by using the tSGM method [14].

Various edge-constrained dense matching methods have been proposed to enhance the matching performance at object edges [14–16]. However, because these methods do not clarify whether the edges are located in areas with discontinuous disparity and apply the same refinement to all edges, noise will be introduced to the refinement on edges in areas with continuous disparity.

Urban scenes often exhibit numerous line segments along the edges of buildings. Leveraging the geometric and semantic information of these line segments to constrain the dense matching holds great potential for improving the matching performance at building edges. Nonetheless, two challenges arise in the dense matching process when utilizing line segment constraints. The first challenge is how to efficiently obtain reliable line segment matches to furnish dense matching constraints for a broader range of building edges. The second challenge involves leveraging the matched line segments to guide the dense matching, specifically at building edges.

To tackle these challenges, this study develops an efficient and reliable line segment matching method. On the basis of this method, a dense matching method is developed under the constraint of matched line segments. First, a rough disparity map is generated from the coarse-level matching of a hierarchical dense matching framework. Subsequently, line segments are extracted and grouped to form line pairs from the images to be densely matched, and the rough disparity map and epipolar geometry are exploited to establish constraints that enable candidate line pair matches to be identified. A "hypothesis–verification" strategy is then proposed to refine the candidate matches. Specifically, a candidate local homography is derived from a pair of candidate line pair matches, and a disparity map (named "derived disparity" in this paper) corresponding to the local region formed by the line pair is calculated according to the candidate local homography. A similarity measure based on the consistency between the derived disparity and the rough disparity in the local region is constructed to generate the final line segment matches from the candidate line pair matches. Finally, based on the rough disparity map, building structural line segments are extracted from the matched line segments, and these line segments are used to constrain the cost aggregation of the fine-level matching in the hierarchical dense matching.

Different penalty strategies are applied to pixels on each side of the building structural line segments to enhance the distinctiveness of building edges in the dense matching. Our main contributions are as follows.

First, a reliable and efficient line segment matching method that can correctly match line segments on building edges with discontinuous parallax is proposed. There is no need to calculate grayscale feature descriptors in the matching procedure, resulting in high efficiency. Moreover, because the proposed disparity consistency-based similarity measure is sensitive to discontinuous areas, the proposed method can correctly match line segments on building edges.

Second, a dense matching method is developed under the constraint of the matched line segments. The cost propagation and penalties are applied under the constraint of the line segments in the neighborhood of building edges, which helps to improve the dense matching accuracy and feature preservation in these areas.

The remainder of this paper is organized as follows. Section 2 provides an overview of related work. Section 3 introduces our proposed method, which focuses on line segment matching and the line segment constrained hierarchical dense matching approach. Section 4 presents the experimental results and analysis of satellite and aerial datasets. Section 5 concludes this study.

2. Related Work

The SGM method [9,10] is a milestone in the field of dense matching because it combines the advantages of global and local matching methods. The tSGM method [14] incorporates a hierarchical matching framework on the basis of the SGM algorithm, which transfers the matching results from coarser to finer levels. As a result, the memory requirements and time overhead are greatly reduced, and the matching ambiguity in areas with weak textures is effectively alleviated, enhancing the overall integrity of the disparity results. However, because SGM and its variants employ a fixed support window and ignore the discontinuous areas in the window, false matches can be generated at the edges of buildings.

To improve the matching accuracy on building edges, many improvements have been made in the cost aggregation and post-processing stages. Introducing edge features into the cost aggregation process is effective in reducing matching ambiguities [14–16]. Several local methods construct adaptive edge-aware windows based on local properties such as grayscale [17–19] or employ edge-preserving filtering methods with adaptive weights [20,21] to improve the matching performance in regions with depth discontinuities. The tSGM method assigns smaller penalty parameters to edge pixels in the path optimization process of SGM, which improves the matching performance of edge pixels with discontinuous disparity. Chuang et al. [15] incorporated edge constraints into the local support window aggregation and introduced a weight to handle edge pixels. Tatar et al. [22] proposed a superpixel-based weighted cost aggregation that reduces the influence of depth-discontinuous paths. These methods contribute to the generation of sharp edges in building reconstruction. However, as these methods assume depth discontinuities for all edge pixels, incorporating all edges as prior constraints inevitably introduces erroneous matches, particularly for images with complex textures. Kim et al. [16] selected edges located at discontinuous areas based on a pre-estimated disparity map and penalized small disparity differences between edge pixels and neighboring pixels. However, this method does not differentiate between the foreground and the background. In reality, pixels in foreground regions should have similar disparities to edge pixels. Thus, this approach may introduce some erroneous clues for edge pixel matching, leading to over-sharpening of edges.

Post-processing methods can be applied to address the disparity map generated from dense matching. These algorithms either perform global optimization based on grayscale features [23] or refine the disparity map using higher-order features such as line segments or planes [24–26]. While these methods often achieve good results, they typically

assume a discontinuous disparity at edge pixels or coplanarity of planes, which can reduce the matching accuracy when used inappropriately. Additionally, post-processing incurs additional computational costs.

Compared with grayscale edges, line segments contain richer semantic and structural information. Incorporating correctly matched line segments as prior constraints into the cost aggregation of dense matching can enhance the reconstruction accuracy of building edges. Line segment matching methods are typically divided into individual line segment-based matching [27,28] and line group-based matching. The former relies on photometric information in the vicinity of the line segments and requires a considerable overlap between the extracted corresponding line segments. Challenges are encountered when matching parallel line segments with similar textures and line segments from satellite images with seasonal landscape variations. The line group-based methods introduce more geometric constraints to alleviate ambiguities [29–31]. However, existing line group-based methods are prone to false matches of building structural line segments in discontinuous areas, and these structural line segments are meaningful features in the constrained dense matching. Qin et al. [26] used a coarse disparity map to guide the matching of line segments, which improves the matching accuracy while reducing the time cost of line segment matching. However, the need to construct complex grayscale feature descriptors when calculating the similarity measure of line segments entails a large time cost. As a preprocessing step for disparity refinement, this excessive time cost is unacceptable.

Given that urban scenes contain rich linear structures, this study mines corresponding line segments to constrain the dense matching. However, there are obvious differences between this study and existing methods for line segment matching and line segment-constrained dense matching. For line segment matching, this study employs epipolar geometry and rough disparity maps to greatly narrow the search range of candidate matches. There is no need to calculate grayscale feature descriptors in the similarity measurement, which enhances the matching efficiency. Additionally, because the rough disparity and the derived disparity are significantly different in discontinuous areas, the similarity measurement constructed based on the consistency of the two disparities enables improved matching of building structural line segments in discontinuous areas. For line segment-constrained dense matching, a guiding parameter for distinguishing background from foreground is proposed. This parameter penalizes the large disparity difference between foreground pixels and line segment pixels and encourages large disparity differences between background pixels and line segment pixels. Overall, this approach contributes to the correct handling of the matching problem of depth discontinuities and generates clear building boundaries.

3. Methodology

This study proposes a hierarchical dense matching strategy. First, an image pyramid is constructed for each of the two images to be matched, and the disparity map of the coarse level is successively transferred to finer levels until the upper level of the original image is reached. The disparity map of this level is denoted as the rough disparity map. Line segments are then extracted from the original images, and a reliable and efficient line segment matching method is proposed based on the rough disparity map to obtain corresponding line segments. The line segments located in discontinuous areas are screened out from the matched line segments through a rough disparity map-based consistency check. Most of these line segments are located at the edges of buildings, which are recorded as building structural line segments. Finally, the cost aggregation of the last layer of the hierarchical dense matching is performed under the constraint of the matched building structural line segments to obtain the final dense matching result. Figure 2 illustrates the workflow of the proposed method.

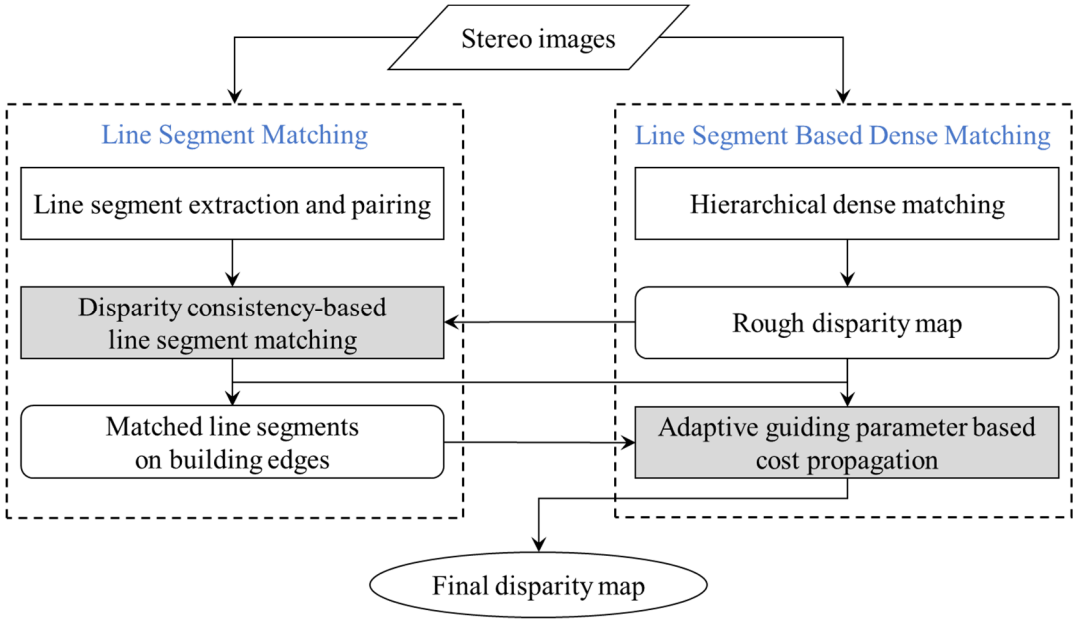


Figure 2. Workflow of the proposed method. The gray filled ones are the important steps of the proposed method.

