*remote sensing*

MDPI

*Article*

# A Multi-View Dense Image Matching Method for High-Resolution Aerial Imagery Based on a Graph Network

**Li Yan, Liang Fei *, Changhai Chen, Zhiyun Ye and Ruixi Zhu**

School of Geodesy and Geomatics, Wuhan University, 129 Luoyu Road, Wuhan 430079, China; lyan@sgg.whu.edu.cn (L.Y.); chench@whu.edu.cn (C.C.); 2008301610260@whu.edu.cn (Z.Y.); 2010301610088@whu.edu.cn (R.Z.)

* Correspondence: lfei@whu.edu.cn; Tel.: +86-27-6877-8166

Academic Editors: Lizhe Wang, Gonzalo Pajares Martinsanz, Xiaofeng Li and Prasad S. Thenkabail
Received: 28 June 2016; Accepted: 22 September 2016; Published: 26 September 2016

**Abstract:** Multi-view dense matching is a crucial process in automatic 3D reconstruction and mapping applications. In this paper, we present a robust and effective multi-view dense matching algorithm for high-resolution aerial images based on a graph network. The overlap ratio and intersection angle between image pairs are used to find candidate stereo pairs and build the graph network. A Coarse-to-Fine strategy based on an improved Semi-Global Matching algorithm is applied for disparity computation across stereo pairs. Based on the constructed graph, point clouds of base views are generated by triangulating all connected image nodes, followed by a fusion process with the average reprojection error as a priority measure. The proposed method was successfully applied in experiments on aerial image test dataset provided by the ISPRS of Vaihingen, Germany and an oblique nadir image block of Zürich, Switzerland, using three kinds of matching configurations. The proposed method was compared to other state-of-art methods, SURE and PhotoScan. The results demonstrate that the proposed method delivers matches at higher completeness, efficiency, and accuracy than the other methods tested; the RMS for average reprojection error reached the sub pixel level and the actual positioning deviation was better than 1.5 GSD.

**Keywords:** high-resolution aerial images; dense image matching (DIM); MVS; SGM; graph network

## 1. Introduction

With the improvement of camera technologies and the rise of new matching approaches, more and more image-based software for 3D modeling and reconstruction have been developed, such as Street Factory, Accute3D, PhotoScan, in which a multi-view dense image matching (DIM) algorithm is one of the most crucial processes. The goal of multi-view dense matching is to extract dense point clouds from multiple images that have known orientated parameters. Image-based surveying and 3D modeling can now deliver point clouds with accuracy comparable to those produced by laser scanning [1] for many terrestrial and aerial applications in a reasonable time. Since the inherent nature of multi-spectral images, rich texture information can be accessed. Moreover, the absolute accuracy of a point cloud can be assessed based on the redundant measurements extracted from imagery. Image-based 3D reconstruction is widely used for 3D modeling, mapping, robotics, and navigation because it is lightweight, convenient, cost effective, and generates textured point clouds comparable to those produced by LiDAR systems.

Dense image matching algorithms can be divided into stereo pair based algorithms [2–5] and multi-view based algorithms [6–12] according to type of processing units. They can also be divided into local and global algorithms according to different optimization strategies [13]. Local (window-based) methods compute disparity at a given point using intensity values with implicit smoothing assumptions, while global methods make explicit smoothness assumptions and solve for a global optimization problem using an energy minimization approach, based on regularized Markov fields (MRFs), graph-cut, dynamic programming, or max-flow methods [1]. The global methods often yield better performance than local methods, but with a more complex algorithm. Semi-global matching (SGM) algorithms [2,3,14] use pixelwise, Mutual Information (MI)-based matching supported by a smoothness constraint expressed as a global cost function, and performs a fast approximation by pathwise optimizations from all directions, thus representing a compromise between performance and complexity. An extensive study of different matching costs [15] showed that a census matching cost is the most robust technique for stereo vision, and performs better for matching local radiometric changes than the MI cost in many real-world applications.

The four state-of-art software; SURE, MicMac, PMVS, and PhotoScan were evaluated in terms of similarities and differences in approach [1]. SURE [5,16] is based on SGM method followed by a fusion step in which redundant depth estimations across single stereo models are merged. Micmac [7,17] implements a coarse-to-fine extension of the maximum-flow image-matching algorithm across multi-resolution and multi-image presented in [18]. PMVS [10] is a region growing method that expands a sparse set of matched key points to nearby pixel correspondences before using visibility constraints to filter out false matches. PhotoScan uses a stereo SGM-like method, a proprietary commercial package. Experiments showed that the SURE and PhotoScan approaches that adopted a SGM or SGM-like method are more accurate and efficient than the PMVS and Micmac approaches.

The stereo DIM algorithm in our implementation is based on the tSGM (SGM with a tube shaped disparity range) algorithm in SURE; it is an improvement on the SGM algorithm in memory demand and processing time. The tSGM algorithm provides more a complete reconstruction of lightly textured objects and objects with a repetitive texture. However, edges are not reconstructed as clearly as in SGM algorithm, particularly in the case of close range photogrammetry. For this reason, a guided median filter was introduced to preserve depth discontinuities instead of the median filter deployed in the current method. Within MVS methods in image space, disparity estimations for single stereo models are typically merged in order to increase the reliability and precision of the final depth maps or point clouds. The original SGM algorithm [3] proposes an orthographic 2.5D projection method to fuse several disparity images from different viewpoints for aerial applications; this method is not applicable for close range and oblique aerial applications because too much information would be lost. SURE [5] presents a method to transform all disparity maps of a base image to the distance maps between the camera center and object points on the optical ray, then fuses them by minimizing the reprojection error. However, the point clouds generated from different base images result in small offsets since they are triangulated separately using different set of images. In our approach, a multi-view DIM point cloud generation method based on a graph network is introduced, integrating a correspondence linking technique as presented in [19]. The main steps of the current method are as follows: First, a graph network is built considering the overlap relationship and intersection angle between adjacent images. Second, epipolar images are generated and a modified tSGM algorithm is applied to obtain dual direction disparity maps across stereo pairs. A multi-view triangulation approach based on the graph network is adopted to calculate a dense point cloud for a single base view. Finally, multi-view dense point clouds are fused with average reprojection error as a priority measure to generate a final point cloud.

The main contribution of this paper are: (1) Proposing a graph network based multi-view DIM method that computes fewer stereo pairs than traditional MVS methods like SURE and PhotoScan, while obtaining better completeness and higher accuracy. The triangulated point clouds from different base images in the proposed method share the same correspondences with no offset and can be seamlessly integrated; (2) We introduce a guided median filter into the tSGM algorithm to preserve depth discontinuities for better performance in difficult areas with sharp discontinuities, weak textures, or repeated textures.

This paper is organized as follows: Section 2 describes the graph network based multi-view DIM algorithm, where Section 2.1 describes the construction of graph network. Section 2.2 presents the improved tSGM algorithm for stereo dense matching. Section 2.3 shows the graph network based triangulation method of base view. Section 2.4 introduces the multi-view point-cloud fusion approach. Section 3 presents experiments and results, followed by the discussion in Section 4. Section 5 draws some conclusions.

## 2. Graph Network Based DIM Methodology

In this section, an effective graph network based approach for multi-view dense matching of high-resolution imagery is presented. The proposed method can be divided into four main steps. In the first step, candidate stereo pairs are selected considering overlap relationships and intersection angle to construct the multi-view graph network. In the second step, epipolar images are generated followed by processing with the modified tSGM algorithm to calculate dual-direction disparity maps. In the third step, a graph network based triangulation is adopted to generate a dense point cloud of base views from all the connected images, including those outside the candidate matches. Redundancy is exploited through the correspondence linking technique to eliminate error and increase the accuracy of triangulation, while enabling assessment of the accuracy of the average reprojection error. Finally, the multi-view point clouds are fused with the average reprojection error as a priority measure. Details of those specific principles and implementation processes are provided in the subsequent sub sections.

### 2.1. Graph Network Construction

Figure 1 illustrates the process for constructing the graph network. Let the target graph network be denoted by $G$ then, nodes represent the available images, arcs represent stereo matching from a start node to end node, and paths represent correspondence linking between two images, as shown in the Figure. All the oriented images are first added to the $G$ as single nodes respectively, while arcs are inserted to $G$ by candidate stereo pairs.

