Article

A Novel Stereo Matching Algorithm for Digital Surface Model (DSM) Generation in Water Areas

Wenhuan Yang [1] [iD] **, Xin Li** [1] **, Bo Yang** [2,*] **and Yu Fu** [3]

[1] School of Remote Sensing and Information Engineering, Wuhan University, Wuhan 430079, China; 2016102130021@whu.edu.cn (W.Y.); xli2126@whu.edu.cn (X.L.)

[2] State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 430079, China

[3] Systems Engineering Research Institute of China State Shipbuilding Corporation, Beijing 100036, China; FY07SERI@126.com

* Correspondence: xt000025@whu.edu.cn

check for updates

Abstract: Image dense matching has become one of the widely used means for DSM generation due to its good performance in both accuracy and efficiency. However, for water areas, the most common ground object, accurate disparity estimation is always a challenge to excellent image dense matching methods, as represented by semi-global matching (SGM), due to the poor texture. For this reason, a great deal of manual editing is always inevitable before practical applications. The main reason for this is the lack of uniqueness of matching primitives, with fixed size and shape, used by those methods. In this paper, we propose a novel DSM generation method, namely semi-global and block matching (SGBM), to achieve accurate disparity and height estimation in water areas by adaptive block matching instead of pixel matching. First, the water blocks are extracted by seed point growth, and an adaptive block matching strategy considering geometrical deformations, called end-block matching (EBM), is adopted to achieve accurate disparity estimation. Then, the disparity of all other pixels beyond these water blocks is obtained by SGM. Last, the median value of height of all pixels within the same block is selected as the final height for this block after forward intersection. Experiments are conducted on ZiYuan-3 (ZY-3) stereo images, and the results show that DSM generated by our method in water areas has high accuracy and visual quality.

Keywords: stereo matching; block matching; DSM; water area; ZY-3

1. Introduction

1.1. Background and Related Works

DSM is a critical component of many physical description of the Earth's surface, and provides the surface structures required for myriad applications as diverse as mapping, navigation, military mission and smart city. Shuttle radar topography mission (SRTM) with 30-m or 90-m resolution, has been one of the most important and widely used publicly available spatial datasets in nearly ten years [1,2]. However, there is a mass of data voids in SRTM, and many of them appear in water regions. Post-processes, including water detection, interpolation, and manual editing, are needed before practical applications. Nevertheless, SRTM is no longer suitable for current various applications due to relatively low resolution. Nowadays, with the continuous improvement of optical remote sensing image resolution, DSM with meter and even sub-meter resolution generated by stereo matching is available [3–5].

Stereo matching, the key step to obtain elevation or depth information from images, has always been a research hotspot for decades in both realms of the digital photogrammetry and computer

vision [6]. So far, excellent stereo matching algorithms can achieve almost the same accuracy as LIDAR (light detection and ranging) in elevation estimation [7–11]. Given the good performance in terms of accuracy, efficiency, and cost, they have been adopted by many well-known digital photogrammetry software programs, such as Geomatica [12] and Inpho [13], for DSM generation. However, poor texture is still a huge challenge for image matching. Additionally, the DSMs automatically extracted in poor texture areas still require a lot of manual editing before practical applications [14–16]. Literature [17] confirms that through-water image dense matching is feasibly for mapping shallow water bathymetry under favorable conditions (clear water, calm water surface, bottom texture). This paper aims at solving the stereo matching problem in water areas under poor texture further, and generating reliable DSMs automatically.

The matching primitive lack of uniqueness due to relatively less intensity variation is the major obstacle for reliable image matching in poor texture areas. There are obvious ambiguities between matching candidates when using these primitives to calculate the matching cost. Efforts have been made to address this problem, which could be divided into three categories in terms of the strategy used: (1) improving the uniqueness of the matching primitive, (2) adopting geometrical constraints, and (3) introducing neighbor smoothing constraints.

It is instinctive to improve the uniqueness of the matching primitive to attain reliable matching results in poor texture areas. Usually, area-based matching methods are employed to handle these areas, in which more intensity variations are involved by increasing the size of the matching template. However, it is very likely that depth discontinuity and projective distortion will be introduced with the size of the matching template growing large, which will instead reduce the reliability of the matching result. In order to overcome this drawback, a matching strategy with a shape and size adaptive template is proposed in [18], in which intensity and prior disparity information are used to minimize the uncertainty of the matching template for each pixel with a higher calculation cost. In addition to this, edge features are extracted and used to avoid incorporating depth discontinuity in other works [19,20]. Moreover, Gaussian filtering and bilateral filtering are also employed to improve the reliability of the matching template through reducing the weight of pixels located at depth discontinuity areas [21]. Unfortunately, it is not enough to identify depth discontinuity based solely on edge features.

In addition to improving the uniqueness of the matching primitive, using geometrical constraints is another useful strategy. An integrated point and edge matching method based on the self-adaptive edge-constrained triangulations is carried out, in which the newly matched points and edges are inserted into triangulations and the constrained triangulations are updated dynamically with matching propagation [15]. Usually, this method is able to achieve reliable matching results from images with poor texture but relatively favorable edge features.

Neighbor smoothing constraints have also been introduced to address this problem. The pixels belonging to the same segment are approximately considered with an identical disparity after the image segmentation have been finished. Then the disparity of one pixel which has been successfully matched can propagate to its neighbors [19]. In addition, patch-based multi-view stereo (PMVS) is another well-known stereo matching method, that generates high-quality point clouds by employing a neighbor smoothing constraint but has high computational complexity [11,22,23].

In summary, the aforementioned stereo matching methods still have limits in both accuracy and efficiency to some extent with poor texture areas. After a thorough analysis, a pixel-by-pixel match pattern is the basic reason. Although various extra information and constraints are introduced, the lack of uniqueness of the point matching primitive is still inevitable in relatively large poor texture areas. Compared to point matching, block matching can avoid the influence of deformation while maintaining high uniqueness. Thus, a flexible block matching strategy is necessary and effective for this issue.

1.2. The Proposed Approach

In our opinion, a proper matching primitive is the key to achieve reliable image matching in poor texture areas. To be more specific, the matching primitive should have enough uniqueness and avoid

the depth discontinuity, simultaneously. In conventional approaches, a fixed window centered at a given pixel is widely adopted, which often cause mismatches because of lacking intensity variation in poor texture areas (Figure 1). Simply increasing the size of the window to include more intensity variation is a better choice, but there remain two problems: (1) An excessively large window may incorporate depth discontinuity, so that it is very likely that the minimum matching cost represents an incorrect matching result due to the projective distortion between the left and right images. (2) The window with a fixed size and shape always has difficulty when handling different scenes in one image. In order to obtain reliable matches in poor texture areas, an adaptive matching primitive strategy tends to play a very important role.
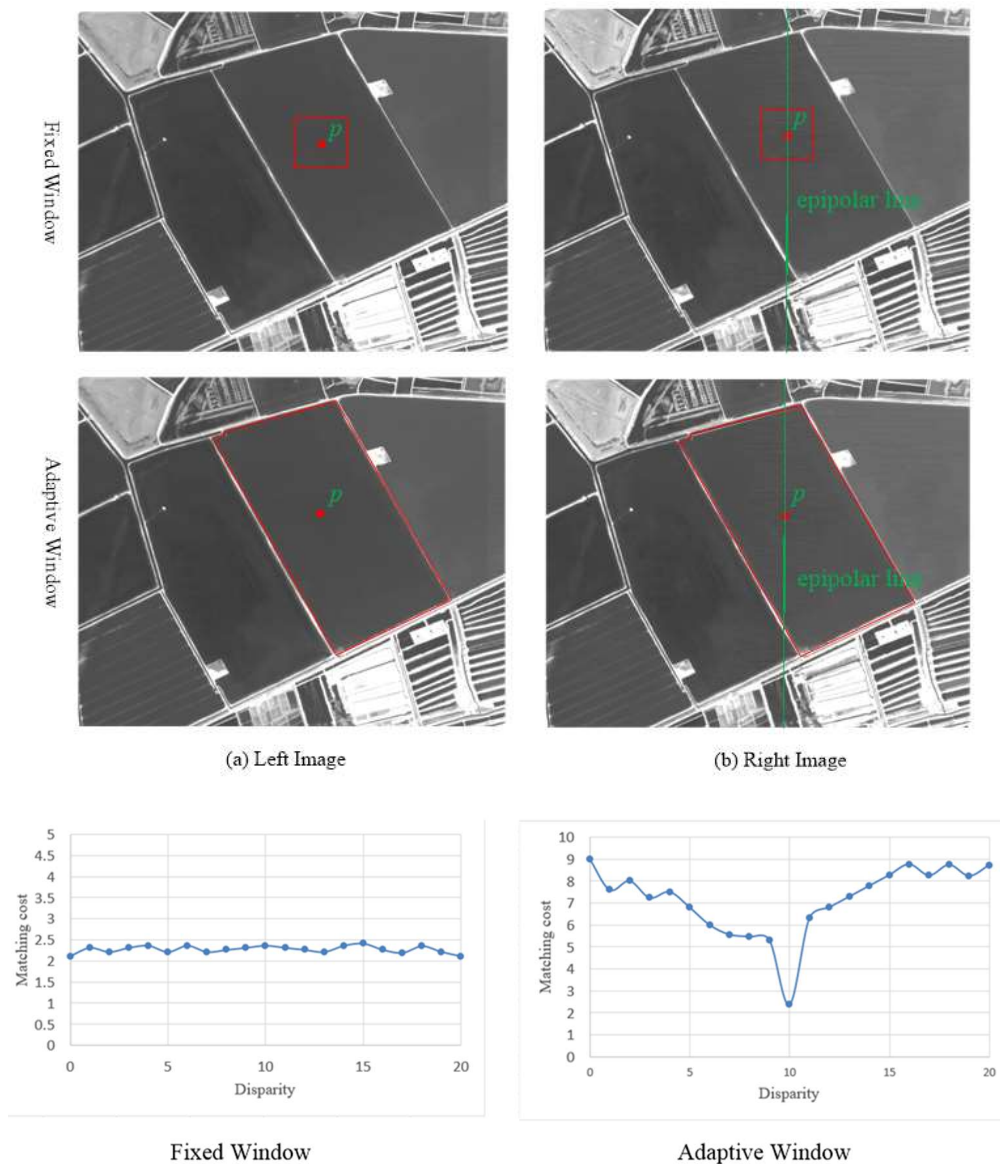

Figure 1. The comparison of "Fixed Window" and "Adaptive Window" as the matching primitive. The matching cost curves were obtained with a disparity range (0,20) along epipolar line. For the "Fixed Window", there is no obvious minimum, while the "Adaptive Window" has an obvious minimum.