3.1. Disparity Consistency-Based Line Segment Matching

In the proposed approach, the LSD method [32] is used to extract line segments. The line segments are then paired following the method of [31]. The images to be densely matched have been rectified, and a rough disparity map has been obtained through the coarse-level matching of the hierarchical dense matching. Thus, for each line pair on the reference image, candidate matches can be obtained through epipolar geometry and disparity range constraints.

Let LP_1^i (formed by line segments l_1^{i1} and l_1^{i2}) and LP_2^j (formed by line segments l_2^{j1} and l_2^{j2}) be corresponding line pairs with intersection points $p_1^i(x,y)$ and $p_2^j(x,y)$ on the reference image and search image, respectively. Then, p_1^i and p_2^j should satisfy the epipolar line constraint, that is, their vertical coordinates should be equal. Additionally, the difference in the horizontal coordinates of the intersection points should fall within a specified range. In this paper, ϵ_1 (set to three pixels) is used as the rectification error tolerance in the epipolar line constraint. Thus, in the matching process, if $\left|p_1^i(y) - p_2^j(y)\right| < \epsilon_1$ and $d_{min} < p_1^i(x) - p_2^j(x) < d_{max}$, LP_1^i and LP_2^j are considered to be a pair of candidate matches, where $[d_{min}, d_{max}]$ represents the disparity range, as shown in Figure 3.

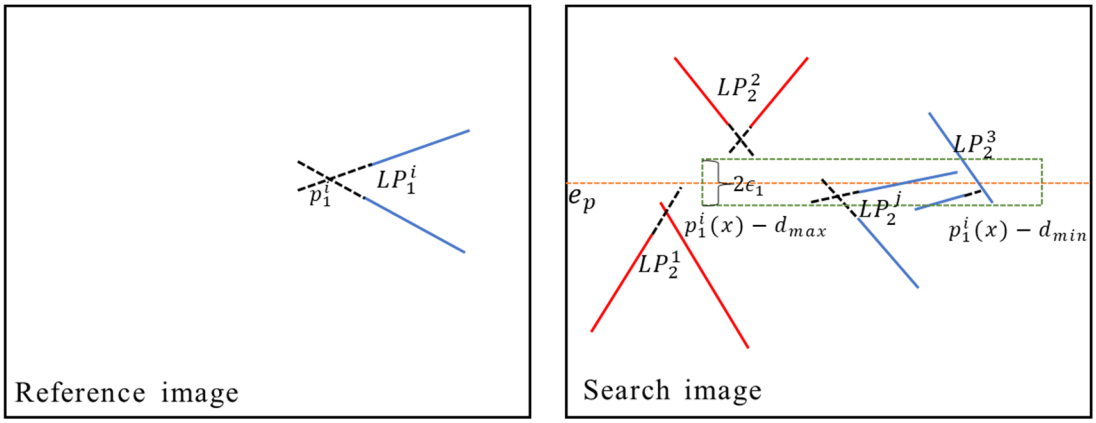


Figure 3. Epipolar line constraint and disparity range constraint.

An impact region $R_1^i\left(p_1^i, p_1^{i1}, p_1^{i2}, p_1^{ig}\right)$ is formed for line pair LP_1^i, where p_1^i is the intersection point; p_1^{i1} is the endpoint of line segment l_1^{i1} that is farther from p_1^i; p_1^{i2} is the endpoint of line segment l_1^{i2} that is farther from p_1^i; p_1^{ig} is determined according to $\overrightarrow{p_1^i p_1^{ig}} = \overrightarrow{p_1^i p_1^{i1}} + \overrightarrow{p_1^i p_1^{i2}}$. If LP_1^i and LP_2^j are two corresponding line pairs, two pairs of line segment matches can be obtained. Here, $\left(l_1^{i1}, l_2^{j1}\right)$ and $\left(l_1^{i2}, l_2^{j2}\right)$ are assumed to be the two pairs of line segment matches. During the pairing of line segments, each line segment is only paired with adjacent line segments that fulfill distance and angle conditions. Hence, the two line segments forming a line pair can be assumed to be coplanar [31]. Consequently, a local homography H_1^i can be estimated based on $\left(l_1^{i1}, l_2^{j1}\right)$ and $\left(l_1^{i2}, l_2^{j2}\right)$ [31,33].

The disparities of pixels in R_1^i can be derived from H_1^i, that is, the derived disparities. At the same time, the rough disparities of the corresponding pixels can be obtained from the rough disparity map of the hierarchical dense matching. If $\left(LP_1^i, LP_2^j\right)$ is a correct match, the distribution of the derived disparity and the distribution of the rough disparity of the pixels in R_1^i should be similar. Therefore, a disparity consistency-based similarity measure is constructed based on this assumption.

Let $X_1^i = \left\{x_{1i}^t\right\}_{t=1}^m$ be the homogeneous coordinates of valid pixels within R_1^i according to the rough disparity map. On the search image, the corresponding coordinates of x_{1i}^t estimated using the local homography matrix are given by $x_{2j}^t = H_1^i x_{1i}^t$. Consequently, the derived disparity d_{1i}^t at x_{1i}^t is calculated as $d_{1i}^t = x_{1i}^t - H_1^i x_{1i}^t$. The rough disparity D_{1i}^t can be obtained from the rough disparity map. Thus, the disparity consistency-based similarity measure is calculated as:

$$sim\left(LP_1^i, LP_2^j\right) = \frac{\sum_{t=1}^m e^{-\left|D_{1i}^t - d_{1i}^t\right|}}{am + (1-a)M}, \tag{1}$$

where m is the number of valid pixels and M is the total number of pixels in the impact region R_1^i; the parameter a is introduced to balance the influence of m and M. The similarity score for an individual pixel is computed using the exponential function to mitigate the impact of erroneous matches in the rough disparity map and disparity noise caused by fine structures in planar scenes. If only the total number of pixels is considered, the similarity measure becomes less sensitive to data variations, particularly in scenarios where the rough disparity map contains numerous invalid pixels. If only the valid pixels are considered, the similarity measure may fail to match line segment pairs due to minor disparity matching errors. To strike a balance between these factors, an empirical value of 0.5 is assigned to a.

For each line pair on the reference image, the similarity to each of its candidate matches is computed according to the method described above, as shown in Figure 4. The candidate match with the largest similarity (and greater than a threshold t_s) among all candidate matches is taken as the final match of the line pair. From the matched line pairs, line segment matches can be determined according to the relative positions of the line segments forming the line pairs, with the clockwise direction as the reference.

3.2. Line Segment Constrained Dense Matching

The SGM method introduces a path cost aggregation approach, which aggregates the matching costs of all disparities for each pixel along one-dimensional paths surrounding the pixel [9,10]. More specifically, the costs are recursively aggregated along a path r as follows:

$$L_r\left(p,d\right) = c(p,d) + min\left\{\begin{array}{c} L_r(p-r,d) \\ L_r(p-r,d-1) + P_1 \\ L_r(p-r,d+1) + P_1 \\ \min\limits_i L_r(p-r,i) + P_2 \end{array}\right\} - \min\limits_i L_r(p-r,i), \tag{2}$$

where $L_r\ (p,d)$ represents the accumulated cost of pixel p with disparity d along the current path; $c(p,d)$ represents the cost of pixel p with disparity d, and $p-r$ represents the previous pixel on the current path; P_1 and P_2 are smoothing parameters, where $P_1 < P_2$. To ensure local smoothness of the disparity map, a small penalty P_1 is assigned for a disparity change of 1 between neighboring pixels, while a higher penalty P_2 is assigned for larger disparity changes. The aggregated costs along all paths are summed, and the disparity D_p of pixel p is obtained by minimizing this sum.

$$D_p = \underset{d}{argmin} \sum_r L_r\ (p,d) \tag{3}$$

The line segments matched from Section 3.1 are incorporated as cues to constrain the fine-level path optimization stage in the hierarchical dense matching. First, the rough disparity map is utilized as a prior and only the matched line segments on discontinuous edges are retained. The direction of each line segment is determined based on the reference of occlusions between foreground and background. Subsequently, an adaptive guiding parameter is introduced and the cost aggregation is optimized to enhance the matching performance on building edges. The optimization reduces the disparity between foreground pixels and line segment pixels, while allowing larger disparities between background pixels and line segment pixels.