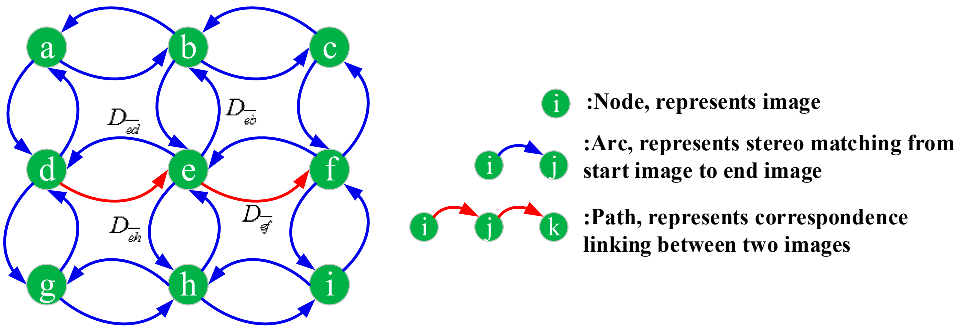

a
b
c
d
e
f
g
h
i
$D_{\overline{ed}}$
$D_{\overline{eb}}$
$D_{\overline{eh}}$
$D_{\overline{ef}}$
i
:Node, represents image
i j
:Arc, represents stereo matching from start image to end image
i j k
:Path, represents correspondence linking between two images

**Figure 1.** The schematic diagram of multi-view graph network structures. A node represents images of the scene. Arc represents stereo matching from start image i to end image j with bound disparity map $D_{ij}$. Path represents correspondence linking between two images.

Overlap ratio and intersection angle are then used to determine the candidate stereo pairs. The overlap ratio can be defined in two ways as illustrated in Figure 2.

In an aerial situation, the corners of image frame are projected to a common plane; then the overlap ratio can be calculated by

$$\varepsilon = S_o / \min(S_1, S_2) \tag{1}$$

where $S_o, S_1, S_2$ represent the area and overlap area respectively. In more general cases, the matched features obtained after a SFM (Structure from Motion) or BA (Bundle Adjustment) process are used to define the overlap ratio as

$$\varepsilon = \min(S_{o1}/S_1, S_{o2}/S_2) \tag{2}$$

where $S_{o1}, S_{o2}$ represent the minimum bounding box of matched features. When the matches are not available, coarse-image matching can be executed with low-resolution images for fast processing.

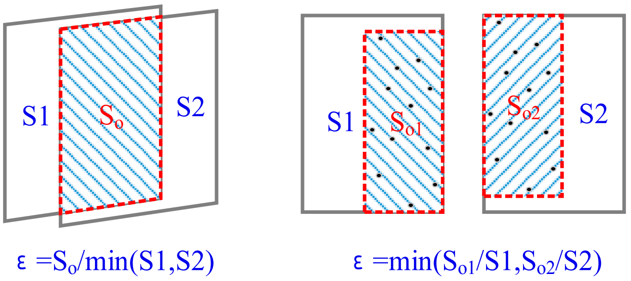


**Figure 2.** Two definitions of the overlap ratio: on the left, the definition in an aerial situation; and on the right the overlap ratio as defined for general purposes.

The intersection angle can be calculated by the two optical axis vectors of the cameras, and only those with an intersection angle less than a threshold (ten degrees is the default) are considered. More occlusion appears between images with a larger intersection angle; adding difficulty to dense image matching and reducing the quality of matching results. Stereo pairs outside of target graph $G$ however, can be connected by certain paths through image pairs with high overlap, thus making full use of the redundant measurements and therefore improving matching accuracy.

## *2.2. Dense Stereo Matching*

In this section, the implemented module for dense stereo matching is described. In Section 2.2.1, the details of epipolar rectification are presented based on the algorithms in [20–23]. In Section 2.2.2, a modified hierarchical strategy based on tSGM algorithm is proposed by introducing the guided median filter that efficiently preserves local structure.

### 2.2.1. Epipolar Rectification

Epipolar rectification must be done before stereo dense matching, which reduces the disparity search space to one dimension. Suppose the camera intrinsic matrix is $K$, the extrinsic parameters are $R_l, T_l$ and $R_r, T_r$; and the relative transformation is $T = T_r - R \cdot T_l$ where $R = R_r \cdot R_l^{-1}$; then, the rotation matrix $R_{rect}$ for rectification between an original image and epipolar image can be calculated as follows:

$$\left\{ \begin{array}{l} e_1 = \frac{T}{||T||}, e_2 = \frac{1}{\sqrt{T_x^2 + T_y^2}} [-T_y, T_x, 0]^T, e_3 = e_1 \times e_2 \\ R_{rect} = \left( \begin{array}{ccc} e_1^T & e_2^T & e_3^T \end{array} \right)^T \end{array} \right. \tag{3}$$

The new rotation matrix for the rectified image is:

$$\left\{ \begin{array}{c} R_l' = R_{rect} \cdot R \\ R_r' = R_{rect} \end{array} \right. \tag{4}$$

Thus, the homographic matrix between the original image and epipolar image is:

$$\left\{ \begin{array}{c} x_{rl} = H_l \cdot x_{ol} \\ x_{rr} = H_r \cdot x_{or} \end{array} \right. \tag{5}$$

After deriving the homographies $H_l, H_r$, the gray values for pixels at integer positions in the rectified frames are calculated by interpolation in Equation (5) using the correspondence coordinates.

### 2.2.2. Modifications of the tSGM Algorithm

The SGM algorithm is often the technique of choice in real world applications as it generates dense reconstruction results with highly robust parameterization and capability for real-time processing. This algorithm however, creates extensive memory demands given the large format frames required and the high number of potential correspondences.

The tSGM algorithm in SURE provides a hierarchical coarse-to-fine solution for the SGM method to limit disparity search ranges and decreases the memory demand as well as processing time. The key advantage of tSGM algorithm is that it makes full use of disparity images matched at low-resolution pyramids to limit the disparity search range at next level pyramids, which dramatically decrease the memory demands and processing times. However, edges are not reconstructed as clearly as in SGM algorithm.

Inspired by the joint bilateral filter [24] and guided filter [25], which have the edge-preserving smoothing property, a guided median filter [26] was introduced to replace the traditional median filter in this paper, with left epipolar image as the guidance image and the output of a pixel is a weighted average of nearby pixels in the guidance image. The pixels are weighted in the local histograms:

$$h(x,i) = \sum_{x' \in \mathrm{N}(x)} w(x,x')\delta(V(x') - i) \tag{6}$$

where $N(x)$ is a local window near $x$, $V$ is the pixel value, $i$ is the discrete bin index, $\delta(\cdot)$ is the Kronecker delta function: $\delta(\cdot)$ is 1 when the argument is 0, and is 0 otherwise, and $w$ is the bilateral weight that suppresses the pixels with different color from the center pixel. It is straightforward to pick the median value through accumulating this histogram. Comparing the stereo matching results by guided median filter and traditional median filter as shown in Figure 3, the disparities in areas with sharp discontinuities obtained better results due to the introduction of the guided median filter.

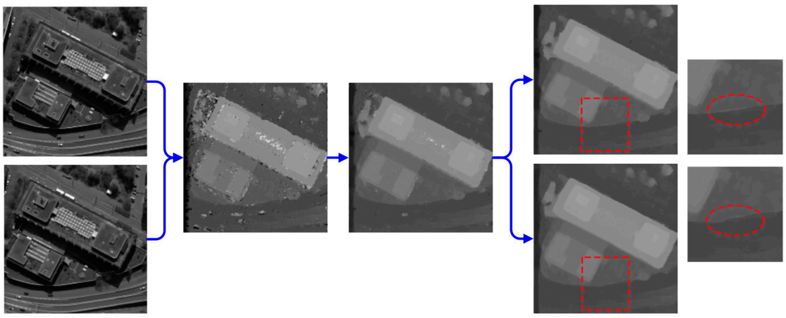


**Figure 3.** The dense matching results of different processing stages for a local area (Column 1: original stereo pair, Column 2: cost aggregation results, Column 3: speckle-filtering results, Column 4: above shows the results from guided median filtering and below shows the results from traditional median filtering).