Based on the analysis above, we propose a novel stereo matching method for DSM generation from images containing water areas. The significant difference between this method and other conventional methods is that water areas and other areas are handled separately. First, water areas are extracted by seed point growth. Then, an adaptive block matching strategy is employed to achieve accurate disparity

estimation for pixels belonging to the water areas. Additionally, the disparity estimation for other pixels is conducted with classical dense matching methods, such as SGM. After all pixels' disparities have been obtained, DSM is generated by forward intersection and a series of processing steps.

2. Method

Generally, most methods of extracting DSM from optical remote sensing images can be divided into two parts: disparity estimation and DSM generation. The input and output of the former part are a rectified image pair and a disparity map. The input for the later part is a disparity map and rational polynomial coefficients (RPC) while the output is a DSM. A flowchart of the proposed method, named SGBM, is presented in Figure 2. Disparity estimation is the focus of this method, which includes three main stages: (1) water block extraction, (2) disparity range calculation, and (3) image matching and refinement. For the first stage, a seed point detection and growth strategy is employed to identify water blocks. In the second stage, a pyramid matching constraint is introduced to obtain reasonable disparity ranges. The third stage involves block matching and dense matching, which is the main process of this method. Specifically, the disparity estimation of pixels belonging to water blocks is addressed by adaptive block matching, and the disparity estimation of other pixels is achieved by dense matching. Then, disparity refinement is conducted to generate a disparity map. Finally, the disparity map and RPCs are used to generate the DSM by forward intersection and a series of processing steps.
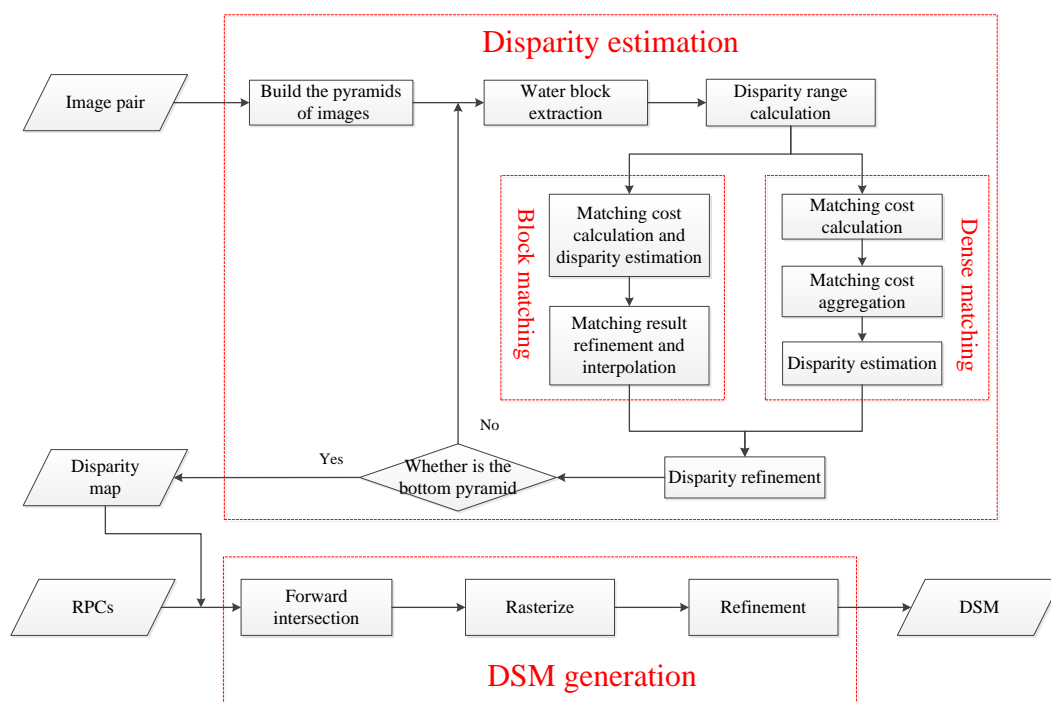

Figure 2. Flowchart of SGBM.

2.1. Water Block Extraction

As previously mentioned, conventional dense matching methods have difficulty with water areas due to unobvious intensity variation. Actually, neither image dense matching nor interferometric synthetic aperture radar (InSAR) can handle water areas well, and there is a high possibility of data voids in the final DSM. Lots of efforts have been made to address these data voids, including water area recognition and void interpolation strategy with auxiliary data and manual editing [24], the optimal interpolation strategy considering terrain and area [25,26], and interpolation approach preserving first and second moments of the slope area relationship [27]. However, all these methods can only be used after the DSM has been generated. We prefer to generate reliable DSM in water areas automatically

with an effective block matching and interpolation strategy. Water block extraction is a pre-step, which directly affects the reliability of the subsequent block matching. Many water area detection approaches have been put forward in the related researches, including a nonparametric expert system based on multiple data sources [28], a water detection method based on multispectral images, terrain data and water masks [29], threshold classification based on gray histogram of panchromatic image [24]. Considering that only panchromatic images are used in this study and the characteristic that the intensity variation is usually very slight in water areas, a simple and effective water block extraction strategy is employed, which consists of two steps: seed point detection and seed point growth.

Seed point detection is to identify pixel p located at water areas as a seed point, which is the input for seed point growth. Thus, the reliability of seed point directly affects the water block extraction result. To avoid misdetection of water blocks as possible, a robust seed point detection strategy is carried out. As Figure 3a shows, a square template centered at pixel p with a certain size (set as 3) is used to identify seed point. If all pixels' grads within it are less than a certain value (set as 5), then p is saved as a seed point. In order to further eliminate the influence of radiation change and noise, the seed point detection is only conducted on the top pyramid image which is the least affected by radiation change and noise. Moreover, the seed points on other pyramid images are obtained from the seed points on the top pyramid image based on geometric relationship between different pyramids.
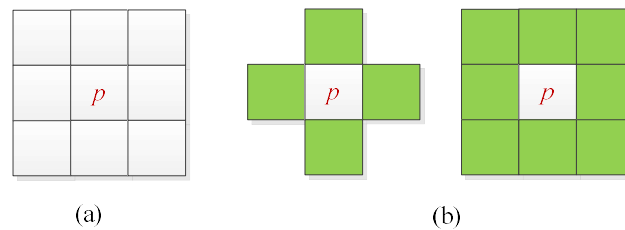
(a)　　　　　　　　　　　　　　　　　(b)

Figure 3. (**a**) Square template for seed point detection. (**b**) 4-neighborhood and 8-neighborhood.

After obtaining the seed points, the seed point growth [30] aims to obtain a water block, B, which contains the current seed point p. Pixel q, one of p' neighborhoods, will be accepted as a seed point, if the difference between the gray value of p and q is smaller than a certain value (set as 3). When seed point growth has been finished, the extracted blocks exceeding a certain size (set as 1000) will be saved. Moreover, 4-neighborhood strategy (the left part of Figure 3b), which has been proven to have better performance in avoiding overgrowth, is picked instead of 8-neighborhood strategy (the right part of Figure 3b) in this study.

Based on the analysis mentioned above, an automatic water block extraction strategy is introduced as below.

(1) Seed point detection is conducted on the whole image with a fixed sampling interval (set as 5), and then all detected seed points are recorded.

(2) Seed point growth for all unprocessed seed points in turn to extract water blocks. If one seed point has been processed or belongs to a water block already detected, it will be marked as processed immediately.

2.2. Disparity Range Calculation

A reasonable disparity range is crucial for both block matching and dense matching. With a pyramid matching strategy, the disparity range for each pixel on different pyramid images will be calculated separately. For the top pyramid image, this is calculated based on the matching results of scale-invariant feature transform (SIFT) [31] feature points with Equation (1).

$$D_p = \left[\ min(d_i) - \Delta d \quad max(d_i) + \Delta d \ \right]_{i \in F_L} \tag{1}$$

where Dp stands for the disparity range of all pixels, d_i is the disparity of pixel i, F_L represents the set of feature points, and Δd is a constant, which is valued at 2.

For other pyramid images, the disparity range is determined by the dense matching result of the upper pyramid image with Equation (2).

$$
\begin{cases}
D_p = \begin{bmatrix} d_p - \Delta d & d_p + \Delta d \end{bmatrix} \\
D_L = \begin{bmatrix} \min(d_p) - \Delta d & \max(d_p) + \Delta d \end{bmatrix}_{p \in B_L} \\
d_p = d_q \times 2
\end{cases}
\tag{2}
$$

where D_p refers to the disparity rang of pixel p, and D_L represents the disparity range of block L, d_p is the disparity of pixel p, d_q is the disparity of pixel q, q stands for the corresponding pixel of p on the upper pyramid image, B_L indicates the current water block, and Δd is a constant, which is valued at 2.