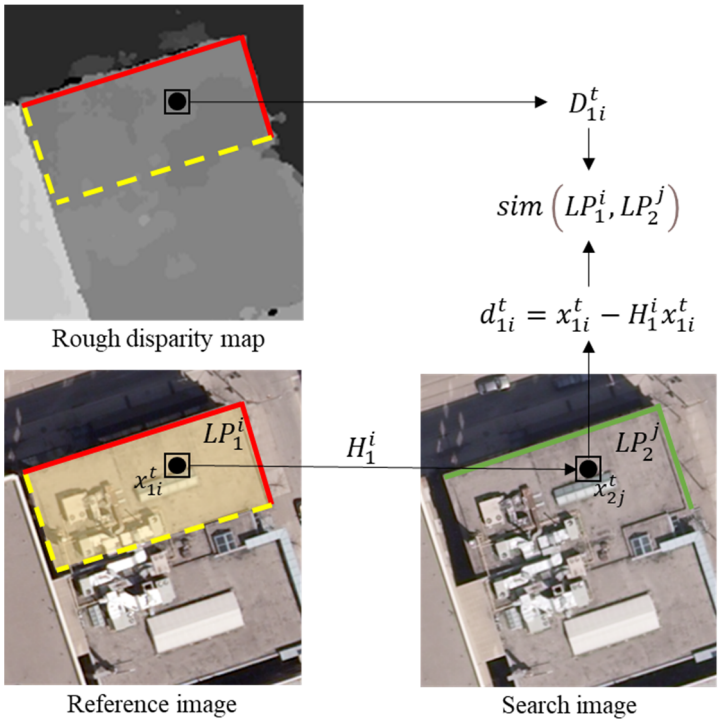


Figure 4. Calculation of similarity of LP_1^i to its candidate matches LP_2^j, taking a pixel x_{1i}^t in the impact region as an example.

3.2.1. Direction of Line Segments in Discontinuous Areas

Line segments on building edges can be extracted from the matched line segments using a line segment buffer method [26]. The process involves establishing a buffer around each line segment, comparing the disparities on both sides of the buffer using the grayscale-weighted median, and selecting line segments with discontinuous depths. However, the fixed width of the buffer does not consider the morphology and size of buildings, which may result in pixels from different planes being included within the buffer.

The impact region generated in the line pair matching contains more semantic information than the general buffer, which alleviates cases where the buffer contains multiple non-coplanar regions. Therefore, as shown in Figure 5, the impact region is used as one

side of the buffer and a certain distance is extended vertically on the other side of the line segment to create the other side of the buffer. Line segments with a disparity difference that exceeds the threshold $thre_D$ between the two buffer zones are regarded as belonging to a discontinuous area. The buffer zone with a larger disparity is treated as the foreground buffer zone, while that with a smaller disparity is considered the background buffer zone. The direction of a line segment is defined such that the foreground buffer is located on the right side of the line segment.



Figure 5. Determination of line segment buffer. A line pair formed by the red line segment and the yellow line segment is located on the edge of the building roof. For the red line segment, the buffer zone, denoted as R_1, is the impact region of the line pair. Buffer zone R_2 is formed by extending R_1 in the direction perpendicular to the line segment by an additional 20 pixels. R_1 and R_2 are the foreground and background buffer zones, respectively.

The line segment direction, the disparities of the two buffer zones, and the aggregation paths are used to determine the foreground and background along the line segment during the cost aggregation. In SGM, the aggregation cost is obtained by summing the one-dimensional path costs calculated for each pixel. Consider the example of four-directional SGM cost aggregation: the aggregation path direction θ_r is 0° from left to right, 180° from right to left, 90° from top to bottom, and 270° from bottom to top. The direction of a line segment θ_l is determined as $\theta_l = \arctan\left(\frac{y_2 - y_1}{x_2 - x_1}\right)$, where (x_1, y_1) and (x_2, y_2) are the coordinates of the two endpoints of the line segment. The direction $\theta_l \in \left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$ is ambiguous. To set the direction of the line segment based on the principle that the foreground buffer zone is on the right side of the line segment, the range of θ_l changes to $(-\pi, \pi)$. In the algorithm, if $\sin(\theta_l - \theta_r) > 0$, the aggregation path passes from the background buffer zone to the foreground buffer zone. Conversely, if $\sin(\theta_l - \theta_r) < 0$, the aggregation path passes from the foreground buffer zone to the background buffer zone.

3.2.2. Adaptive Guiding Parameter for Cost Propagation

The intersection points of line segments and epipolar lines are added into the dense matching as disparity control points. In the neighborhood of a line segment in a discontinuous area, there are notable disparity differences between the background pixels and the line segment pixels. Conversely, the foreground pixels tend to exhibit similar disparities to the line segment pixels. Hence, a guiding parameter is proposed based on the line segment pixel disparity and the relative positions of the buffer zones on the two sides of the line segment. As shown in Figure 6, this parameter constrains the cost propagation among neighboring pixels along the line segment, which restricts the cost propagation from line segment pixels (in red) to the background pixels (in green), and encourages the cost propagation from line segment pixels to the foreground pixels (in yellow).

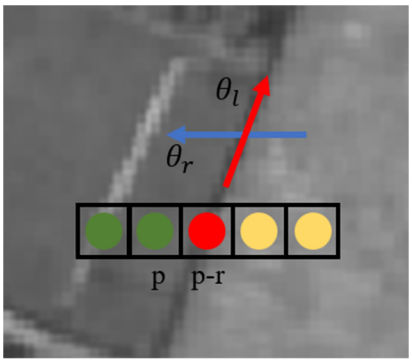


Figure 6. Illustration of cost aggregation on a line segment. The red point is a pixel sampled along the line segment, while the green and yellow pixels represent background and foreground pixels, respectively. The blue arrow indicates the direction of the aggregation paths, while the red arrow represents the direction of the line segment.

Incorporating the guiding parameter changes Equation (3) to Equation (4), which leverages the disparities of line segment pixels to constrain the propagation among neighboring pixels along the line segment.

$$\begin{cases} L_r(p,d) = c(p,d) + T(p-r,d)min\begin{Bmatrix} L_r(p-r,d) \\ L_r(p-r,d-1)+P_1 \\ L_r(p-r,d+1)+P_1 \\ \min\limits_i L_r(p-r,i)+P_2 \end{Bmatrix} - T_b(p-r,d)\min\limits_i L_r(p-r,i) \\ T(p-r,d) = max(\frac{|d-d_L|+1}{thre_D+1},\varepsilon)^{\sin(\theta_l-\theta_r)P} \\ T_b(p-r,d) = min(\varepsilon^{\sin\theta P},(thre_D+1)^{\sin(\theta_l-\theta_r)P}) \end{cases}, \quad (4)$$

where $T(p-r,d)$ is a guiding parameter that corrects the aggregation cost of neighboring pixels by comparing the current computed disparity d with the prior disparity d_L of the line segment pixel. The term $\sin(\theta_l-\theta_r)$ appears in the exponential term of the function and adaptively reflects the direction of the aggregation path, employing different penalty strategies accordingly. When the current pixel on the aggregation path is a background pixel and the previous pixel is a line segment pixel, $\sin(\theta_l-\theta_r)<0$. According to the assumption that the disparity difference between the background pixel and the line segment pixel is relatively large, if the disparity difference between the current pixel and the line segment pixel is less than the discontinuity threshold $thre_D$, the guiding parameter $T(p-r,d)>1$, and this situation will be suppressed. Conversely, $T(p-r,d)<1$ will encourage this situation. When the current pixel on the aggregation path is a foreground pixel and the previous pixel is a line segment pixel, then $\sin(\theta_l-\theta_r)>0$. If the disparities of the current pixel and the line segment pixel are similar, the guiding parameter $T(p-r,d)<1$, which will encourage this situation; otherwise, $T(p-r,d)>1$ and the situation will be suppressed. Here, ε represents a truncation threshold, typically set to 0.5, which ensures that $T(p-r,d)$ falls within a suitable range ($T(p-r,d)\in[0.5,2]$). The parameter P represents the confidence of the line segment match. A smaller degree of confidence makes $T(p-r,d)$ closer to 1, and the penalty effect will be reduced, thus avoiding the mismatch caused by rigidly adding the constraints into the dense matching. In this study, the similarity measure $sim\left(LP_1^i, LP_2^j\right)$ in the line segment matching is regarded as the confidence. The term $T_b(p-r,d)$ is introduced to prevent negative values in the aggregation cost when $T(p-r,d)<1$.