The hierarchical approach in our implementation is a little different from tSGM algorithm. One advantage of guided median filter in the tSGM algorithm is that a canny edge detector is no longer necessary to optimize the penalty parameter P2, thus increasing the efficiency of the algorithm. When matching an image pair fin a certain pyramid level $l$, the disparity between base and match images is computed followed by a consistency check to obtain valid disparity pixels with high confidence. The guided median filter is adopted to generate a median disparity image. If a pixel is valid, a small 15 × 15 window is searched; else, a larger 31 × 31 window is used to obtain the median value of all valid disparities. After these steps, a speckle-filtering algorithm in OpenCV library [27] is exploited to extract the regions of noise or holes followed by interpolation, again using a guided median filter. The maximum and minimum disparity images $R^l_{max}, R^l_{min}$ are calculated from the final median disparity image $D^l$ with the search ranges for valid and invalid pixels by 16 and 32 according to

$$d_{\min} = \begin{cases} 0, & if \quad d(x_b) - r < 0 \\ d(x_b) - r, & if \quad d(x_b) - r \geq 0 \end{cases} \qquad d_{\max} = \begin{cases} D_{\max}, & if \quad d(x_b) - r > D_{\max} \\ d(x_b) + r, & if \quad d(x_b) + r \leq D_{\max} \end{cases} \tag{7}$$

where $D_{max}$ is the prior maximum disparity of the image pair and $r$ is the half size of search ranges, then path accumulation keeps the same with tSGM algorithm. Considering the direction of edges in graph network, the depth maps of the reference image and matched image are both computed and output for subsequent processing, as shown in Figure 4.

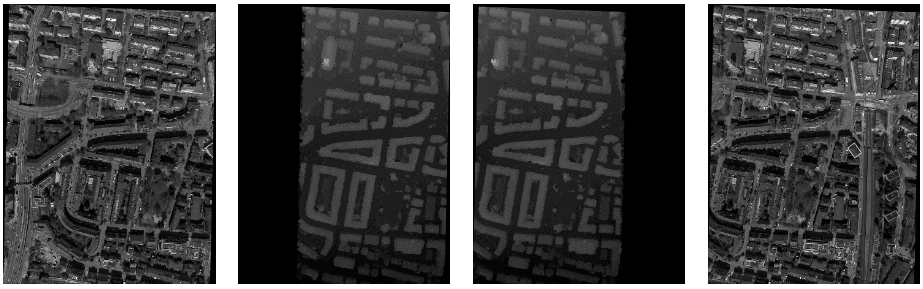


**Figure 4.** Stereo dense matching results for epipolar images (From left to right: left epipolar, left disparity, right disparity, and right epipolar images).

### 2.3. Graph Network Based Triangulation

The geometric correspondence relationship across multi-view images based on graph network is shown in Figure 5.

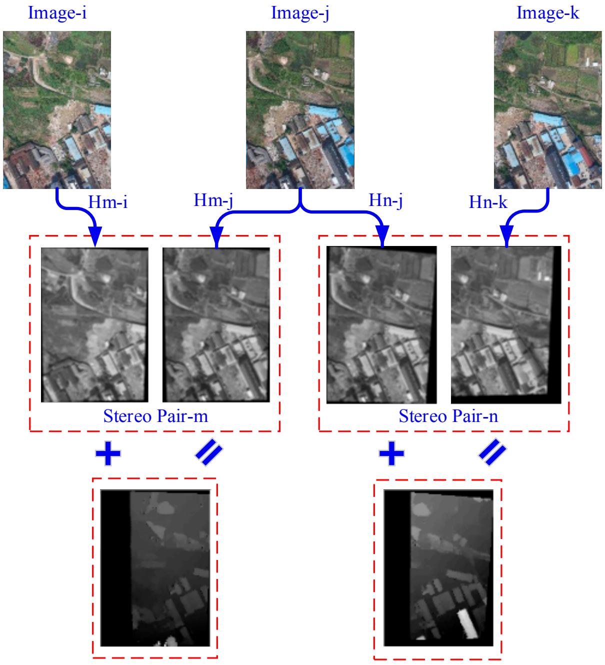


**Figure 5.** The geometric correspondence relationship across multi-view images based on a graph network. First, a pixel on base image can be transformed to left epipolar image by homographic matrix $H_l$. Then, the correspondence on the right epipolar image can be calculated using the disparity image of the base image. Finally, a pixel on a matched image can be obtained by multiplying inverted homographic matrix $H_r$.

For a pixel $x_i$ on image $I_i$, the correspondence $x_j$ on image $I_j$ can be calculated as follows: First its correspondence coordinate on left epipolar image is calculated using the homographic matrix $H_{m-i}$. Then, the matched pixel on right epipolar image is easily obtained from the corresponding disparity of base image. Finally, the correspondence pixel $x_j$ is determined by multiplying the inverted homographic matrix $H_{m-j}$. The whole transformation is as follows:

$$\begin{cases} x_i' = H_{mi} \cdot x_i \\ x_j' = x_i' - D_{ij}\left(x_i'\right) \\ x_j = H_{mj}^{-1} \cdot x_j' \end{cases} \tag{8}$$

Since the transformed coordinate is not integer, bilinear interpolation will be exploited to calculate the value on the disparity image. Moreover, the disparity image of a base image or matched image has to be chosen according to the direction of arcs in graph network. For a path with length of $d$ from image $I_i$ to image $I_n$, the correspondence on image $I_n$ can be computed along the path by applying the transformation in Equation (8), for $d$ times. To reconstruct the dense results for a reference image, a full graph search is utilized to find the candidate images for triangulation. However, only those that meet the minimum overlap conditions (10% as the default) and intersection angle (10 degrees as the default) are maintained. After all the matches across the multi-view images in the graph network are computed, the corresponding object point $X_i$ can be calculated using multi-view triangulation according to

$$x_i = P_i \cdot X_i \tag{9}$$

where $P_i$ represents the camera projection matrix of image $I_i$, and the average reprojection error can be calculated by

$$\sigma_{\mathrm{r}} = \sum_{i=0}^{n} \sqrt{\left(x_i' - x_i\right)^2 + \left(y_i' - y_i\right)^2} / n. \tag{10}$$

Despite the use of consistency checking, speckle removal, and guided median filtering, there is still some error. By setting the minimum number of image matches and the threshold for reprojection error during triangulation, potential error can be efficiently eliminated, which greatly reduces the complexity and runtime in the subsequent filtering process. Compared to direct triangulation from stereo pairs, graph network based triangulation makes full use of redundant measurements, resulting in higher accuracy for reconstructed dense point clouds. Due to the use of graph network, the triangulation of the same object point from different base images will get the same results. The dense point clouds extracted from different images therefore, will have a consistent coordinate system and can be more easily fused.

Benefiting from the transitivity of graph network, there is no need to compute stereo matching for image pairs with low overlap directly, resulting in different matching configurations for the proposed method. Three different matching configurations were designed and tested.

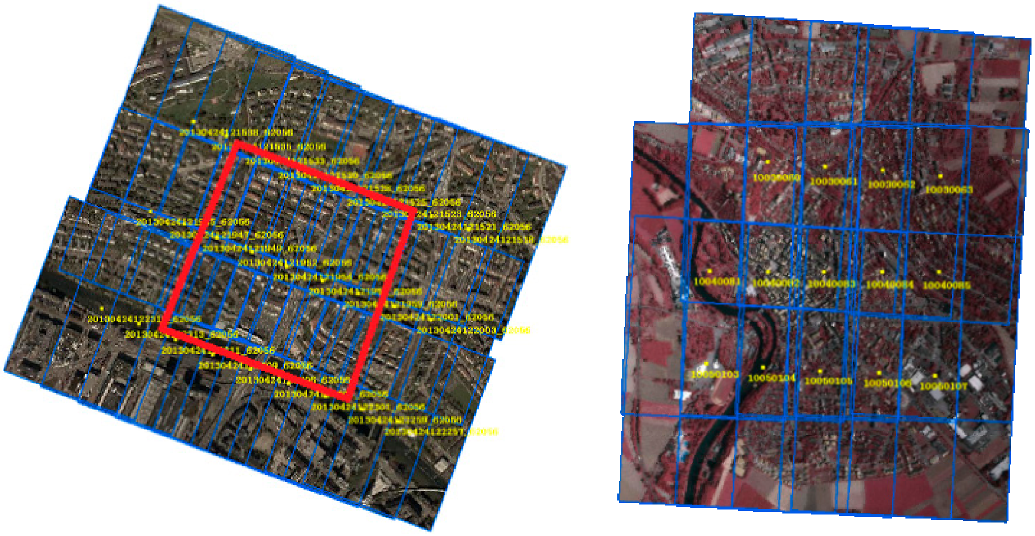
### *2.4. Multi-View Point Cloud Fusion*

Considering the influence of perspective, occlusion, and mismatching, point clouds generated from different views need be fused and filtered so that the scenario results are complete. Since there is a high density of points in overlapping areas, resampling is applied to generate a final point cloud. The main steps of point cloud fusion in the current implementation are as follows: (1) Using noise filtering algorithm in Point Cloud Library (PCL) [28] to remove the remaining mismatches of every single point clouds. Since most mismatches are excluded during the multi-view consistency checking process, the number of those to be eliminated is small; (2) Calculating the minimum bounding box of the target survey areas and partition task units by a tiling mechanism to make the task parallelization and accelerate the progress; (3) Within each task unit, point clouds that intersect with a bounding box are fused and resampled, using the K-d tree algorithm. The average reprojection error is the priority measure.