2.3. Block Matching

Following water block extraction and disparity range calculation, block matching is the most important step of disparity estimation for water blocks. The block matching results will directly affect the reliability of disparity estimation. A novel block matching method, neither a feature-based method [31,32] nor an area-based method [33,34], but between them, is proposed. On one hand, as is similar to classical feature-based methods, features, which are water blocks in this study, will be extracted at first. However, there are two different parts between them: (1) feature extraction is only conducted on the reference image and (2) feature descriptors are not needed in our method. On the other hand, as with area-based methods, a block is the matching primitive in our method. However, different from those methods, in which the matching primitive is often a fixed window, our method has matching primitives with adaptive shape and size, which is the key to obtain reliable matches.

Compared to general image matching issue, the water block matching in this study has some unique characteristics which will be introduced below. By making full use of these characteristics, this matching problem can be greatly simplified, and thus the reliability of the matching result can be effectively improved.

(1) Considering that the water block should be a plane, it is reasonable to assume that all pixels belonging to the same block share the same disparity with high probability. Thus, it is natural to use the automatically detected blocks as the matching primitives in our method.
(2) The image pair in this study has been already geometrically rectified, so it is allowed that the search space is limited to the one-dimension direction based on the epipolar constraint.
(3) In this study block matching strategy is embedded into dense matching processing. Thus, a reasonable disparity range, which is crucial to low complexity and high accuracy, can be derived from the dense matching result.
(4) Considering the two images of the image pair in this study are obtained by the same satellite simultaneously, it can be accepted that there is no complex but only linear radiation difference between them, which is very useful for reliable matching cost design.

Overall, water block matching consists of two stages: (1) matching cost calculation and disparity estimation; and (2) mismatch elimination and interpolation. The two stages will be expanded in detail below.

2.3.1. Matching Cost Calculation and Disparity Estimation

Matching cost directly depends on the similarity between the two primitives, the more reasonable the matching cost, the more reliable the matching result. Given that there is only a linear radiation difference between the left image and right image, the variance of the gray difference of corresponding pixels should be relatively small when the block is matched correctly. Based on the above analysis,

a simple but effective matching cost called variance of the gray difference (VGD) is designed, as Equation (3) shows.

$$\begin{cases} \overline{G} = \frac{\sum_{i \in B_L, j \in I_R}(G_i - G_j)}{n} \\ VGD = \frac{\sum_{i \in B_L, j \in I_R}\left((G_i - G_j) - \overline{G}\right)^2}{n} \end{cases} \tag{3}$$

where B_L represents the water block on the left image, I_R is the right image, pixel i and j are a couple of corresponding points with a given disparity, n counts the number of corresponding points, and G_i and G_j are the gray value of i and j, respectively.

According to the disparity range calculated in the former stage, a set of VGD will be calculated. Then, the optimal disparity of the block is estimated by the winner takes all (WTA) principle.

2.3.2. Geometrical Deformation Process

In the above, we assume that the disparities of all pixels belonging to the same block should be equal. In other words, the block is treated as a rigid body on both the left and right images. However, this assumption may not hold when the size of the block exceeds a certain value. That is because there are geometrical deformations between the left and right images due to different imaging angles. To eliminate the influence of the geometrical deformations by controlling the size of the matching primitive, two strategies are introduced to address the deformation along the epipolar line and perpendicular to the epipolar line separately. As Figure 4 shows, we first divide the extracted block (the red frame) into several sub-blocks along the horizontal direction (in this study, epipolar line is assumed to be in the vertical direction). The width of each sub-block will not exceed a given value *TH* (set as 50). More specifically, only the width of the right-most sub-block may be smaller than *TH*, while the widths of other sub-blocks are all equal to *TH*. Then, these sub-blocks will be processed one by one. Subsequently, two end-blocks (the yellow frames) are selected by extending the up-edge and bottom-edge of the middle sub-block with a certain step (set as 5), respectively. The two small-size end-blocks are used as the matching primitives instead of the middle sub-block. After all these end-blocks have been matched, the matching results of all pixel belonging to the original block can be obtained by interpolation.
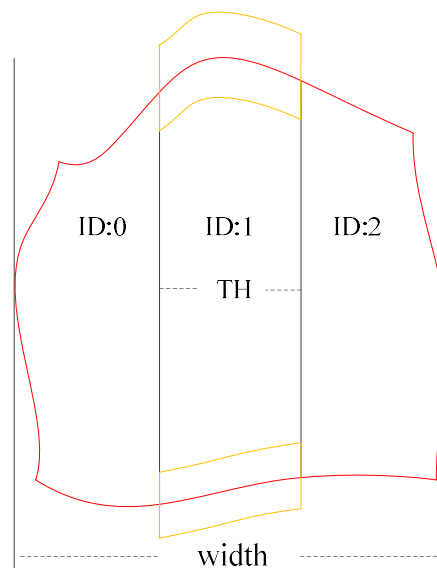

Figure 4. Geometrical deformation process. The red frame is the extracted block, which is divided into three sub-blocks. The yellow frames are the end-blocks of the middle sub-block. *TH* is a given value.

2.3.3. Mismatch Elimination and Interpolation

The end-block matching can be seen as a kind of feature matching, so RANSAC [35] can be used to eliminate the mismatches. Moreover, if one end-block is considered wrongly-matched, it will be interpolated linearly based on the two nearest valid end-block matches on both sides of it. In the absence of valid end-block match on one side, the disparity of the nearest valid end-block match on the other side will be selected as the disparity of this end-block. For each sub-block, the disparity of the pixel between the up and bottom edges are interpolated linearly based on the two end-block matches.

Based on the analysis mentioned above, an automatic water block matching strategy, called end-block matching (EBM), is introduced, as shown in Figure 5. There are three main steps: disparity range calculation; matching cost calculation and disparity estimation; and mismatch elimination and interpolation, which have been introduced in detail above. It is noteworthy that if the width of one extracted block exceeds a given threshold, the block should be divided into several sub-blocks, which will subsequently be processed, one by one.
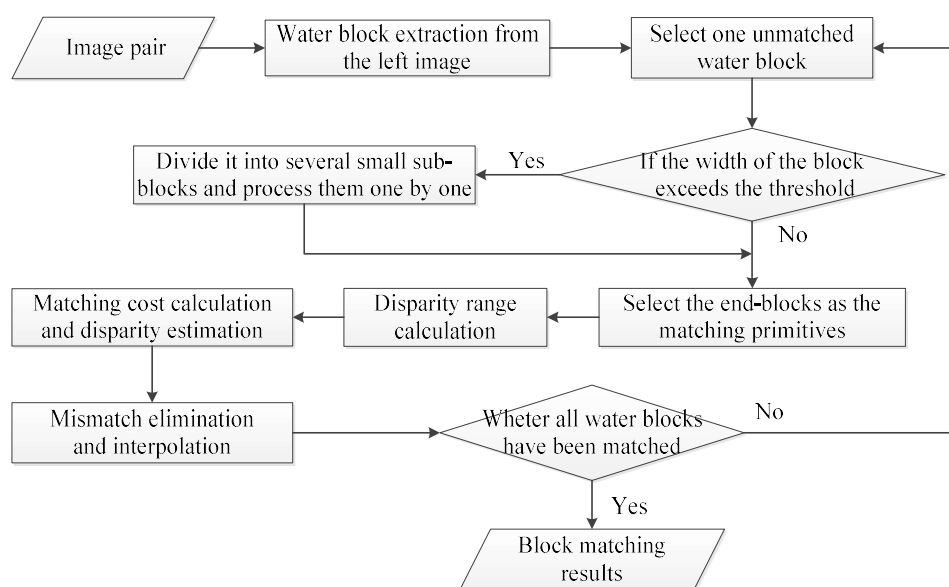

Figure 5. End-block-matching (EBM) strategy.

In addition, it is noted that our method can be seen as a variation of SGM, which integrates block detection and matching. Like SGM, our method is symmetric, which means both the left image and right image can be selected as the reference image. Literature [8] points that a more accurate result can be obtained by fusing the two matching results with left image as the reference image and the right image as the reference image respectively. This fusion strategy is also suitable for the end-block matching. Our experiments indicate that more accurate and complete matching results can be obtained with this fusion strategy, especially for the case where there are obvious differences between the water block extraction results from the left image and that from the right image. In fact, for the sake of efficiency, only the left image or right image is selected as reference image in most of stereo matching algorithms. For comparison, only the left image is selected as the reference image for all experiments in this paper.

2.4. Dense Matching

As mentioned above, for pixels belonging to water blocks, the disparity estimation is realized by EBM, while for pixels beyond these blocks, the disparity estimation is accomplished by dense matching. Dense matching usually consists of four steps, including matching cost computation, cost aggregation, disparity computation and disparity refinement. In our method, census [36], for its high reliability and efficiency, is adopted as matching cost measure. In addition, the boundary of the water blocks are also

treated as the boundary of the reference image in the cost aggregation. As with SGM, more details about these steps can be found in [8].

2.5. DSM Generation

After obtaining the disparity map, 3-dimension point clouds can be achieved based on the RPC model by forward intersection [37]. Considering that the water block should be a plane, then for each water block, the median of the elevation values of all points within it, which has the maximum confidence, will be selected as the final height. To further improve the quality of the generated DSM, a range of processes, including median filter, peak elimination, and interpolation, will be involved [9].

3. Results

3.1. The Experimental Platform and Data

The proposed algorithm is implemented in Visual C++ and a PC with Intel®CoreTM i7-5600U CPU 2.6GHz processors, 8 GB RAM, and Microsoft Windows 2010.