4. Experimental Results and Analysis

In this section, the experimental datasets involved in this study are first introduced, and then the performance of the proposed line segment matching method is evaluated.

On this basis, the dense matching results before and after adding the line segment-based constraints are compared to demonstrate the effectiveness of line segments in improving dense matching performance.

4.1. Experimental Datasets

Experiments are conducted by using satellite imagery and aerial imagery with a focus on urban buildings, as shown in Figure 7. The satellite imagery was obtained from the US3D dataset provided by Johns Hopkins University [34], and consists of WorldView-3 satellite images with a ground sampling distance (GSD) of approximately 0.3 m. The provided stereo image pairs have undergone epipolar-based rectification. Each image block is a pansharpened fusion image with dimensions of 1024 × 1024 pixels. Ground truth disparity maps generated from LiDAR data are also provided. The aerial dataset used in the experiments is the Vaihingen aerial dataset provided by ISPRS [35]. The images were captured using the Intergraph/ZI DMC camera, with a GSD of 0.08 m. The dataset includes a high-resolution Digital Surface Model (DSM) with a grid of 25 cm, obtained through interpolation of airborne LiDAR point clouds, which serves as the ground truth. A pair of image blocks with a size of 4980 × 4300 pixels, mainly containing urban building areas, is cropped from the image pairs for the experiments.

4.2. Evaluation of Line Segment Matching

The proposed line segment matching method is evaluated on the pair of satellite images and pair of aerial images shown in Figure 7a,c, and is compared with the state-of-the-art line segment matching methods MSLD [28] and LJL [31]. There are few line segments in areas with discontinuous disparity in image pair 2, which is not conducive to distinguishing the performance between the two line segment matching methods. Thus, only image pair 1 is used to conduct the line segment matching experiment. Using our approach, the similarity threshold for line pair matching is empirically determined to be 0.25. The LSD algorithm [32] is adopted for all matching methods, ensuring that the input line segments of our proposed algorithm are the same as those for MSLD and LJL.

The number of correct matches (NCM), matching precision (MP), and execution time (ET) are used to evaluate the performance of the line segment matching algorithms. The MP represents the ratio of NCM to the total number of matched line segments. Table 1 presents the line segment matching results of the proposed method and the compared methods.

Table 1. Experimental results of line segment matching.

Dataset	NCM			MP (%)			ET (s)		
	MSLD	**LJL**	**Proposed**	**MSLD**	**LJL**	**Proposed**	**MSLD**	**LJL**	**Proposed**
Pair 1	97	380	362	83.5	83.5	97.5	0.7	225	3
Pair 3	1814	2536	2528	94.6	93.3	99.4	11.5	3910	52

It can be seen from Table 1 that although MSLD has the highest time efficiency, it obtains far fewer correct matches than LJL and the proposed method. In addition, its MP values are comparable to LJL and far lower than the proposed method. The difference in the performance of LJL and the proposed method in terms of NCM is very small. LJL achieves a slightly higher NCM value because it extends the individual line segments using local homography constraints after matching based on local line pair descriptors. In terms of MP, our method significantly outperforms LJL, achieving a high level of over 97% in both types of imagery. LJL exhibits poor discrimination between line segments in areas with discontinuous disparity and those that are close and parallel, leading to erroneous matches, as indicated in the enlarged local areas in Figures 8 and 9. Fortunately, the proposed method reduces the ambiguity through the use of the disparity consistency-based similarity measure. In addition, the MP values of MSLD and LJL on image pair 1 is about 10% lower

than those on image pair 3. This is mainly because the images in pair 3 cover a large scene range. In addition to some multi-story buildings with large parallax changes, there are also a large number of low targets on the images. The line segments on these targets are easier to match. The images in pair 1 correspond to a small urban building area, and it is more difficult to match the line segments in the area. Compared with MSLD and LJL, the MP difference of the proposed method on the two pairs of images is less than 2%. It shows that the proposed method is more robust to different scenarios than MSLD and LJL.

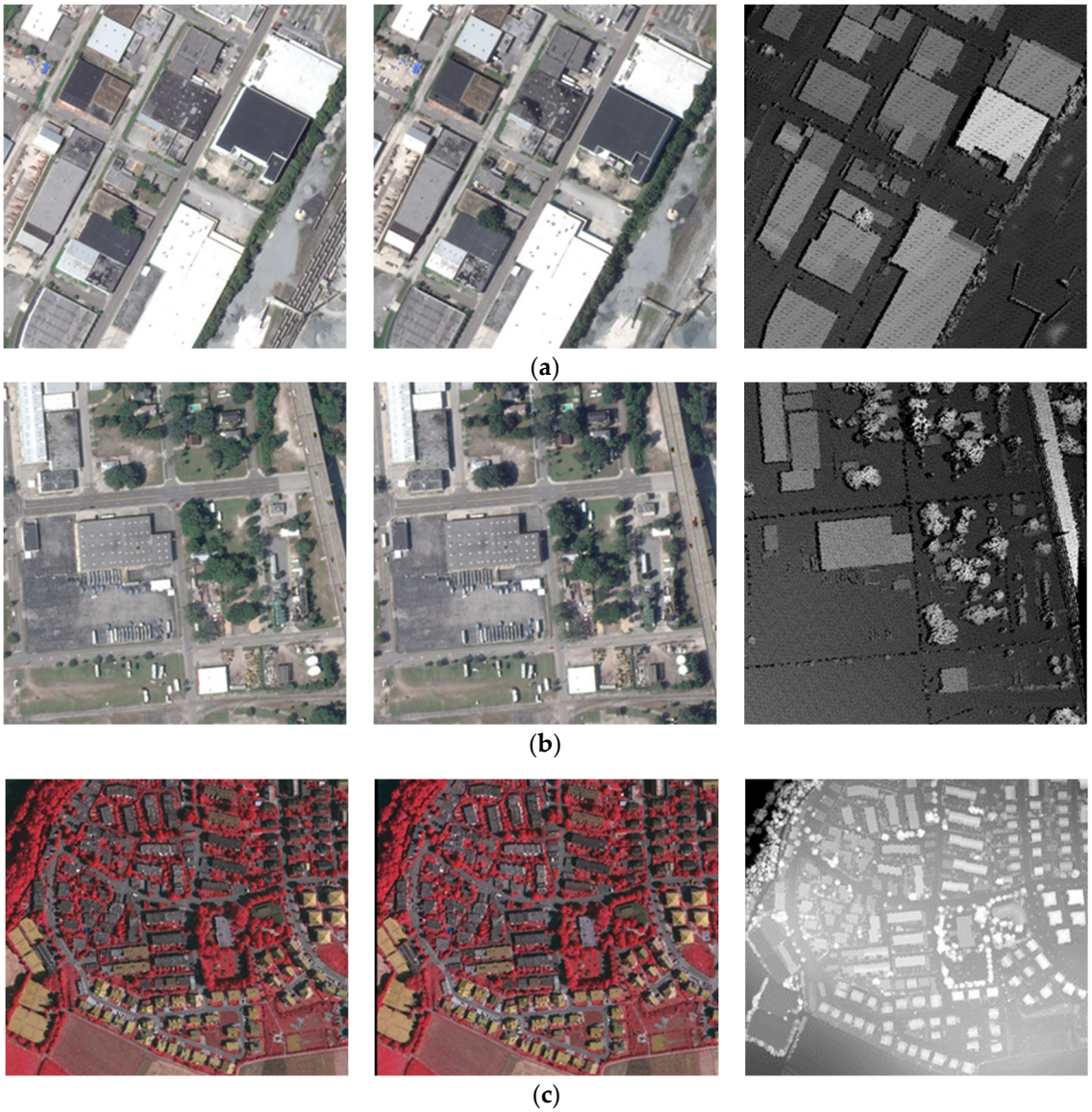


Figure 7. Experimental datasets. (**a**,**b**) show two satellite image pairs with the corresponding LiDAR-based ground truth disparity maps, named image pair 1 and image pair 2, respectively. (**c**) shows the aerial image pair with the corresponding ground truth DSM, named image pair 3.

In terms of ET, because our method relies on the rough disparity map, the time statistics for our method include the time required for the preceding hierarchical dense matching. All methods were executed in the same hardware and software environment (Intel Core i7 2.50 GHz Windows 10 PC). The proposed method exhibits excellent time efficiency, with an average ET that is around 74 times faster than that of LJL. According to the NCM, MP, and ET indicators, the overall performance of the proposed line segment matching method is far better than that of the MSLD and LJL methods.