## 3. Experiments and Results

### *3.1. Description of the Test Dataset and Experiments*

To verify the effectiveness of the graph network-based multi-view DIM method proposed in this paper, in our experiments two public image datasets in an ISPRS test project were used as shown in Figure 6.



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

**Figure 6.** The coverage and overlap relationships of test image datasets: (**a**) Zürich dataset (red rectangle for test area); (**b**) Vaihingen dataset.

The first dataset was captured over the city of Zürich with a Leica RCD30 Oblique Penta medium format camera with a GSD (Ground Sample Distance) of 6 cm. The test area to be processed is depicted in Figure 6 by a red rectangle overlaid on an ortho image. In these experiments, only the nadir images of an oblique image block were used. The second dataset was captured over Vaihingen in Germany; it is a part of an Intergraph/ZI DMC block at 8 cm ground resolution. This dataset was collected with a Leica ALS50 system with the average point density of about 4 points/m$^2$. This was used as reference data to evaluate actual positioning accuracy. Descriptive information for the two datasets is shown in Table 1.

**Table 1.** Data description of test image datasets.

| Dataset | Image Resolution | Pixel Unit/um | f/mm | GSD/cm | Image Number | Height/m | Overlap Ratio | | With Reference Data |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | Longitudinal | Lateral | |
| Zürich | 9000 × 6732 | 6.0 | 53.0 | 6.0 | 27 | 500 | 70% | 50% | × |
| Vaihingen | 7680 × 13,824 | 12.0 | 120.0 | 8.0 | 14 | 900 | 60% | 60% | √ |

The proposed algorithm was implemented by VC++ 2010 and integrated to our Photogrammetric Package; Mogas, which adopted a Census and MI matching cost for computing the disparity volume. In our experiments, for better performance, Census cost is used to calculate the disparity space image. The related algorithm parameters are set as follows: P1, P2 are set to 30 and 150 in SGM algorithm, the radius of guided median window is set to 19, and the speckle size is set to 256, corresponding to original resolution. All the parameter settings were kept the same for both Zürich and Vaihingen dataset. To reduce memory usage and processing time, the images are down-sampled by half before the dense matching process. The experiments were conducted on a Dell Precision T7600 workstation with a 64-bit Window 7 operation system, a sixteen Intel Xeon E5-2650 M CPU, 2.0 GHz and 64 GB memory.

To explore the most efficient approach with the highest accuracy for multi-view dense matching based on a graph network, three kinds of matching configurations were designed and tested, as illustrated in Figure 7.

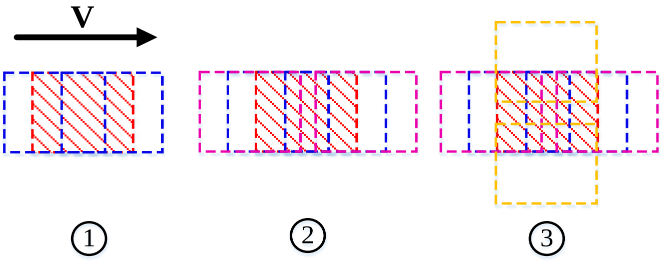


**Figure 7.** Three kinds of matching configurations for multi-view dense matching based on a graph network: ① considers only adjacent image pairs in flight ② considers all possible image pairs with overlap, in flight ③ considers overlapping image pairs both in flight and across flights.

## 3.2. Results

### 3.2.1. Multi-View DIM Results of the Proposed Method Using the Zürich Dataset

To validate the effectiveness and geometric accuracy of the proposed algorithm, the dense point cloud of image-954 of the Zürich dataset is shown together with its height map and accuracy map produced using first configuration, generated according to the average reprojection error during the triangulation process, shown in Figure 8.

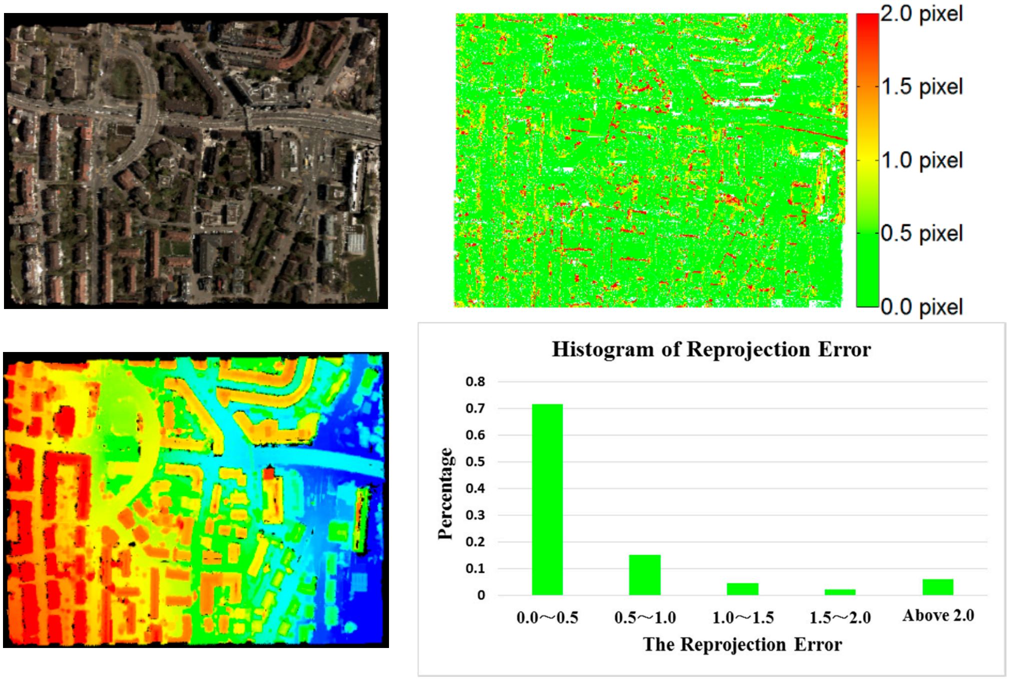


**Figure 8.** The dense matching results of image-954. (Top left: the colorized point cloud, Bottom left: the height map of the point cloud, Top right: the accuracy map rendered by average reprojection error, Bottom right: the histogram for the reprojection error).

Figure 8 shows that the proposed algorithm is effective for multi-view dense matching with high completeness and precision; the RMSE of average projection error was around 0.60 pixel. Furthermore, it can be seen that most of the points with large error are located on the edge of tall buildings, areas of vegetation, and occluded areas as seen in the accuracy map. This result was consistent with the results reported in the literature [13,29]. For these areas with large error, dense image matching often produces results with incorrect disparity or without disparity because there are no correspondences or because there is a great deal of noise, which results in lower confidence. These sources of error will be eliminated during the triangulation process.

To evaluate the efficiency and accuracy of different matching configurations, the number of stereo pairs and points (point cloud of image-954), the processing time, and the RMSE of the average reprojection error were used for comparison; these results are shown in Table 2.

**Table 2.** Comparative results for all three configurations.

| Configuration Solutions | Number of Stereo Pairs | Number of Points (Image-954) | Runtime/min | RMSE/Pixel |
|---|---|---|---|---|
| ① | 24 | 11,574,821 | ≈58 | 0.60 |
| ② | 45 | 11,018,062 | ≈101 | 0.70 |
| ③ | 62 | 12,136,501 | ≈135 | 1.08 |

Table 2 shows that integration of stereo pairs with lower overlap; whether in flight or across flights during MVS processing for the current algorithm, has little effect on the number of reconstructed points. The accuracy slightly decreased when the number of stereo pairs increased. Typically, in aerial photogrammetry, a larger base-height ratio will yield higher intersection accuracy. In a dense matching situation, increasing the base-height ratio also means more occluded areas between stereo pairs, and significantly influences the quality of the SGM dense matching results. Points in occluded areas often produce matching results with more noise, which leads to low accuracy during triangulation process. Meanwhile, correspondence in image pairs outside the candidates selected during the construction stage of the graph network can be matched using the connections in the graph network, resulting in more observations for triangulation and therefore, higher accuracy.