To validate the effectiveness of the proposed method, image blocks, which contain water areas with different sizes, from ZY-3 image pairs were used for evaluation. ZY-3, launched in January 09, 2012, is equipped with three-line scanners (nadir (NAD), backward (BWD) and forward (FWD)) for stereo mapping, the resolutions of the panchromatic (PAN) stereo mapping images are 2.1-m at nadir looking and 3.5-m at tilt angles of ±22° forward and backward looking, respectively. The stereo base-height ratio is 0.85-0.95 [5,38]. The actual heights of these areas are obtained by stereo measurement software. Table 1 lists the information of ZY-3 image pairs. Figure 6 depicts the image pair blocks and Table 2 lists the sizes and the heights of the water areas. We use ROI to represent the water area in the remainder of Section 3 for simplification.

Table 1. Information of ZY-3 image pairs.

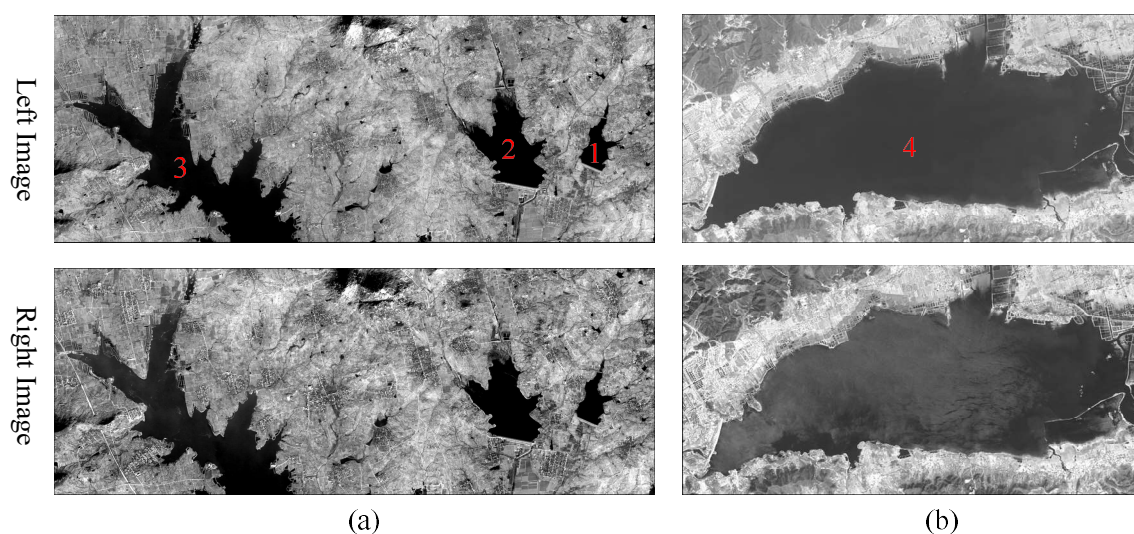|	Satellite Image Pairs
Sensor	ZY-3 FWD/BWD
Image size(pixel)	16281 × 16364
Principal distance(mm)	1700
Pixel size(um)	10
Resolution(m)	3.5
Overlap	>90%

Figure 6. Experimental image pair blocks. (**a**) Image pair block 1. (**b**) Image pair block 2.

Table 2. Sizes and heights of ROIs.

ROI	Pixels	Area (km^2)	Height (m)
1	34329	0.42	116.87
2	136791	1.68	115.11
3	484545	5.94	99.61
4	5376695	65.87	13.00

3.2. Experiment on Seed Point Extraction and ROI Extraction

Accurate ROI extraction result is crucial to accurate disparity estimation of pixels belonging to the ROI. In order to verify the reliable of seed point extraction and ROI extraction. Figure 7 denotes the extracted seed points on the top pyramid image and extracted ROIs on all pyramid images. As the seed points on the lower pyramid image are located based on the seed points on the upper pyramid image, only the seed points on the top pyramid image have been shown. As Figure 7 shows, the seed points and ROIs are both reliable. There is no false seed point and missed seed point on the top pyramid image, which ensures accurate ROI extraction. Furthermore, the consistency between ROIs on the upper pyramid image and ROIs on the lower pyramid image ensures the rationality of the disparity range transmission.
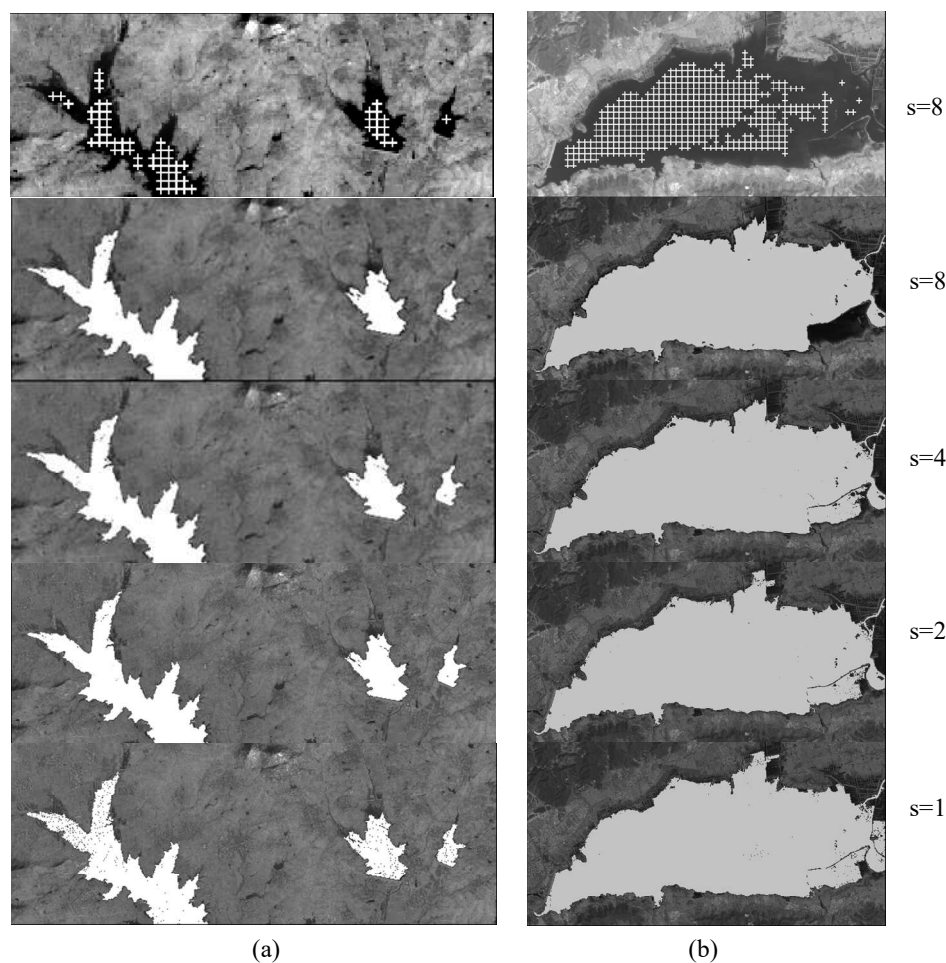

Figure 7. The seed points on the top pyramid image and ROIs on all pyramid images, s is the down sampling scale. (**a**) Image pair block 1. (**b**) Image pair block 2.

The ROI extraction results also were compared with Finer Resolution Observation and Monitoring of Global Land Cover (FROM-GLC) 10, a publicly available global land cover data with 10-m

resolution [39,40]. As shown in Figures 8 and 9, the FROM-GLC10 water masks comparatively accurately cover the water areas. However there are many local boundary mismatches between FROM-GLC10 water masks and the water areas, while the boundaries of the ROIs and the water areas match very well, such as the red frame 1 and red frame 2 in Figures 8 and 9, respectively. The imaging time difference may be a reason for the former local boundary mismatch issue. Moreover, more details similar to dams were saved in the ROIs, which may be related to the improvement of resolution, such as the red frame 3 and red frame 4 in Figure 8 and the red frame 3 in Figure 9. In addition, there are some small voids in the ROIs, which are inevitable in the process of seed point growth. Fortunately, these small voids will be processed by subsequent interpolation and filtering, and have little effect on the final DSM. Generally, publicly available water masks with relatively high accuracy will help to identify water areas to some extent, but inevitably there will be local boundary mismatches which will reduce the accuracy of DSM. Thus, we prefer to directly and automatically extract accurate water areas with elegant strategies.
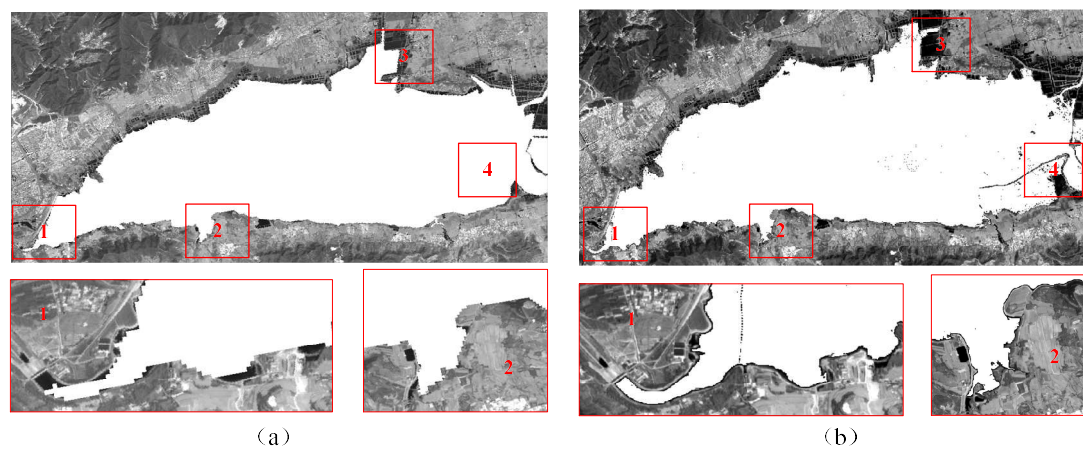