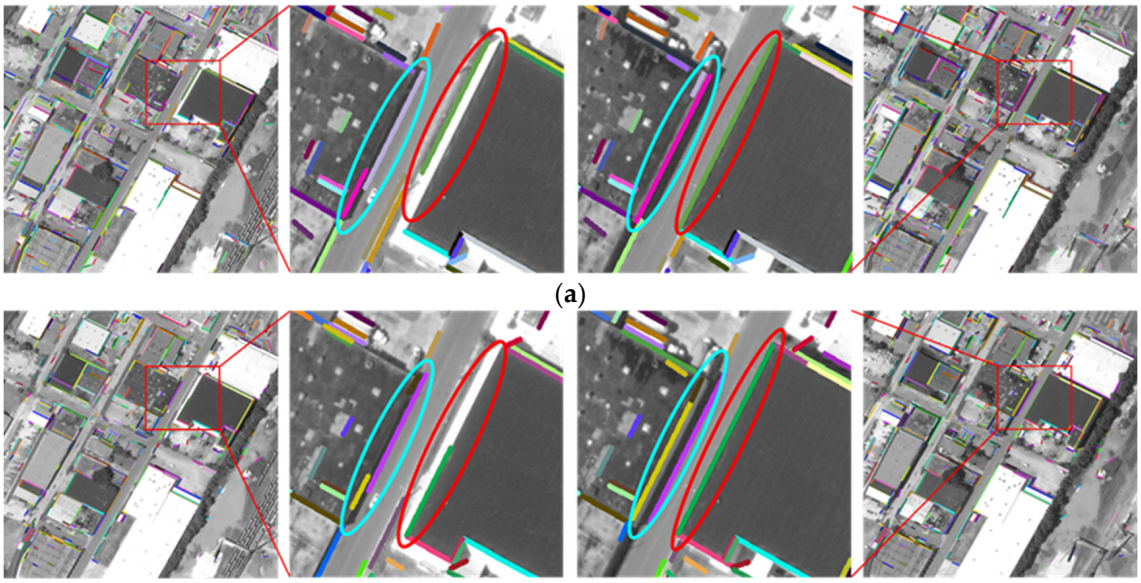


Figure 8. Line segment matches on the satellite images obtained by (**a**) LJL and (**b**) the proposed method.

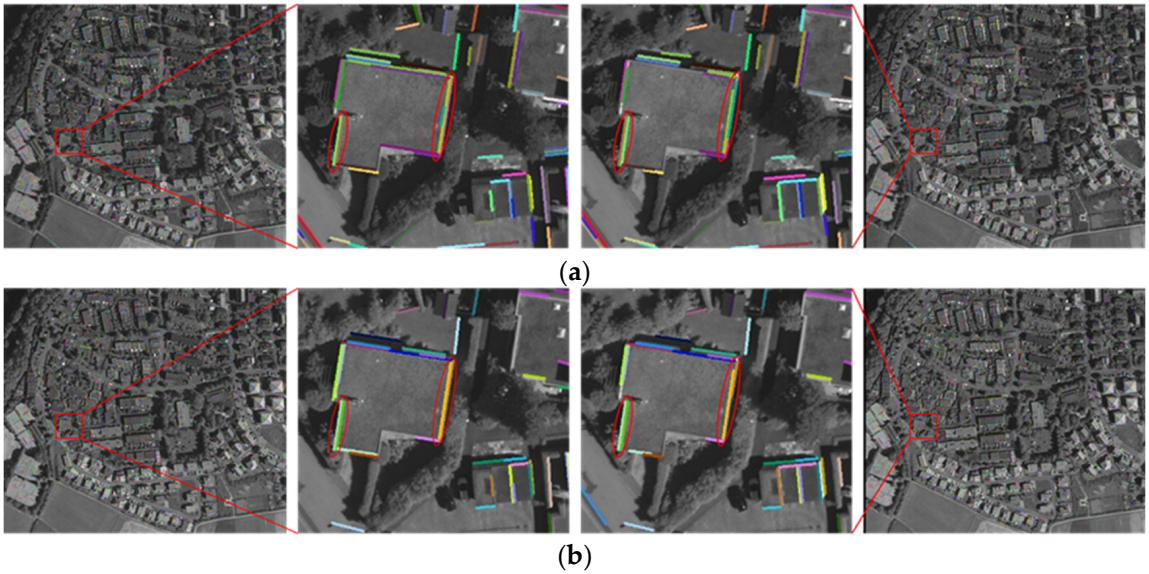


Figure 9. Line segment matches on the aerial images obtained by (**a**) LJL and (**b**) the proposed method.

4.3. Evaluation of Line Segment Constrained Dense Matching

This section examines the effectiveness of the proposed line segment constrained dense matching method on satellite stereo images and aerial stereo images. A comprehensive analysis is conducted and quantitative and qualitative results are presented.

4.3.1. Performance on Satellite Imagery

The proposed method was compared with the SGM, tSGM, and tSGM-pyr methods. tSGM-pyr only uses the pyramid hierarchical matching strategy in tSGM, but does not improve the penalty parameters for the matching of edge pixels in tSGM. All algorithms employed the Census transform based on a 5×5 neighborhood for cost computation, and the eight-directional path aggregation was performed using P_1 and P_2 values of 24 and 96,

respectively. In the tSGM algorithm, the Canny operator [36] was used to extract edges, and P_2 was adaptively updated. If a pixel exhibited a positive response to edge detection, its penalty parameter P_2 was set to 27. To ensure a fair comparison, line segment pixels were considered to be edge pixels. A consistent postprocessing procedure was applied to all disparity map results, including the intermediate levels of tSGM hierarchical matching. A 3×3 median filter was employed to achieve smoother results, followed by a consistency check and the removal of regions containing fewer than 50 pixels.

A line segment buffer zone with a width of five pixels was defined and quantitative analysis was performed within the buffer zone following the approach described in [26]. The invalid pixel error (IPE), bad pixel error (BPE), and total error (TE) metrics are adopted [37]. The IPE is the percentage of invalid pixels in the estimated disparity map that are valid pixels in the ground truth disparity map. The BPE is defined as the percentage of valid pixels in the estimated disparity map for which the absolute difference between the estimated disparity and the ground truth disparity exceeds a threshold (set to two pixels in this experiment). The TE represents the sum of all defined errors. In the assessment of building edge accuracy, the IPE measures the inward displacement of building edges. However, it does not reflect the outward expansion caused by the mismatch of building edges. Therefore, an additional metric called the occluding pixel error (OPE) is included in the evaluation. The OPE represents the percentage of valid pixels in the estimated disparity map that are invalid pixels in the ground truth disparity map.

The results in Table 2 show that the proposed method achieves the best overall performance from the perspective of TE, while the SGM method gives the worst results. Although SGM performs very well in terms of OPE and BPE, it has the highest IPE. This is because there are numerous invalid pixels on buildings that have not been matched.

Table 2. Dense matching performance comparison on satellite images.

Method	IPE (%)	OPE (%)	BPE (%)	TE (%)
SGM	57.80	2.49	3.91	64.20
tSGM-pyr	7.67	12.77	19.16	39.60
tSGM	6.87	12.64	18.71	38.22
Proposed	8.14	10.17	15.96	34.26

Compared with the tSGM-pyr and tSGM methods, the proposed method achieves the best performance in all evaluation metrics except for IPE. The reason that the proposed method does not score as well in IPE is that the ground truth disparity map is generated from LiDAR point clouds with low resolution. Consequently, the accuracy of the ground truth disparity map is low around the building edges. There are some invalid pixels near the edges on the ground truth disparity map, while the proposed method improves the matching rate and reduces the number of invalid pixels near the edges. Therefore, it is not conducive to obtaining a better IPE that accurately reflects the performance of the proposed method. Although the proposed method does not give the highest IPE, it achieves the best performance near the building edges, as shown in Figures 10 and 11.

Figures 10 and 11 show that the SGM algorithm produces disparity maps with a large number of invalid pixels. These are caused by the presence of numerous weak-textured areas in the satellite imagery and differences introduced by the acquisition time. Relatively, the hierarchical dense matching methods yield better disparity map results by narrowing the disparity range and increasing the certainty of disparity selection.