Furthermore, when compared with the third configuration adopted in SURE and PhotoScan, the total processing time was reduced by 57% using the first configuration, as recommended in this paper. The number of images however was small and the overlap relationships were simple in the test dataset, nevertheless the total runtime can be reduced by multiples when applying the proposed algorithm to scenes with larger amount of images. Increasing the in-flight overlap with the minimum required overlap across flights for aerial photogrammetry balances between reconstruction accuracy and efficiency of field operations.

According to the method presented in Section 2.4, the multi-view point clouds within the test area were fused and resampled with the average reprojection error as a priority measure, the results are shown in Figure 9.

Figure 9 shows that the presented strategy of multi-view point cloud fusion is effective, with the RMSE of average projection error of the whole target area at around 0.55 pixel, which is superior to the single view accuracy. The proportion of points with large error was reduced when compared to single view, as evident in the accuracy map from the histogram of reprojection error. During the fusion process, points in the neighborhood with the minimum average reprojection error are reserved as resampled points, which can reduce the difficulty and complexity of post-processing and further improving the accuracy of the final point cloud. This approach can be easily extended to oblique photogrammetry, where the dense point clouds from different directions must also be fused.

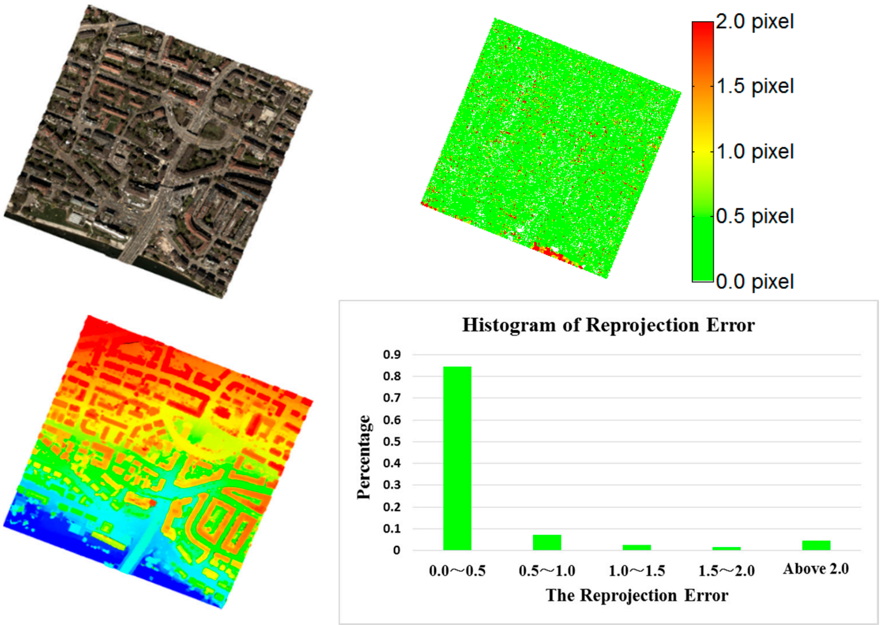


**Figure 9.** The fused result of multi-view point clouds are presented (Top left: the colorized point cloud, Bottom left: the height map of point cloud, Top right: the accuracy map of average projection error, Bottom right: the histogram of reprojection error).

### 3.2.2. Robustness for Matching in Difficult Areas

Regions with sharp discontinuities, weak textures or repeated textures like tall buildings, and areas of vegetation, are considered difficult areas in dense matching [13,29]. Complete and accurate matching results for those kinds of regions can be obtained using the proposed algorithm as illustrated in Figure 10. Due to the introduction of guided median filter in tSGM algorithm, sharp edge characteristics of the target can be maintained, which ensures accurate dense matching results at sharp discontinuities. Reasonable disparities in areas with weak textures or repeated textures can be computed according to the surrounding valid pixels by guided median filtering and interpolation.

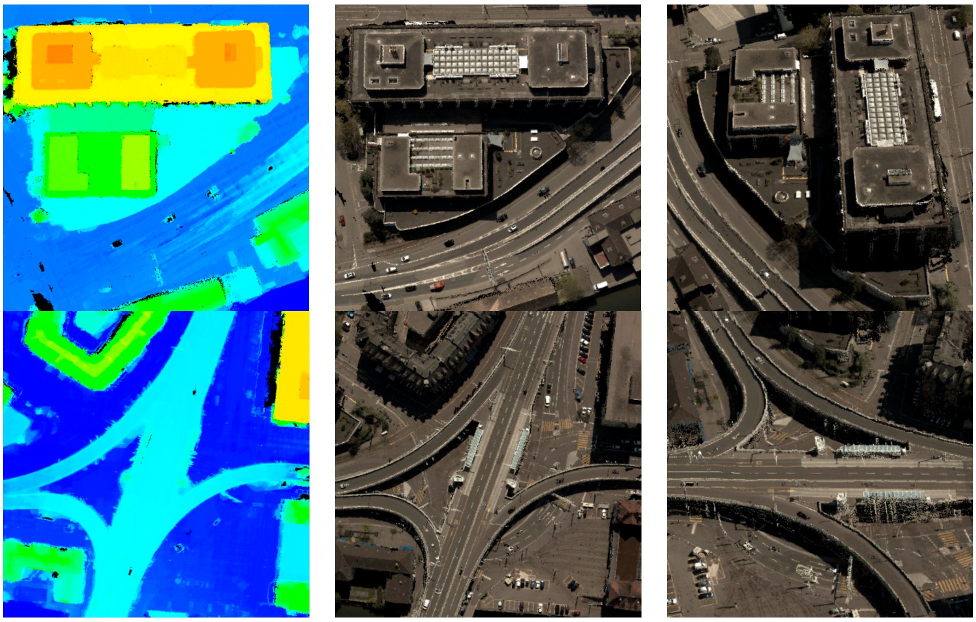


**Figure 10.** *Cont.*

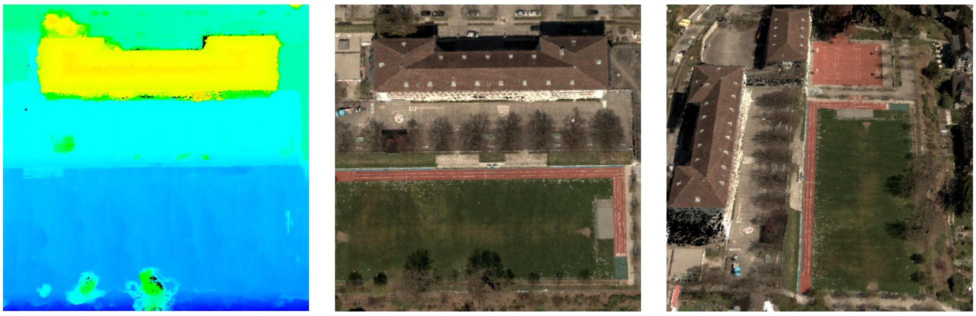


**Figure 10.** Multi-view dense matching results for difficult areas with sharp discontinuities, weak textures, or repeated textures.

## 4. Discussion

### 4.1. Accuracy Analysis with Reference ALS Data

Since the actual positioning accuracy is affected by both matching accuracy and base-height ratio, reprojection error cannot fully evaluate the result of dense matching, so the Vaihingen dataset with reference ALS point cloud was used to validate the RMSE for positioning accuracy in the target area with vegetation and buildings. The dense matching results produced by the tested three configurations are compared to the reference data. Deviation maps were generated, as shown in Figure 11.