Figure 8. The comparison of the water masks and the ROIs for image pair block 1. (**a**) FROM-GLC10 water masks. (**b**) ROIs.
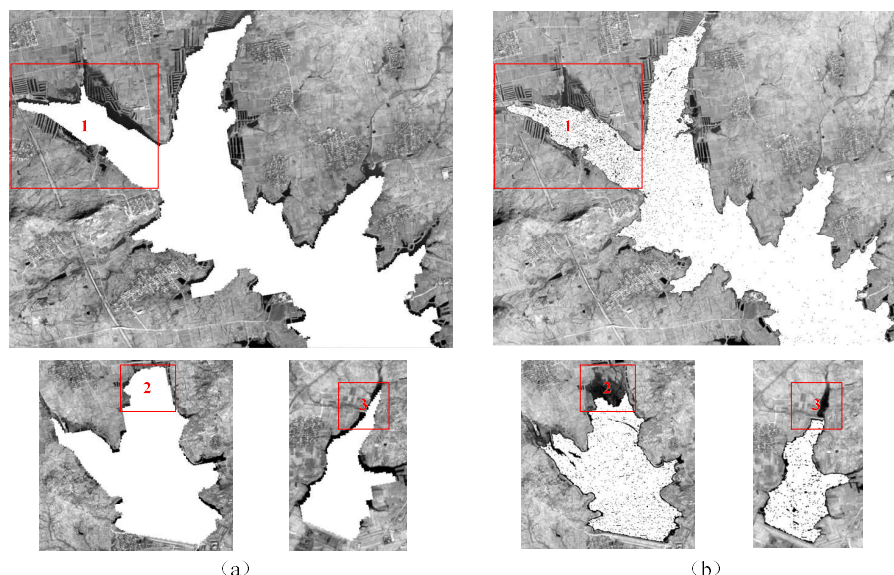

Figure 9. The comparison of the water masks and the ROIs for image pair block 2. (**a**) FROM-GLC10 water masks. (**b**) ROIs.

There are geometrical deformations between the left and right images, and it becomes more and more significant with the increase of size of the image. As shown in Figure 10, the ROIs on the left

image were embedded into the right image with a certain disparity. In Figure 10a, the selected disparity suits to the red frame 1, but does not suit to red frame 2 and 3. In Figure 8b, the selected disparity suits red frame 1, but does not suit red frames 2, 3, and 4. Both results illustrate that there are geometrical deformations in the horizontal direction between the left and right images. Therefore, it is necessary to divide the extracted block into some small sub-blocks to eliminate the deformation. In Figure 10b, the selected disparity suits red frame 5, but does not suit red frame 2. These results demonstrate that there are geometrical deformations in the vertical direction between the left and right images. Thus, the end-block matching strategy is necessary to compensate for the deformation.
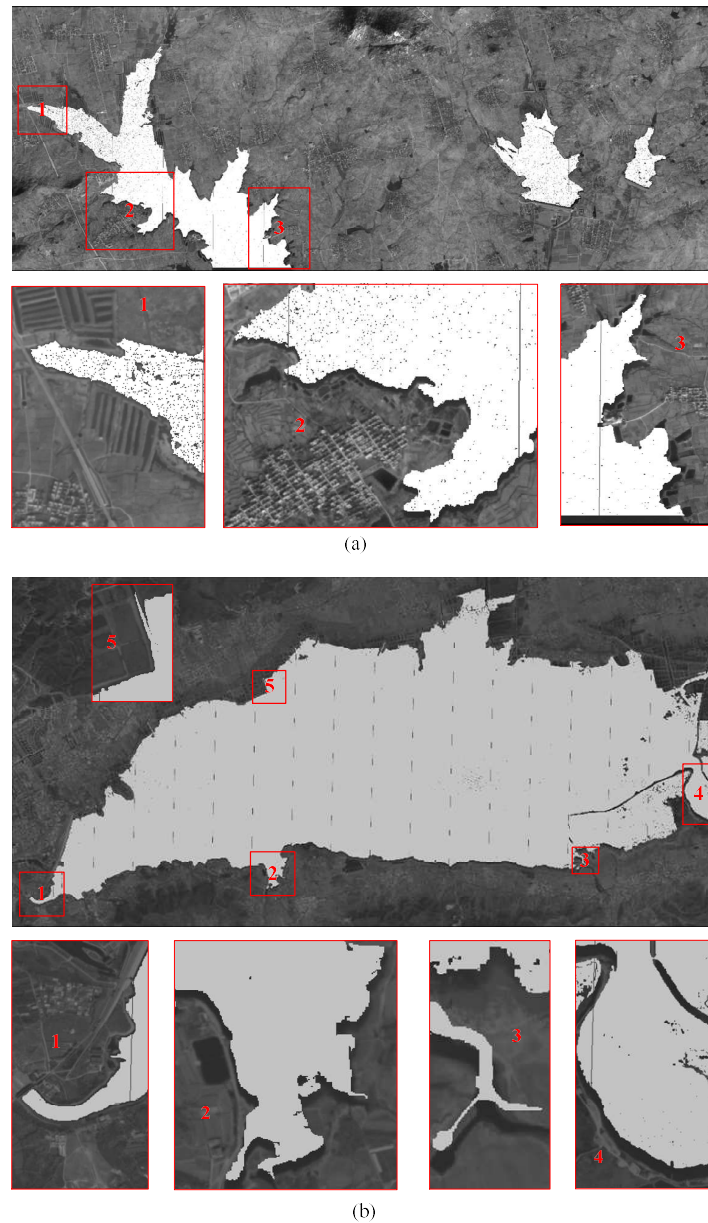

Figure 10. Left image ROIs were embedded into right image. (**a**) Image pair block 1. (**b**) Image pair block 2.

3.3. Experiment on ROI Matching

This section was designed to verify the effectiveness of the proposed block matching strategy and the advantage of handling geometrical deformations. Figures 11 and 12 denote the image pair block 1 ROI matching results and image pair block 2 ROI matching results separately.
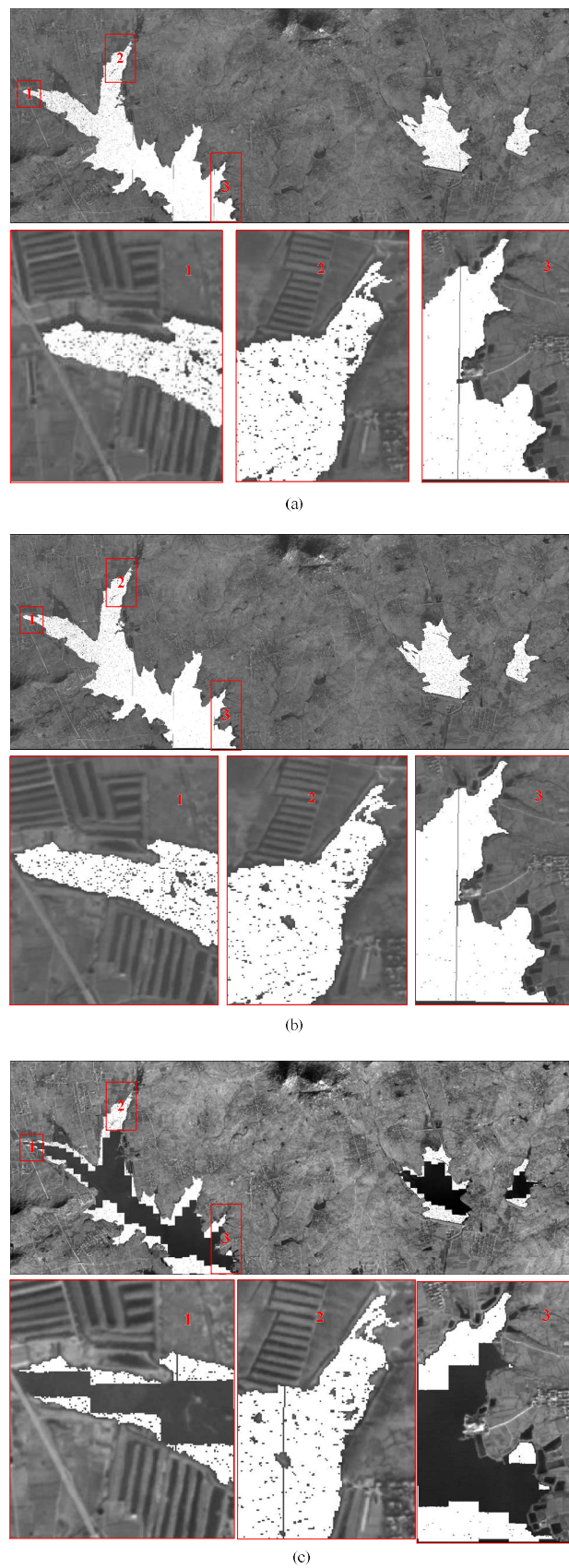
(a)

(b)

(c)