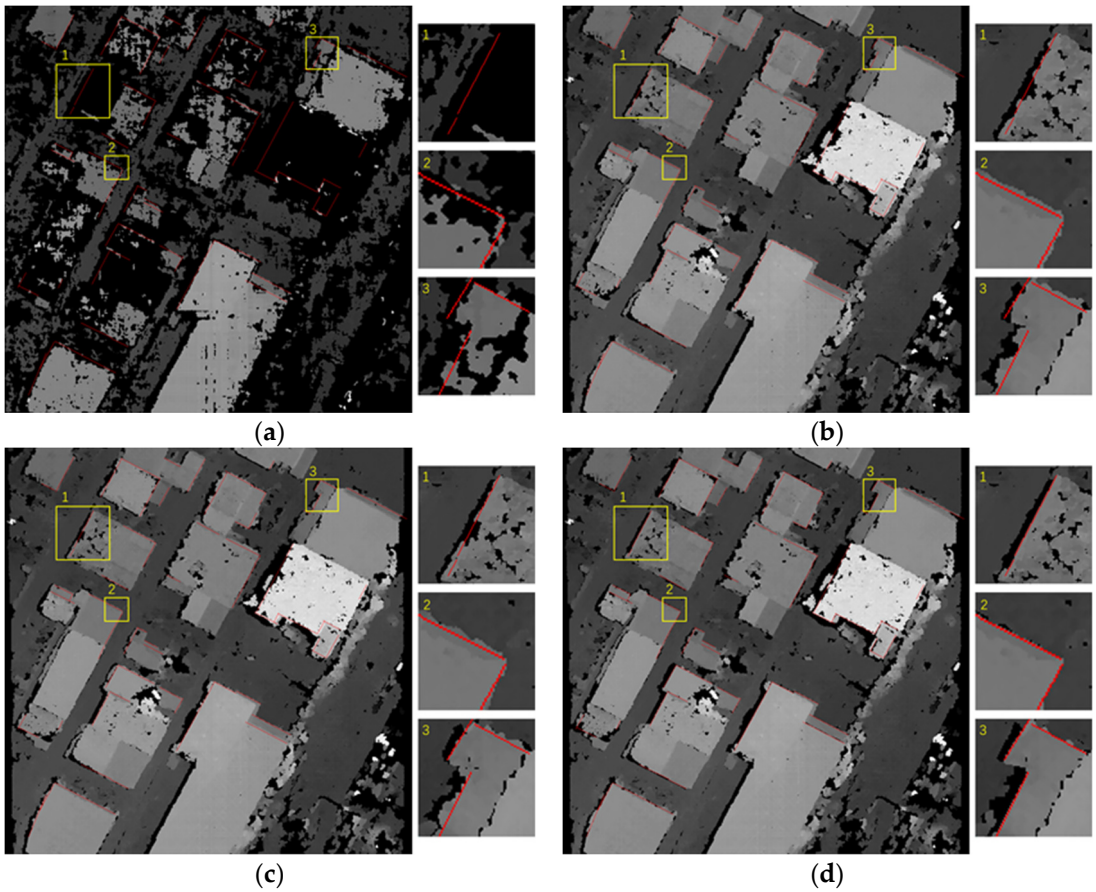


Figure 10. Disparity maps from satellite image pair 1 using (**a**) SGM, (**b**) tSGM-pyr, (**c**) tSGM, and (**d**) the proposed method. The red lines indicate the matched line segments.

Among the hierarchical dense matching methods, the tSGM-pyr method produces numerous erroneous matches near building edges. Even after performing left–right consistency checks, the algorithm fails to address the distortions and inward recesses in building edges. Although the edge optimization strategy of tSGM based on smoothness parameters can partially mitigate inward recesses, its effectiveness is limited and new erroneous matches are sometimes introduced, as shown in the rectangular region "3" in Figure 11. Conversely, the disparity maps generated by the proposed method display a significant improvement in the dense matching results as they effectively compensate for the inward recesses and preserve the building edges.

Local regions "2" and "3" in Figure 10 and local regions "1" and "3" in Figure 11 show examples of building edge expansion being refined by the proposed method. At these edges, the disparity map obtained by the edge refinement strategy in tSGM still exhibits expansion, while the disparity map generated by the proposed method almost coincides with the line segments, and the edge features are well maintained.

The experimental results demonstrate that the proposed method not only fills in missing building edges, but also reduces the outward expansion of building edges, thereby improving the accuracy of 3D reconstruction. This is not simply achieved by considering the disparity values of line segment pixels, but also requires the areas on either side of the building line segments to correspond to disparity-consistent areas and disparity-transition areas, thereby allowing different penalty strategies to be employed.

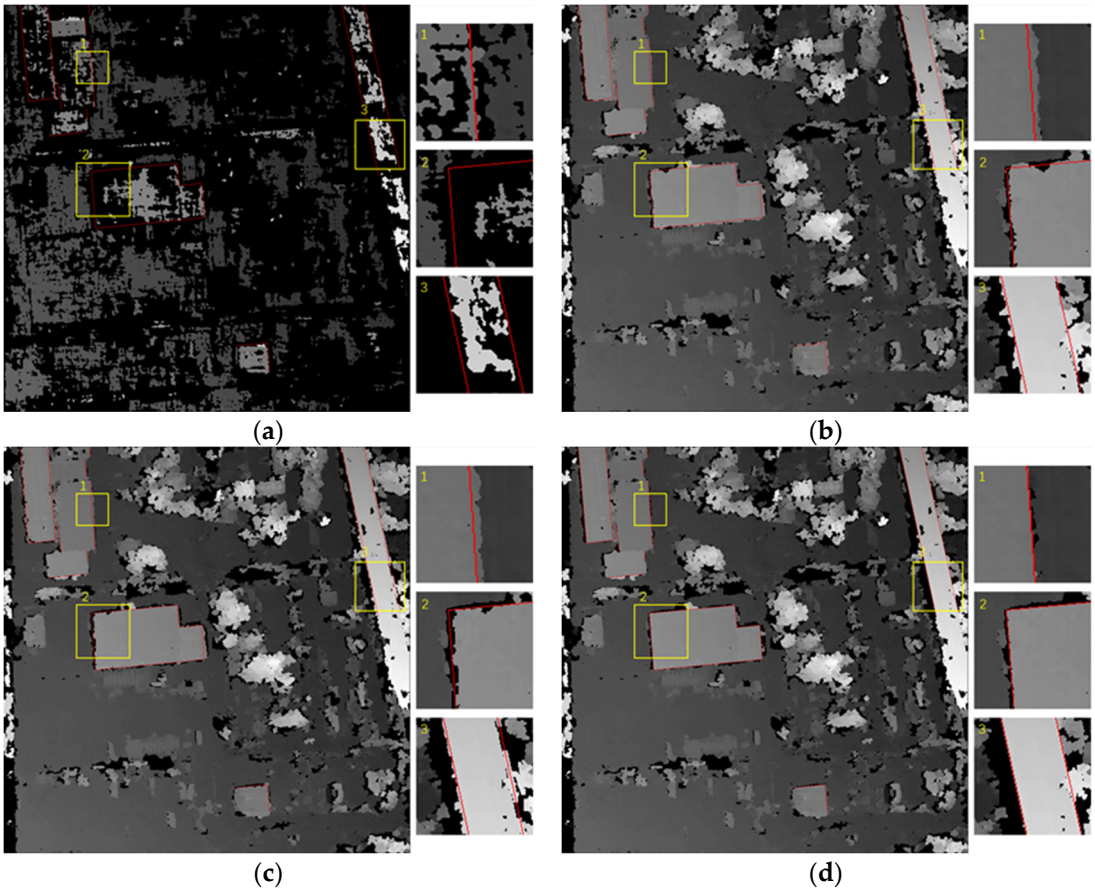


Figure 11. Disparity maps from satellite image pair 2 using (**a**) SGM, (**b**) tSGM-pyr, (**c**) tSGM, and (**d**) the proposed method. The red lines indicate the matched line segments.

4.3.2. Performance on Aerial Imagery

The aerial imagery dataset does not include ground truth disparity maps, so a DSM interpolated from LiDAR point clouds is taken as the ground truth. Dense matching was performed on the aerial stereo images, and erroneous points were removed using left–right consistency checks. The corresponding points were then projected into a dense point cloud and interpolated to generate a DSM. The DSM generated by the proposed method was compared with those generated by the SGM and tSGM methods. The parameters were the same as for Section 4.3.1. The root mean square error (RMSE) and BPE are used as evaluation criteria, and the same buffer pixel statistics method as in Section 4.3.1 is employed with a buffer radius of eight pixels. The evaluation results are presented in Table 3, where the proposed method achieves the best performance in terms of both RMSE and BPE.

Table 3. Dense matching performance comparison on aerial images.

Method	RMSE (m)	BPE (%)
SGM	1.91	25.80
tSGM-pyr	1.88	25.10
tSGM	1.73	22.24
Proposed	1.65	20.51

Two representative examples are shown in Figures 12 and 13. In these examples, the buildings reconstructed using our method exhibit regular and well-aligned contours. The edge features are relatively pronounced, and there are fewer cases in which the roofs extend to the ground. Notably, in the detailed areas, some edges appear even more regular than in

the DSM from LiDAR point clouds; this is due to the lower sampling rate of the LiDAR point clouds. Furthermore, cross-sectional analysis is conducted by selecting cross-sectional profiles represented by red lines in Figures 12b and 13b of the LiDAR DSM. The results reveal that our method achieves the minimum RMSE for the pixels along the cross-sections, indicating its superior performance.