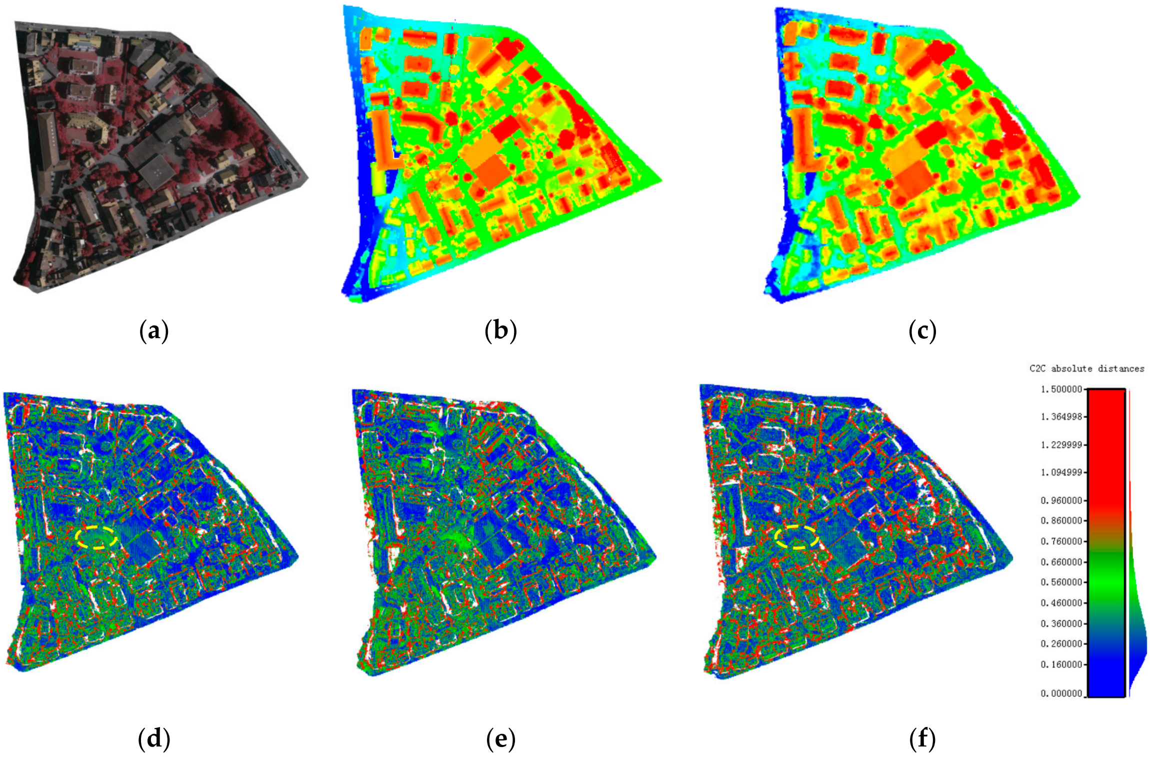


**Figure 11.** The dense matching results for the target area in the Vaihingen dataset: (**a**) The color image of the target area; (**b**) The height map of the reference ALS point cloud; (**c**) The height map of the DIM result from the proposed method; and (**d**–**f**) represent the distance deviation maps for the DIM results from the three tested configurations.

Figure 11 shows that the dense matching results from the proposed method were consistent with the reference ALS point cloud. The actual positioning accuracy was 1.50 GSD, 1.75 GSD and

1.87 GSD for three configurations, respectively; similar to the accuracy as determined using reprojection error as discussed in Section 3.2.1. Comparing the positioning error maps of the three configurations, configuration ② was less accurate than configuration ① due to the impact of the low matching accuracy of wide-baseline image pairs using the same image collection for triangulation. On the other hand, image pairs across flights were added in configuration ③ for triangulation, which was more accurate in some local areas than configuration ① as shown in the area denoted by a yellow ellipse. However, the matching accuracy in the areas with sharp discontinuities like tall buildings was further reduced due to the big intersection angle between image pairs across flights, this resulted in less accurate results than configuration ①, which considered completeness and overall accuracy.

### *4.2. Comparison with SURE and PhotoScan*

Since our method is close to SURE, the point clouds obtained by the two methods were first compared through visual inspection. Figure 12 (red ellipse areas) displays the results for the Zürich dataset. Ghosting effects are visible in the SURE results, while the results produced using our proposed method show no evidence of Ghosting. In the SURE method, point clouds for the same scene are generated from different views during the triangulation process with different correspondences. This results in a slight offset across the multi-view point clouds. In contrast, since a graph network is used in our algorithm, the point clouds from different views share the same correspondences with no offset and can be seamlessly integrated.

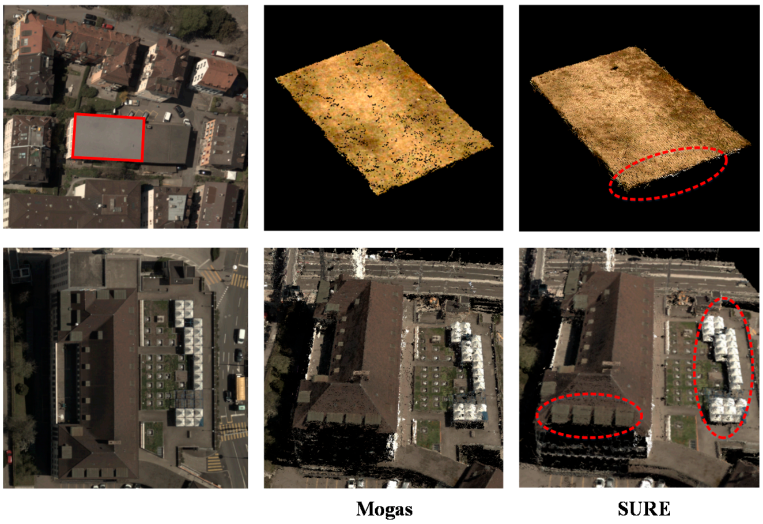


**Figure 12.** Multi-view dense matching results from the proposed method and SURE (From left to right: Origin images, matching results from the proposed method, and matching results from SURE).

To further validate the performance and accuracy of the proposed method, two state-of-the-art software product (SURE and PhotoScan) results are compared at three target areas from easy to challenging in the Vaihingen dataset. Target-1 is a planar playground, target-2 is a single building and target-3 is a continuous complex building. Affected by the plant growth due to time gap between image capture and LiDAR flights, and given differences in measurement principles, the reference ALS data cannot exclusively match the DIM results. Areas of vegetation therefore, were not used in these experiments.

To make a fair comparison, the three methods are excuted at the same image resolution using configuration three, which is adopted as the solution in SURE and PhotoScan. In order to reduce the requirements for professional skills, SURE and PhotoScan are designed ina one-button processing mode, so the matching results are generated with default settings. In this paper, point clouds obtained from the three methods were first resampled with the grid width of the GSD for the respective imagery to get the same density, so points with the median elevation were retained when the neighborhood of a target grid point contained matching results. Then, deviation maps were generated by computing the euclidean distance using reference data. The RMSE and mean error were also calculated along with the correctness and completeness rates that represent the percentage of pixels with an error <n × GSD, as shown in Figures 13 and 14. Correctness was computed with the value of three GSD, completeness with ten GSD.