Figure 11. The image pair block 1 ROI matching results. (**a**) Block matching. (**b**) Sub-block matching. (**c**) End-block matching.
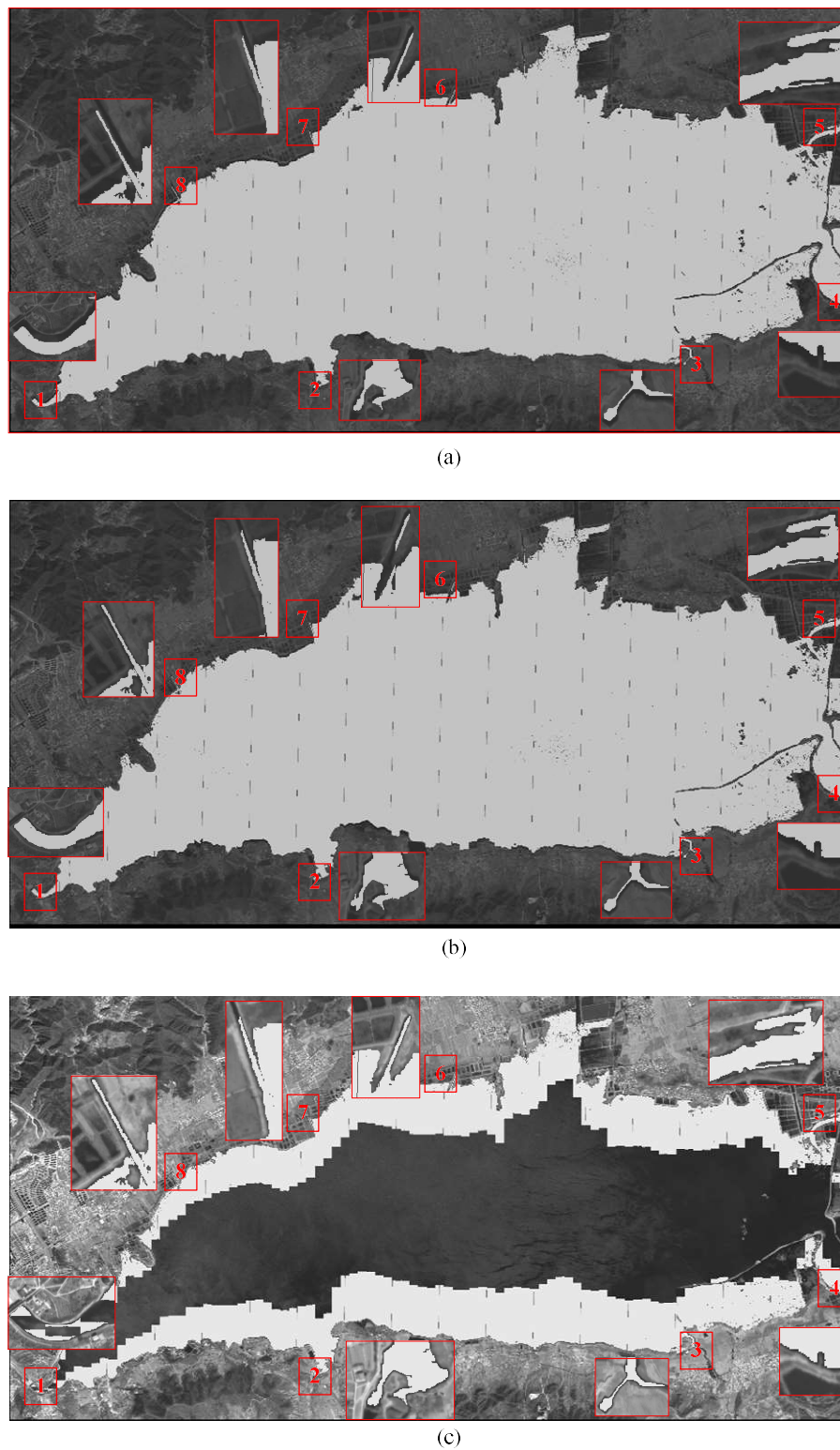
(a)

(b)

(c)

Figure 12. The image pair block 2 ROI matching results. (**a**) Block matching. (**b**) Sub-block matching. (**c**) End-block matching.

It can be seen from Figures 11a and 12a that the block matching strategy cannot take good care of the geometrical deformations between the left and right images. In Figure 11a, only red frame 2 was matched correctly, while red frames 1 and 3 have obvious deviation. In Figure 12a, only the red frame 5 was matched correctly, all other frames have obvious deviation. Moreover, the sub-block matching

strategy can handle the deformation in the horizontal direction well, as Figure 11b shows, red frames 1 and 2 and the top of red frame 3 were matched correctly.

As Figure 12b shows, red frames 1, 5, 7, and 8 were matched correctly. However, the bottom of the red frame 3 (Figure 11b) and red frames 3, 4, and 6 (Figure 12b) still have deviations. This is because that the sub-block matching strategy cannot compensate the deformation in the vertical direction. Eventually, the end-block matching strategy can effectively handle the deformation in both the horizontal and vertical direction. Almost all red frames in Figures 11c and 12c were matched correctly. Only the bottom of red frame 3 (Figure 11c) has some slight deviations, and this is because this part is located at the image boundary and has not enough intensity variation. The proportion of these mismatches to the whole block is small and they will be eliminated in DSM generation step by the median strategy.

The right two ROIs in Figure 11 have relatively small sizes, and they are less affected by geometric deformations. Thus, there is no obviously difference between the different matching strategies for them.

Figure 13 shows the ROI matching results of different block matching strategies and can be used for a further analysis of the performance of them. All ROI matching results were sampled and scaled for easy of drawing, however, the relationship between them has not changed. From Figure 13a,b, the result of block matching strategy is significantly affected by geometrical deformations. Moreover, the sub-block matching strategy performs much better in terms of the deformation in the horizontal direction but still has difficulty with the deformation in the vertical direction. By comparing Figures 13a and 13b, we can conclude that the effect of the deformation in the vertical direction will get strong when the height of ROI becomes larger. Fortunately, the end-block matching strategy can handle the deformation in both the horizontal and vertical directions well, and is insensitive to the size of the ROI.
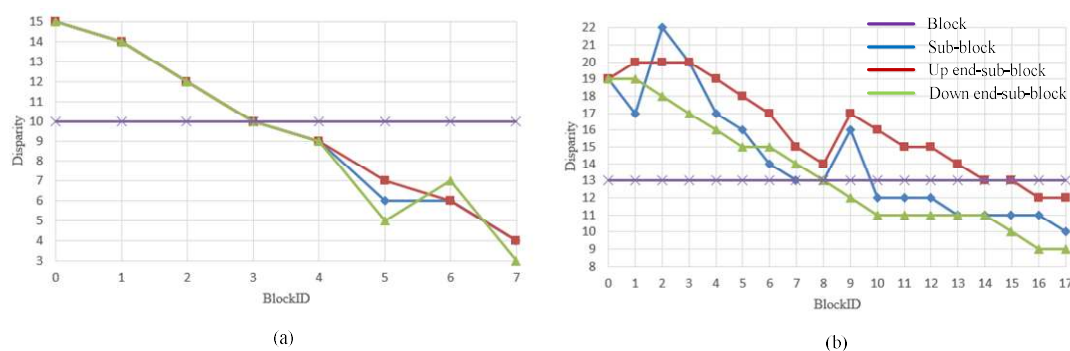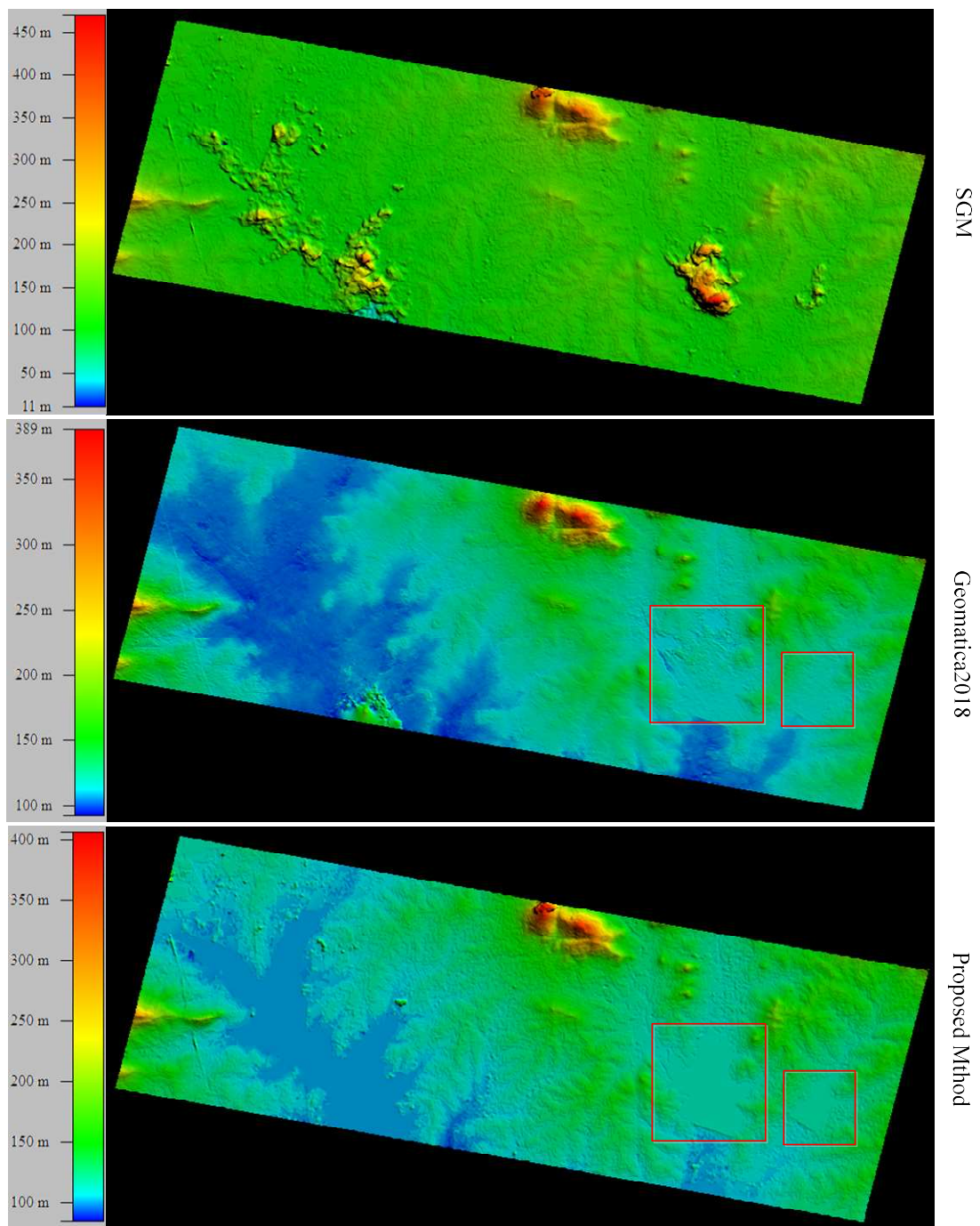