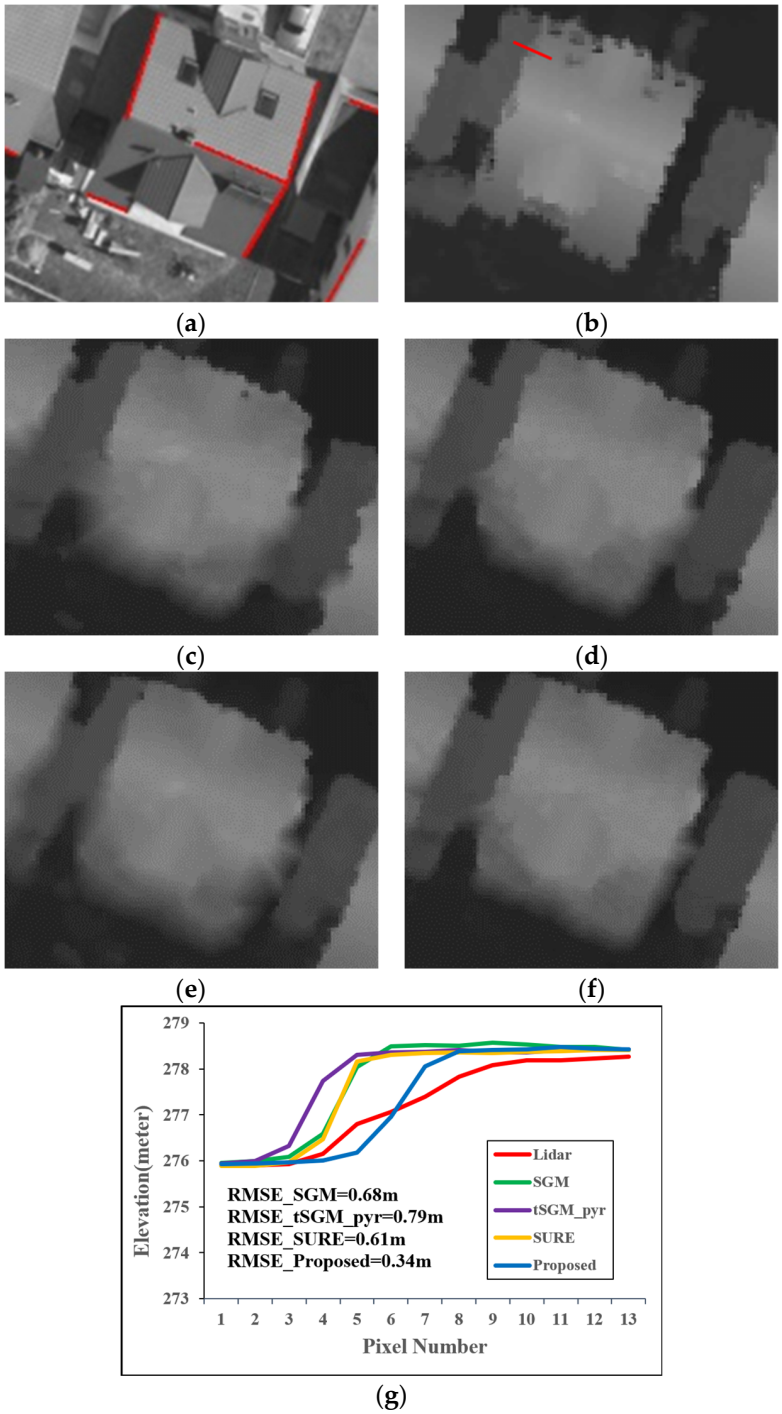


Figure 12. Example 1 of DSMs from aerial images using different methods. (**a**) Original image, (**b**) ground truth DSM from LiDAR point clouds, (**c**–**f**) DSMs generated by SGM, tSGM-pyr, tSGM, and the proposed method, respectively, and (**g**) cross-sectional analysis. The red lines in (**a**) are the matched line segments. The red line in (**b**) indicates the range of cross-sectional analysis.

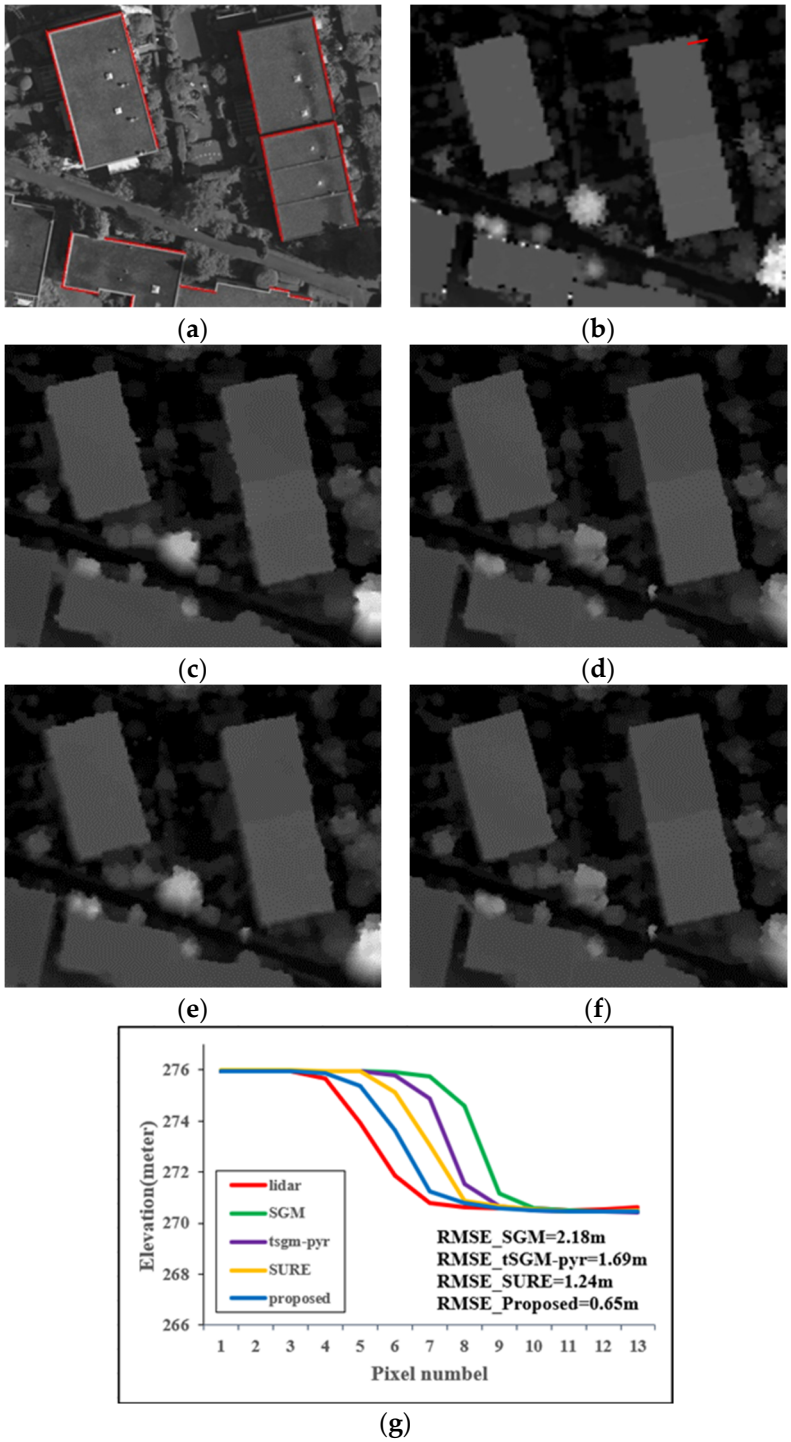


Figure 13. Example 2 of DSMs from aerial images using different methods. (**a**) Original image, (**b**) ground truth DSM from LiDAR point clouds, (**c**–**f**) DSMs generated by SGM, tSGM-pyr, tSGM, and the proposed method, respectively, and (**g**) cross-sectional analysis. The red lines in (**a**) are the matched line segments. The red line in (**b**) indicates the range of cross-sectional analysis.

Based on comparative analysis using both satellite and aerial images, the proposed method demonstrates superior performance in reconstructing building edges where line segments have been successfully matched, surpassing the performance of the SGM, tSGM, and tSGM-pyr methods. Moreover, the proposed line segment-constrained refinement strategy can be seamlessly integrated into other SGM-like techniques, making it highly versatile.

5. Conclusions

This study has considered edge-preserving dense matching. A novel efficient and reliable line segment matching method based on a disparity consistency-based similarity measure is proposed, and then a line segment constrained dense matching method that sets different penalty strategies by distinguishing whether the disparity transition pixels are in the background or foreground is introduced. Experimental results demonstrate that the proposed line segment matching method achieves a high MP (greater than 97%) and significant advantages in terms of time efficiency. Analysis of the disparity maps from satellite imagery and DSMs from aerial images reveals that the proposed line segment constrained dense matching method outperforms SGM and its variants, especially in areas near building edges. Furthermore, the proposed line segment constrained method can be integrated into other dense matching techniques. One limitation of the proposed method is that it relies on individual line segments to improve the dense matching performance. Future work will attempt to establish the topological relationship between line segments to construct constraints with a wider coverage, which should improve the overall dense matching performance.