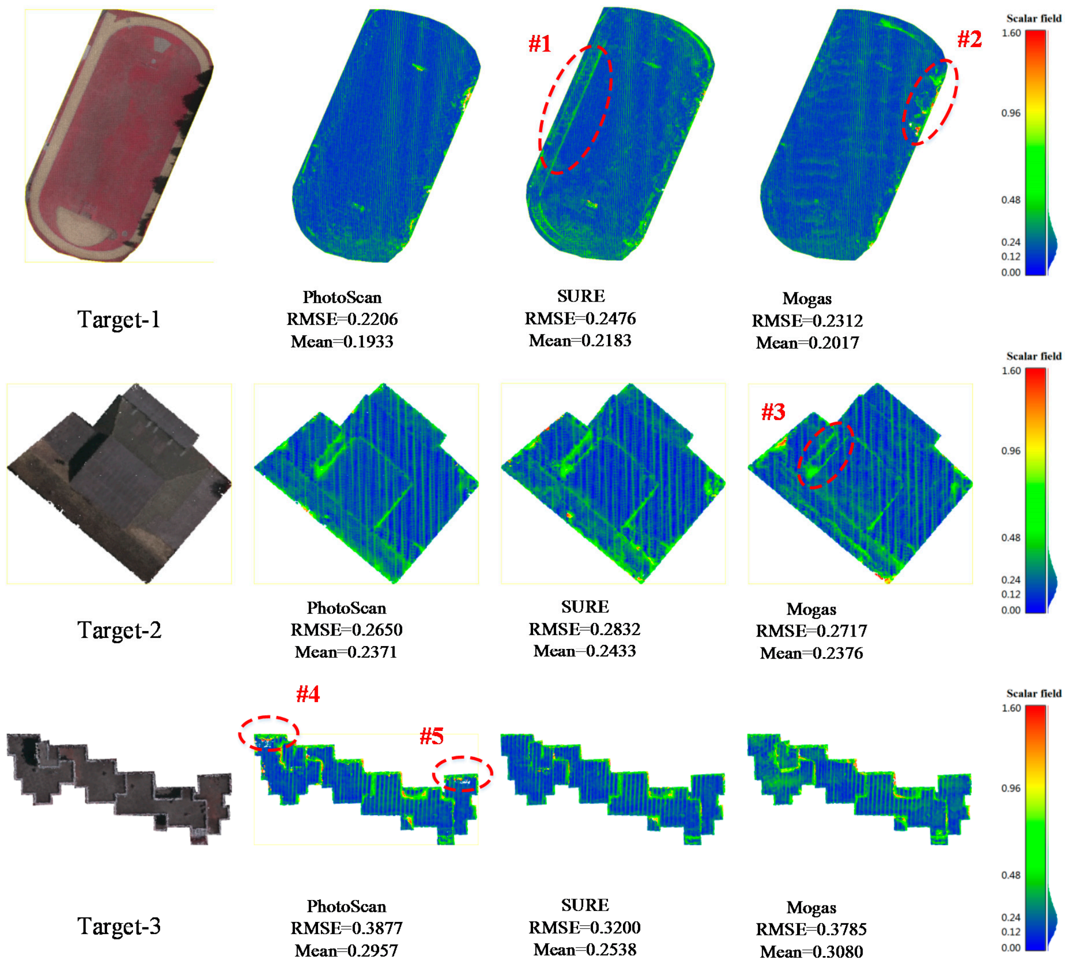


**Figure 13.** The deviation maps for the three target areas from the proposed method, PhotoScan, and SURE (Column 1: the original color images, Column 2: deviation maps of PhotoScan method, Column 3: deviation maps of SURE method, and Column 4: deviation maps of the proposed method).

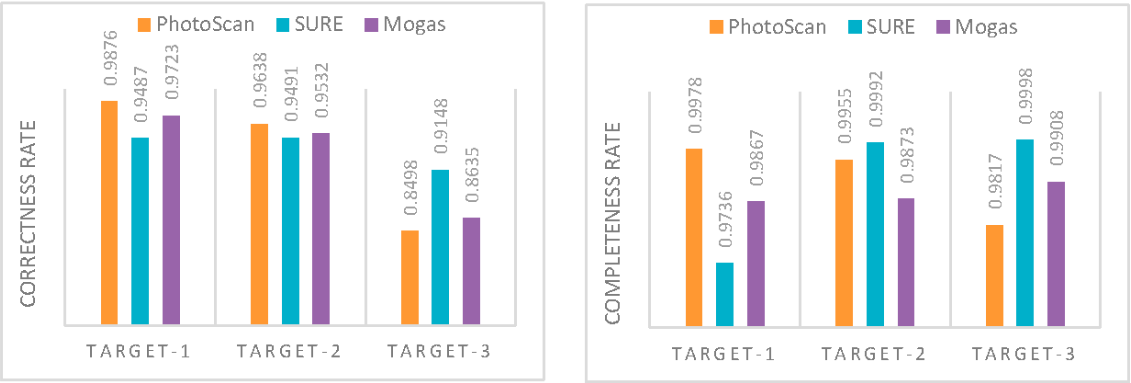


**Figure 14.** The statistics of correctness rate and completeness rate in target areas.

To compare the performance of three methods in flat areas, target-1 was selected for the simple case, as shown in the top row of Figure 13. These results show that the proposed method is comparable to PhotoScan, but better than SURE, considering the RMSE and mean error. SURE has a large deviation in the contour region, #1 for a runway, while areas with large deviations in the results from the proposed method are mainly located in areas of cast shadows, #2. Because the image intensity values in those regions are inconsistent with the surrounding areas, the cencus cost and guided median filter are not viable, so the match fails. In the more difficult cases, target-2 and target-3 were selected to evaluate performance on buildings in urban areas. From the deviation maps of target-b as depicted in the middle of Figure 13, the results are similar to those obtained for target-1 by the RMSE and mean error. This structure was effectively captured by all three methods. However, it can be seen that the proposed method works best on the edge details of #3 showing a building covered by a cast shadow. The most challenging case in target-3, where buildings are of different heights with more occlusionsis presented at the bottom of Figure 13, PhotoScan no longer the delivers the best performance due to missed matches on the both ends of the buildings in region #4 and #5. In contrast to PhotoScan, the proposed method and SURE demonstrate more complete results. However, all the solutions from three methods appear relatively noisy in the outlined areas of buildings. Considering the correctness rate as dipicted on the left of Figure 14, the statistics are consistent with the previous analysis. As for the completeness rate as shown in right of Figure 14, all three methods perform well and have a completeness rate of over 95%.

### *4.3. Strengths and Limitations*

A graph network-based multi-view dense matching algorithm is proposed in this study. The new method was successfully applied for dense point cloud generation from multi-view high-resolution aerial images. The main advantage of the proposed approach is improvement in the efficiency and completeness for stereo pair based multi-view dense matching with high accuracy. Experimental results show that the average reprojection error is at the subpixel level, and actual positioning accuracy is about 1.5 GSD.

The time consumed by dual-depth map generation for a single stereo pair in 15 Megapixel was about 2 min in our tSGM implementation. Fewer stereo pairs need to be computed in the proposed method than in the conventional practices found in the SURE and PhotoScan software, which decreased the overall processing time by 57% in our experiment with the Zürich dataset. In the selected aerial scenario, the overlap ratio was low and the number of images was small. For large scale scenarios with high overlap, the execution time could be further reduced, thereby considerably improving the efficiency and applicability of the MVS algorithm.

Within multi-view stereo methods based on image space, disparity estimations of single stereo models are typically merged in order to increase the reliability and precision of the final depth maps or point clouds. The original SGM [3] proposes an orthographic 2.5D projection method to fuse several disparity images from different viewpoints for aerial applications, and SURE [5] presents a method to transform all disparity maps of base image to the distance maps between camera center and object point on the optical ray, then fuses them by minimizing the reprojection error. However, point clouds from different base images will have small offsets since in these methods they are generated separately without connection. Those produced by the proposed method show higher consistency, therefore these point clouds can be seamlessly integrated without extra processing.

Considering that the performance of stereo matching decreases sharply with the increased base-height ratio using a SGM algorithm, image pairs with wide baselines can obtain higher correspondence using the proposed method than the direct matching result. Our experimental results show that configuration ① has a higher performance than configuration ② and ③. The proposed method performs better than traditional practices found in SURE and PhotoScan since the point clouds from different base images have consistent coordinates, with competitive efficiency.

The proposed method can be combined with an object-based MVS algorithm such as the plane sweep stereo method [30,31] to achieve the optimal performance and efficiency. A minimum number of image pairs are computed to obtain initial matching results with high accuracy, so that the depth search space can be constrained in a small range for the object-based MVS method to refine the results.

## 5. Conclusions

A novel graph network based multi-view dense matching algorithm for high-resolution aerial images is proposed in this paper. In this method, a modified tSGM algorithm is first used to compute dual direction disparity maps across stereo pairs, and point clouds of single view are then generated by triangulation based on graph network, followed by multi-view point cloud fusion with average reprojection error as a priority measure.

The aerial image dataset of Vaihingen and oblique nadir image block of Zürich from the ISPRS test project are used to verify the effectiveness and accuracy of the proposed method. Experimental results demonstrate the completeness, efficiency, and high precision of this algorithm as results at the sub pixel level can be achieved. To explore the most efficient approach for multi-view dense matching based on a graph network, three kinds of configurations were designed and tested. Considering the numbers of image pairs, runtime and accuracy, our comparative experiments confirm that the first configuration is superior to the other two configurations as adopted in the SURE and PhotoScan approaches. Extensive experiments show that the proposed method can derive results comparable to SURE and PhotoScan, with no ghosting and improves the accuracy of the final point cloud and can be extended to oblique photogrammetry. Furthermore, experiments verified the effectiveness of this strategy for multi-view point cloud fusion and difficult areas with sharp discontinuities, weak textures, or repeated textures can be matched by the proposed algorithm.

This research provides a good foundation for oblique multi-view dense matching; our future research will focus on further improvements to the efficiency and accuracy of the proposed algorithm. Given that random noise of final point cloud may still exist, we will attempt to exploit a post-processing process to refine the results to make it smoother.