Figure 13. The ROI matching results. (**a**) Image pair block 1. (**b**) Image pair block 2.
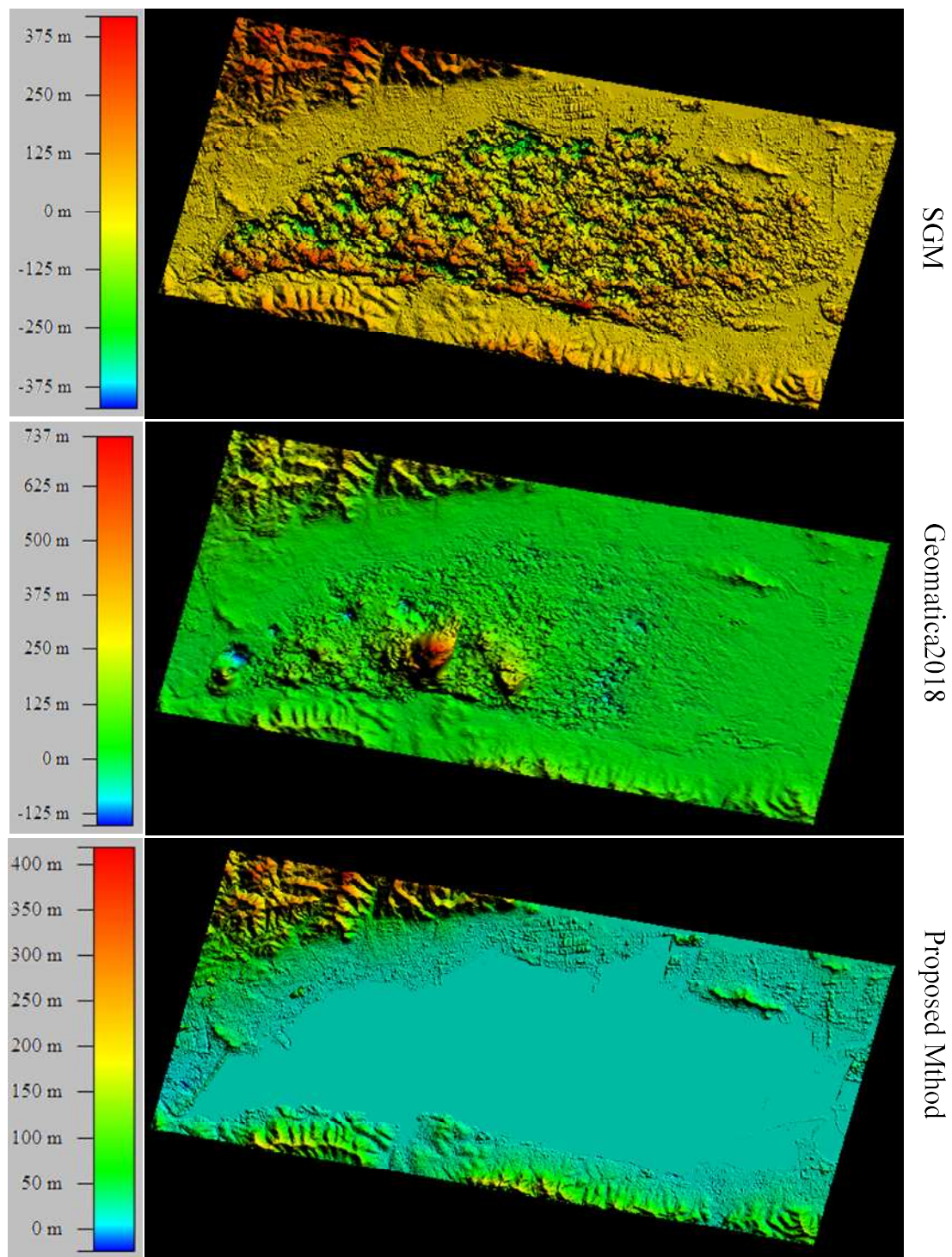
3.4. DSM Quality Assessment

For quality assessment, the DSMs generated by our method were compared with the DSMs generated by both SGM and the commercial software Geomatica2018. Geomatica2018 is one of best performing softwares, in terms of DSM generation from satellite image pairs, and the core algorithm is an improved SGM. The proposed method is somewhat similar to SGM, however employs block matching instead of dense matching in ROIs. Therefore, a comparison with both SGM and Geomatica2018 is necessary.

Figure 14 shows the DSMs generated by the three approaches. There are obvious errors in the DSMs generated by SGM in ROIs. This is because the matching primitive used by SGM does not has enough uniqueness. Although matching cost aggregation is adopted, these errors still cannot be eliminated. Geomatica2018 has better performance in these ROIs, even better in the ROIs with small size. However, the program still has difficult with the ROIs of large size, and the result is over-smooth. The proposed SGBM has obviously better performance than the two other methods Benefiting from the adaptive block matching strategy, it can obtain smooth elevation surface in ROIs and keep the boundaries simultaneously. Experiments show that the DSMs in ROIs generated by SGBM almost do not need extra manual editing.

(a)

Figure 14. *Cont.*

(b)

Figure 14. The DSMs generated by SGM, Geomatica2018 and the proposed SGBM. (**a**) Image pair block 1. (**b**) Image pair block 2.

To quantitatively evaluate the performance of our method, three typical indicators were selected, including root mean square error (RMSE), variance, and time consumption. RMSE represents the difference between the DSM height and the actual height of all points in the ROI. Variance reflects whether the height between pixels belonging to the same ROI are consistent. In fact, pixels belonging to the same ROI should typically have the same height.

The results are shown in Table 3. For the four ROIs in this experiment, SGM has the worst accuracy ranging from tens of meters to hundreds of meters. Geomatica2018 has relatively good accuracy in terms of meters in ROI1 and ROI2, which have small size, but has poor accuracy in terms of tens of meters in ROI3 and ROI4, which have relatively large sizes. SGBM has the best accuracy in terms of tens of centimeters, and the variance is always zero. Moreover, the accuracy of SGBM is insensitive to the size of the ROI. In terms of efficiency, Geomatica2018 has the lowest time consumption. SGBM is faster than SGM, as block detection and matching require less computation than dense matching in the ROI, and even less when the size of the ROI is large.

Table 3. Quantitative Evaluation.

Method	Indicator	Image Pair Block1 Raster: 4010 × 1510 pix			Image Pair Block2 Raster: 5010 × 2510 pix
		ROI1	ROI2	ROI3	ROI4
SGM	RMSE/m	25.31	125.85	51.26	84.65
	Variance/m	19.14	86.42	37.92	83.05
	Time/min		6		14
Geomatica2018	RMSE/m	2.13	2.87	11.68	79.71
	Variance/m	1.36	2.34	10.09	73.92
	Time/min		2		4
SGBM	RMSE/m	0.33	0.54	0.38	0.28
	Variance/m	0	0	0	0
	Time/min		4		9

4. Discussion

4.1. Advancements of The Proposed Method

In contrast to the traditional dense matching methods that obtain the elevation pixel by pixel, this paper introduces a DSM generation approach SGBM, aiming at improving the elevation estimation accuracy of water areas by adaptive block matching. The proposed SGBM was based on SGM. But an effective strategy for water areas was performed and verified by the experimental results. To be more specific, water areas were extracted, then an adaptive block matching strategy was employed to achieve accurate disparity estimation for these areas. Disparity estimation for other pixels was conducted by SGM. As analyzed in Section 1.2, for poor texture areas, the typical matching primitive, a window centered at a selected pixel with a certain size, does not have enough uniqueness, leading to high ambiguity between the candidates. Fortunately, the proposed EBM strategy can handle this problem well and achieve accurate disparity estimation in water areas.

Our adaptive block matching strategy has three advantages: (1) The blocks only be extracted in the left image. (2) The effect of geometrical deformations can be eliminated by EBM, which was elaborated in Section 2.3.2. (3) The computation cost of disparity estimation by the block matching strategy is much-lower than by dense matching in water areas. Consequently, the DSMs in water areas generated by our method have high quality and subsequent laborious manual editing could be significantly reduced.

4.2. Limitations of the Proposed Method

According to the analysis of block matching in Section 2.3 and the experimental results, the proposed method is based on the assumption that the pixels belonging to the same block have almost the same height, therefore, it is only applicable for water areas and other plain terrain areas. Moreover, the DSMs in water areas generated by the proposed method depend on the block extraction results, which are directly related to the seed point detection and growth strategy. With different parameters, the corresponding results greatly vary from each other, thus, the parameters

may need to be adjusted when it comes to different remote sensing stereo images, which potentially impedes the applicability of our method. Nevertheless, the water block extraction module is designed as a plug-in in this method, which means that other more effective water block extraction methods or reliable water masks can be introduced conveniently.

5. Conclusions

This paper proposes a novel DSM generation approach called SGBM, which effectively solves the height estimation problem in poor texture areas that have almost constant height, represented by the water area. Different from conventional methods that fill the data voids by manual editing and auxiliary data after DSM generation, our method directly acquires reliable DSM by handling water areas and other areas separately and almost requires no manual editing. An adaptive block matching strategy is employed instead of a dense matching strategy to achieve accurate disparity estimation in water areas, and the computational redundancy is reduced. The experimental results demonstrate that the precision of DSMs in these areas generated by SGBM have a significant improvement compared with the DSMs generated by SGM and Geomatica2018, and that they can meet the requirements of actual production. Moreover, SGBM is robust and insensitive to the size of the water area. It is an alternative component to facilitate the quality of SGM for DSM generation.