Author Contributions: Conceptualization, Y.Y. and M.C.; methodology, Y.Y., T.F., and M.C.; software, Y.Y.; validation, T.F., W.L., and M.C.; formal analysis, B.X.; investigation, X.G.; resources, H.H.; data curation, Z.Z.; writing—original draft preparation, Y.Y.; writing—review and editing, M.C. and T.F.; visualization, Y.Y.; supervision, M.C.; funding acquisition, M.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China (No. 41971411 and No. 42371445), the Sichuan Science and Technology Program (No. 2023NSFSC0247), and the Open Fund of Key Laboratory of Urban Land Resources Monitoring and Simulation, Ministry of Natural Resources (No. KF-2021-06-012).

Data Availability Statement: The data presented in this study are openly available in reference number [34,35].

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Braun, C.; Kolbe, T.H.; Lang, F.; Schickler, W.; Steinhage, V.; Cremers, A.B.; Förstner, W.; Plümer, L. Models for photogrammetric building reconstruction. *Comput. Graph.* **1995**, *19*, 109–118. [CrossRef]
2. Brenner, C. Building reconstruction from images and laser scanning. *Int. J. Appl. Earth Obs. Geoinf.* **2005**, *6*, 187–198. [CrossRef]
3. Haala, N.; Kada, M. An update on automatic 3D building reconstruction. *ISPRS J. Photogramm. Remote Sens.* **2010**, *65*, 570–580. [CrossRef]
4. Yu, D.; Ji, S.; Liu, J.; Wei, S. Automatic 3D building reconstruction from multi-view aerial images with deep learning. *ISPRS J. Photogramm. Remote Sens.* **2021**, *171*, 155–170. [CrossRef]
5. Gupta, G.; Balasubramanian, R.; Rawat, M.; Bhargava, R.; Gopala Krishna, B. Stereo matching for 3d building reconstruction. In Proceedings of the Advances in Computing, Communication and Control: International Conference, ICAC3 2011, Mumbai, India, 28–29 January 2011; pp. 522–529.
6. Hamzah, R.A.; Kadmin, A.F.; Hamid, M.S.; Ghani, S.F.A.; Ibrahim, H. Improvement of stereo matching algorithm for 3D surface reconstruction. *Signal Process. Image Commun.* **2018**, *65*, 165–172. [CrossRef]
7. Kendall, A.; Martirosyan, H.; Dasgupta, S.; Henry, P.; Kennedy, R.; Bachrach, A.; Bry, A. End-to-end learning of geometry and context for deep stereo regression. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 66–75.
8. Scharstein, D.; Szeliski, R. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *Int. J. Comput. Vis.* **2002**, *47*, 7–42. [CrossRef]
9. Hirschmuller, H. Accurate and efficient stereo processing by semi-global matching and mutual information. In Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), San Diego, CA, USA, 20–25 June 2005; pp. 807–814.
10. Hirschmuller, H. Stereo processing by semiglobal matching and mutual information. *IEEE Trans. Pattern Anal. Mach. Intell.* **2007**, *30*, 328–341. [CrossRef] [PubMed]
11. Facciolo, G.; De Franchis, C.; Meinhardt, E. MGM: A significantly more global matching for stereovision. In Proceedings of the BMVC 2015, Swansea, UK, 7–10 September 2015.

12. Patil, S.; Prakash, T.; Comandur, B.; Kak, A. A comparative evaluation of SGM variants (including a new variant, tMGM) for dense stereo matching. *arXiv* **2019**, arXiv:1911.09800.
13. Scharstein, D.; Taniai, T.; Sinha, S.N. Semi-global stereo matching with surface orientation priors. In Proceedings of the 2017 International Conference on 3D Vision (3DV), Qingdao, China, 10–12 October 2017; pp. 215–224.
14. Rothermel, M.; Wenzel, K.; Fritsch, D.; Haala, N. SURE: Photogrammetric surface reconstruction from imagery. In Proceedings of the LC3D Workshop, Berlin, Germany, 4–5 December 2012.
15. Chuang, T.-Y.; Ting, H.-W.; Jaw, J.-J. Dense stereo matching with edge-constrained penalty tuning. *IEEE Geosci. Remote Sens. Lett.* **2018**, *15*, 664–668. [CrossRef]
16. Kim, K.-R.; Kim, C.-S. Adaptive smoothness constraints for efficient stereo matching using texture and edge information. In Proceedings of the 2016 IEEE International Conference on Image Processing (ICIP), Phoenix, AZ, USA, 25–28 September 2016; pp. 3429–3433.
17. Kim, G.-B.; Chung, S.-C. An accurate and robust stereo matching algorithm with variable windows for 3D measurements. *Mechatronics* **2004**, *14*, 715–735. [CrossRef]
18. Xu, Y.; Zhao, Y.; Ji, M. Local stereo matching with adaptive shape support window based cost aggregation. *Appl. Opt.* **2014**, *53*, 6885–6892. [CrossRef] [PubMed]
19. Zhu, S.; Yan, L. Local stereo matching algorithm with efficient matching cost and adaptive guided image filter. *Vis. Comput.* **2017**, *33*, 1087–1102. [CrossRef]
20. Chen, D.; Ardabilian, M.; Chen, L. A Novel Trilateral Filter based Adaptive Support Weight Method for Stereo Matching. In Proceedings of the BMVC, Bristol, UK, 9–13 September 2013.
21. Hosni, A.; Rhemann, C.; Bleyer, M.; Rother, C.; Gelautz, M. Fast cost-volume filtering for visual correspondence and beyond. *IEEE Trans. Pattern Anal. Mach. Intell.* **2012**, *35*, 504–511. [CrossRef] [PubMed]
22. Tatar, N.; Arefi, H.; Hahn, M. High-resolution satellite stereo matching by object-based semiglobal matching and iterative guided edge-preserving filter. *IEEE Geosci. Remote Sens. Lett.* **2020**, *18*, 1841–1845. [CrossRef]
23. Cheng, F.; Zhang, H.; Yuan, D.; Sun, M. Stereo matching by using the global edge constraint. *Neurocomputing* **2014**, *131*, 217–226. [CrossRef]
24. Huang, X.; Zhang, Y.; Yue, Z. Image-Guided Non-Local Dense Matching with Three-Steps Optimization. *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.* **2016**, *3*, 67–74. [CrossRef]
25. Jiao, J.; Wang, R.; Wang, W.; Li, D.; Gao, W. Color image-guided boundary-inconsistent region refinement for stereo matching. *IEEE Trans. Circuits Syst. Video Technol.* **2015**, *27*, 1155–1159. [CrossRef]
26. Qin, R.; Chen, M.; Huang, X.; Hu, K. Disparity refinement in depth discontinuity using robustly matched straight lines for digital surface model generation. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2018**, *12*, 174–185. [CrossRef]
27. Schmid, C.; Zisserman, A. Automatic line matching across views. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Juan, PR, USA, 17–19 June 1997; pp. 666–671.
28. Wang, Z.; Wu, F.; Hu, Z. MSLD: A robust descriptor for line matching. *Pattern Recognit.* **2009**, *42*, 941–953. [CrossRef]
29. Chen, M.; Li, W.; Fang, T.; Zhu, Q.; Xu, B.; Hu, H.; Ge, X. An adaptive feature region-based line segment matching method for viewpoint-changed images with discontinuous parallax and poor textures. *Int. J. Appl. Earth Obs. Geoinf.* **2023**, *117*, 103209. [CrossRef]
30. Chen, M.; Yan, S.; Qin, R.; Zhao, X.; Fang, T.; Zhu, Q.; Ge, X. Hierarchical line segment matching for wide-baseline images via exploiting viewpoint robust local structure and geometric constraints. *ISPRS J. Photogramm. Remote Sens.* **2021**, *181*, 48–66. [CrossRef]
31. Li, K.; Yao, J. Line segment matching and reconstruction via exploiting coplanar cues. *ISPRS J. Photogramm. Remote Sens.* **2017**, *125*, 33–49. [CrossRef]
32. Von Gioi, R.G.; Jakubowicz, J.; Morel, J.-M.; Randall, G. LSD: A fast line segment detector with a false detection control. *IEEE Trans. Pattern Anal. Mach. Intell.* **2010**, *32*, 722–732. [CrossRef] [PubMed]
33. Hartley, R.; Zisserman, A. *Multiple View Geometry in Computer Vision*; Cambridge University Press: Cambridge, UK, 2003.
34. Bosch, M.; Foster, K.; Christie, G.; Wang, S.; Hager, G.D.; Brown, M. Semantic stereo for incidental satellite images. In Proceedings of the 2019 IEEE Winter Conference on Applications of Computer Vision (WACV), Waikoloa Village, HI, USA, 7–11 January 2019; pp. 1524–1532.
35. Available online: https://www.isprs.org/education/benchmarks/UrbanSemLab/detection-and-reconstruction.aspx (accessed on 11 August 2022).
36. Canny, J. A computational approach to edge detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **1986**, *6*, 679–698. [CrossRef]
37. Scharstein, D.; Szeliski, R. Middlebury Online Stereo Evaluation. 2023. Available online: https://vision.middlebury.edu/stereo/ (accessed on 22 August 2022).

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.