**Acknowledgments:** This research was supported by the National Public Welfare Foundation for Surveying, Mapping and Geoinformation Industry Research of China with Project Number 201512008 and the Fundamental Research Funds for the Central Universities of China with Project Number 2015214020201. The authors would like to thank the ISPRS/EuroSDR Project and the German Society for Photogrammetry, Remote Sensing, and Geoinformation for providing the Zürich dataset [32]. The author would like to thank Agisoft and Nframes for providing the trial version of PhotoScan and SURE software.

**Author Contributions:** Liang Fei conceived and designed the experiments; Liang Fei and Changhai Chen performed the experiments; Li Yan and Liang Fei and Zhiyun Ye analyzed the data; Liang Fei and Changhai Chen and Ruixi Zhu contributed reagents/materials/analysis tools; Liang Fei and Li Yan wrote the paper; Zhiyun Ye and Ruixi Zhu helped to prepare the manuscript. All authors read and approved the final manuscript.

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

## Abbreviations

The following abbreviations are used in this manuscript:

| | |
|---|---|
| MVS | Multi-View Stereo |
| DIM | Dense Image Matching |
| SGM | Semi-Global Matching |

## References

1. Remondino, F.; Spera, M.G.; Nocerino, E.; Menna, F.; Nex, F. State of the art in high density image matching. *Photogramm. Rec.* **2014**, *29*, 144–166. [CrossRef]
2. Hirschmuller, H. Accurate and efficient stereo processing by semi-global matching and mutual information. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Diego, CA, USA, 20–26 June 2005.

3. Hirschmuller, H. Stereo processing by semiglobal matching and mutual information. *IEEE Trans. Pattern Anal. Mach. Intell.* **2008**, *30*, 328–341. [CrossRef] [PubMed]
4. Gehrke, S.; Morin, K.B.; Downey, M.A.; Boehrer, N.C.; Fuchs, T.C. Semi-global matching: An alternative to LIDAR for DSM generation. In *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Proceedings of the ISPRS Commission I Mid-Term Symposium, Calgary, AB, Canada, 15–18 June 2010.
5. Rothermel, M.; Wenzel, K.; Fritsch, D.; Haala, N. SURE: Photogrammetric Surface Reconstruction from Imagery. 2012. Available online: http://www.ifp.uni-stuttgart.de/publications/2012/Rothermel_etal_lc3d.pdf. (accessed on 26 September 2016).
6. Zhang, L. *Automatic Digital Surfece Model (DSM) Generation from Linear Array Images*; Mitteilungen—Institut fur Geodasie und Photogrammetrie an der Eidgenossischen Technischen Hochschule Zurich: Zurich, Switzerland, 2005.
7. Pierrot-Deseilligny, M.; Paparoditis, N. A multiresolution and optimization-based image matching approach: An application to surface reconstruction from SPOT5-HRS stereo imagery. In *Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, Proceedings of the Commission I Symposium, Paris, France, 4–6 May 2006.
8. Goesele, M.; Snavely, N.; Curless, B.; Hoppe, H.; Seitz, S.M. Multi-view stereo for community photo collections. In Proceedings of the IEEE 11th International Conference on Computer Vision, Rio de Janeiro, Brasil, 14–20 October 2007.
9. Remondino, F.; Menna, F. Image-based surface measurement for close-range heritage documentation. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2008**, *37*, 199–206.
10. Furukawa, Y.; Ponce, J. Accurate, dense, and robust multiview stereopsis. *IEEE Trans. Pattern Anal. Mach. Intell.* **2010**, *32*, 1362–1376. [CrossRef] [PubMed]
11. Vu, H.-H.; Labatut, P.; Pons, J.P.; Keriven, R. High accuracy and visibility-consistent dense multiview stereo. *IEEE Trans. Pattern Anal. Mach. Intell.* **2012**, *34*, 889–901. [CrossRef] [PubMed]
12. Toldo, R.; Fantini, F.; Giona, L.; Fantoni, S.; Fusiello, A. Accurate multiview stereo reconstruction with fast visibility integration and tight disparity bounding. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2013**, *40*, 243–249. [CrossRef]
13. Szeliski, R. *Computer Vision: Algorithms and Applications*; Springer: Berlin, Germany, 2010.
14. Hirschmüller, H. *Semi-Global Matching Motivation, Developments and Applications*; Wichmann/VDE Verlag: Belin/Offenbach, Germany, 2011.
15. Hirschmuller, H.; Scharstein, D. Evaluation of stereo matching costs on images with radiometric differences. *IEEE Trans. Pattern Anal. Mach. Intell.* **2009**, *31*, 1582–1599. [CrossRef] [PubMed]
16. Wenzel, K.; Rothermel, M.; Haala, N.; Fritsch, D. *SURE—The ifp Software for Dense Image Matching*; Wichmann: Berlin/Offenbach, Germany; VDE Verlag: Berlin, Germany, 2013.
17. Deseilligny, M.P.; Clery, I. Apero, an open source bundle adjusment software for automatic calibration and orientation of set of images. In Proceedings of the ISPRS Symposium, Sydney, Australia, 10–15 April 2011.
18. Roy, S.; Cox, I.J. A maximum-flow formulation of the n-camera stereo correspondence problem. In Proceedings of the Sixth International Conference on Computer Vision, Bombai, India, 7 January 1998.
19. Koch, R.; Pollefeys, M.; van Gool, L. Multi viewpoint stereo from uncalibrated video sequences. In *Computer Vision—ECCV'98*; Springer: Berlin, Germany, 1998; pp. 55–71.
20. Loop, C.; Zhang, Z. Computing rectifying homographies for stereo vision. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Fort Collins, CO, USA, 23–25 June 1999.
21. Pollefeys, M.; Koch, R.; van Gool, L. A simple and efficient rectification method for general motion. In Proceedings of the Seventh IEEE International Conference on Computer Vision, Kerkyra, Greece, 20–27 September 1999.
22. Monasse, P.; Morel, J.-M.; Tang, Z. Three-step image rectification. In Proceedings of the BMVC 2010-British Machine Vision Conference, Aberystwyth, UK, 30 August 2010–2 September 2010.
23. Fusiello, A.; Trucco, E.; Verri, A. A compact algorithm for rectification of stereo pairs. *Mach. Vis. Appl.* **2000**, *12*, 16–22. [CrossRef]
24. Kopf, J.; Cohen, M.F.; Lischinski, D.; Uyttendaele, M. Joint bilateral upsampling. *ACM Trans. Graph. (TOG)* **2007**, *26*, 96. [CrossRef]

25. He, K.; Sun, J.; Tang, X. Guided image filtering. *IEEE Trans. Pattern Anal. Mach. Intell.* **2013**, *35*, 1397–1409. [CrossRef] [PubMed]
26. Ma, Z.; He, K.; Wei, Y.; Sun, J.; Wu, E. Constant time weighted median filtering for stereo matching and beyond. In Proceedings of the 2013 IEEE International Conference on Computer Vision, Sydney, Australia, 1–8 December 2013.
27. Bradski, G. The opencv library. *Dr. Dobbs J.* **2000**, *25*, 120–126.
28. Rusu, R.B.; Cousins, S. 3D is here: Point cloud library (pcl). In Proceedings of the 2011 IEEE International Conference on Robotics and Automation, Shanghai, China, 9–13 May 2011.
29. Haala, N. The Landscape of Dense Image Matching Algorithms. Available online: http://www.gtbi.net/wp-content/plugins/gallery/uploads/pdf/6951396883562.pdf (accessed on 16 September 2016).
30. Gallup, D.; Frahm, J.M.; Mordohai, P.; Yang, Q.; Pollefeys, M. Real-time plane-sweeping stereo with multiple sweeping directions. In Proceedings of the 2007 IEEE Conference on Computer Vision and Pattern Recognition, Minneapolis, MN, USA, 7–22 June 2007; pp. 1–8.
31. Zach, C.; Gallup, D.; Frahm, J.M.; Niethammer, M. Fast global labeling for real-time stereo using multiple plane sweeps. In Proceedings of the 13th International Fall Workshop Vision, Modeling, and Visualization, Konstanz, Germany, 8–10 October 2008; pp. 243–252.
32. Cavegn, S.; Haala, N.; Nebiker, S.; Rothermel, M.; Tutzauer, P. Benchmarking high density image matching for oblique airborne imagery. In *ISPRS-International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Proceedings of the 2014 ISPRS Technical Commission III Symposium, Zurich, Switzerland, 5–7 September 2014; Volume 1, pp. 45–52.