For now, SGBM is only suitable to poor texture areas with almost constant height. In future work, a more flexible block matching strategy for height fluctuation should be investigated, and more different stereo images should also be tested. Additionally, other image knowledge, such as line features, are worthy of more study to further improve the DSM quality.

References

1. Rabus, B.; Eineder, M.; Roth, A.; Bamler, R. The shuttle radar topography mission-A new class of digital elevation models acquired by spaceborne radar. *ISPRS J. Photogramm. Remote Sens.* **2003**, *57*, 241–262. [CrossRef]
2. Farr, T.G. The shuttle radar topography mission. *Rev. Geophys.* **2007**, *45*, 1–33. [CrossRef]
3. Toutin, T. Comparison of stereo-extracted DTM from different high-resolution sensors: SPOT-5, EROS-a, IKONOS-II, and QuickBird. *IEEE Trans. Geosci. Remote Sens.* **2004**, *42*, 2121–2129. [CrossRef]
4. Aguilar, M.A.; Saldana, M.D.M.; Aguilar, F.J. Generation and Quality Assessment of Stereo-Extracted DSM from GeoEye-1 and WorldView-2 Imagery. *IEEE Trans. Geosci. Remote Sens.* **2014**, *52*, 1259–1271. [CrossRef]
5. Xue, Y.; Xie, W.; Du, Q.; Sang, H. Performance evaluation of DSM extraction from ZY-3 three-line arrays imagery. *ISPRS Arch.* **2015**, *XL-7/W4*, 247–253. [CrossRef]
6. Scharstein, D.; Szeliski, R. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *Int. J. Comput. Vis.* **2002**, *47*, 7–42. [CrossRef]
7. Zhang, L.; Gruen, A. Multi-image matching for DSM generation from IKONOS imagery. *ISPRS J. Photogramm. Remote Sens.* **2006**, *60*, 195–211. [CrossRef]
8. Hirschmüller, H. Stereo processing by semiglobal matching and mutual information. *IEEE Trans. Pattern Anal.* **2008**, *30*, 328–341. [CrossRef]

9. Hu, T.; Guo, B.; Wu, H.; Gong, J.; Zhang, T. Dense corresponding pixel matching between aerial epipolar images using an RGB-belief propagation algorithm. *IEEE Geosci. Remote Sens. Lett.* **2012**, *9*, 1036–1040. [CrossRef]

10. Labatut, P.; Pons, J.P.; Keriven, R. Efficient multi-view reconstruction of large-scale scenes using interest points, delaunay triangulation and graph cuts. In Proceedings of the International Conference on Computer Vision (ICCV), Rio De Janeiro, Brazil, 14–21 October 2007.

11. Furukawa, Y.; Ponce, J. Accurate, dense, and robust multi-view stereopsis. *IEEE Trans. Pattern Anal. Mach. Intell.* **2010**, *32*, 1362–1376. [CrossRef]

12. GEOMATICA. Available online: https://www.nrcresearchpress.com/loi/cig (accessed on 15 September 2019).

13. Trimbpe Inpho. Available online: https://geospatial.trimble.com/products-and-solutions/inpho (accessed on 10 October 2019).

14. Zhang, Y.; Zhang, Y.; Mo, D.; Zhang, Y.; Li, X. Direct digital surface model generation by semi-global vertical line locus matching. *Remote Sens.* **2017**, *9*, 214. [CrossRef]

15. Wu, B.; Zhang, Y.; Zhu, Q. Integrated point and edge matching on poor textural images constrained by self-adaptive triangulations. *ISPRS J. Photogramm. Remote Sens.* **2012**, *68*, 40–55. [CrossRef]

16. Wang, D.; Liu, H.; Cheng, X. A Miniature Binocular Endoscope with Local Feature Matching and Stereo Matching for 3D Measurement and 3D Reconstruction. *Sensors* **2018**, *18*, 2243. [CrossRef] [PubMed]

17. Mandlburger, G. A case study on through-water dense image matching. *ISPRS Arch.* **2018**, *XLII-2*, 659–666. [CrossRef]

18. Kanade, T.; Okutomi, M. A stereo matching algorithm with an adaptive window: Theory and experiment. In Proceedings of the IEEE International Conference on Robotics and Automation, Washington DC, USA, 10–17 May 2002.

19. Ye, L.; Wu, B. Integrated image matching and segmentation for 3D surface reconstruction in urban areas. *Photogramm. Eng Remote Sens.* **2018**, *84*, 135–148. [CrossRef]

20. Hu, H.; Rzhanov, Y.; Hatcher, P.J.; Bergeron, R.D. Binary adaptive semi-global matching based on image edges. In Proceedings of the 7th International Conference on Digital Image Processing, Los Angeles, CA, USA, 9–10 April 2015.

21. Paris, S.; Durand, F. A fast approximation of the bilateral filter using a signal processing approach. In Proceedings of the European Conference on Computer Vision, Graz, Austria, 7–13 May 2006; Springer: Berlin/Heidelberg, Germany, 2006.

22. Shan, Q.; Curless, B.; Furukawa, Y.; Hernandez, C.; Seitz, S.M. Occluding contours for multi-view stereo. In Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014.

23. Shao, Z.F.; Yang, N.; Xiao, X.; Zhang, L.; Peng, Z. A multi-view dense point cloud generation algorithm based on low-altitude remote sensing images. *Remote Sens.* **2016**, *8*, 381. [CrossRef]

24. Slater, J.A.; Garvey, G.; Johnston, C.; Haase, J.; Heady, B.; Kroenung, G.; Little, J. The srtm data "finishing" process and products. *Photogramm. Eng. Remote Sens.* **2006**, *72*, 237–247. [CrossRef]

25. Reuter, H.I.; Nelson, A.; Jarvis, A. An evaluation of void-filling interpolation methods for SRTM data. *Int. J. Geogr. Inf. Sci.* **2007**, *21*, 983–1008. [CrossRef]

26. Grimaldi, S.; Teles, V.; Bras, R.L. Sensitivity of a physically based method for terrain interpolation to initial conditions and its conditioning on stream location. *Earth Surf Proc Land.* **2004**, *29*, 587–597. [CrossRef]

27. Grimaldi, S.; Teles, V.; Bras, R.L. Preserving first and second moments of the slope area relationship during the interpolation of digital elevation models. *Adv. Water Resour.* **2005**, *28*, 583–588. [CrossRef]

28. Pekel, J.; Cottam, A.; Gorelick, N.; Belward, A.S. High-resolution mapping of global surface water and its long-term changes. *Nature* **2016**, *540*, 1–19. [CrossRef] [PubMed]

29. Feng, M.; Sexton, J.O.; Channan, S.; Townshend, J.R. A global, high-resolution (30-m) inland water body dataset for 2000: First results of a topographic-spectral classification algorithm. *Int. J. Digit. Earth* **2016**, *9*, 113–133. [CrossRef]

30. Wang, Q.; Yuan, J.Y.; Li, B.-L. Matching algorithm integrated feature points and contour information. *Appl. Res. Comput.* **2014**, *31*, 3145–3147.

31. Lowe, D.G. Distinctive image feartures from scal-invariant keypoints. *Int. J. Comput. Vis.* **2004**, *60*, 91–110. [CrossRef]

32. Bay, H.; Tuytelaars, T.; Gool, L.V. SURF: Speeded up robust features. *Comput. Vis. Image Und.* **2006**, *110*, 404–417.

33. Perri, S.; Colonna, D.; Zicari, P.; Corsonello, P. SAD-based stereo matching circuit for FPGAs. In Proceedings of the 13th IEEE International Conference on Electronics, Circuits and Systems, Nice, France, 10–13 December 2006.

34. Pourmohammad, A.; Poursajadi, S.; Karimifar, S. Scene matching NCC value improvement based on contrast matching. In Proceedings of the IEEE 8th Iranian Conference on Machine Vision and Image Processing, Zanjan, Iran, 10–12 September 2013.

35. Bolles, R.C.; Fischler, M.A. A RANSAC-Based Approach to Model Fitting and Its Application to Finding Cylinders in Range Data. In Proceedings of the IJCAI, Vancouver, BC, Canada, 24–28 August 1981; Volume 1981, pp. 637–643.

36. Humenberger, M.; Engelke, T.; Kubinger, W. A census-based stereo vision algorithm using modified semi-global matching and plane fitting to improve matching quality. In Proceedings of the Computer Vision and Pattern Recognition Workshops, San Francisco, CA, USA, 13–18 June 2010.

37. Noh, M.J.; Howat, I.M. Automatic relative RPC image model bias compensation through hierarchical image matching for improving DEM quality. *ISPRS J. Photogramm. Remote Sens.* **2018**, *136*, 120–133. [CrossRef]

38. China Centre for Resource Satellite Data and Application. Available online: http://www.cresda.com/CN/Satellite/3070.shtml (accessed on 10 April 2019).

39. Gong, P.; Liu, H.; Zhang, M.; Li, C.; Wang, J.; Huang, H.; Clinton, N.; Ji, L.; Li, W.; Bai, Y.; et al. Stable classification with limited sample: Transferring a 30-m resolution sample set collected in 2015 to mapping 10-m resolution global land cover in 2017. *Sci. Bull.* **2019**, *64*, 370–373. [CrossRef]

40. Finer Resolution Obsevation and Monitoring-Global Land Cover. Available online: http://data.ess.tsinghua.edu.cn/fromglc10_2017v01.html (accessed on 20 January 2020).